Fourth International Workshop

in

Sequential Methodologies

THE UNIVERSITY OF GEORGIA

DEPARTMENT OF STATISTICS

FRANKLIN COLLEGE OF ARTS AND SCIENCES

July 17 – 21, 2013

Conference Organizers:

T. N. Sriram
Nitis Mukhopadhyay

Local Organizing Committee:

Abhyuday Mandal
T. N. Sriram

Student Volunteers:

Adrijo Chakraborty
Haileab Hilafu
Linwei Hu
Jacob Martin
Yaser Samadi
Wenhui Sheng
Xijue Tan
Guannan Wang
Wenbo Wu

Staff Support:

Tim Cheek
Megan Weatherford

Sponsors:

Fourth International Workshop in Sequential Methodologies acknowledges the generous support from the following sponsors (in alphabetical order)

- Aptiv Solutions
- Department of Statistics, University of Georgia
- Franklin College of Arts and Sciences, University of Georgia
- Institute of Mathematical Statistics
- National Institute of Statistical Sciences
- Office of the Senior Vice President for Academic Affairs and Provost, University of Georgia
- Office of the Vice President for Research, University of Georgia
- Taylor & Francis
- The Graduate School, University of Georgia

List of Sessions

Wednesday, July 17, 2013		
7 - 9pm		
Pecan Galleria		Reception

Thursday, July 18, 2013		
7 - 8am		
Banquet Area		Breakfast
8 - 8:30am		
Masters Hall		Opening Remarks
8:30 - 9:45am		
Masters Hall	Plenary Session 1	Sequential Hypothesis Tests: Historical Overview and Recent Results
9:45 - 10:15am		
Pecan Galleria		Refreshment break
10:15 - 11:45am		
Room TU	ThuAM-InvPapSess 1	Adaptive Designs in Clinical Trials
Room VW	ThuAM-InvPapSess 2	Change Detection in Time Series
Room YZ	ThuAM-InvPapSess 3	Multiple Comparisons in Sequential Experiments
11:45am - 1:15pm		
Banquet Area		Lunch
1:15 - 2:45pm		
Room TU	ThuPM-InvPapSess 1	Design of experiments
Room VW	ThuPM-InvPapSess 2	Applications of Change-Point Detection
Room YZ	ThuPM-InvPapSess 3	Sequential Methods in the Hands of Young Researchers I
2:45 - 3:15pm		
Concourse		Refreshment break
3:15 - 4:45pm		
Room VW	ThuPM-InvPapSess 4	Optimal Stopping and Sequential Statistics
Room YZ	ThuPM-InvPapSess 5	Recent Advances in Sequential Methodologies with Applications - I

Friday, July 19, 2013

7 - 8:30am

Banquet Area		Breakfast

8:30 - 9:45am

Masters Hall	Plenary Session 2	Exact Distributions of Stopping Times in Two-Stage and Sequential Sampling

9:45 - 10:15am

Pecan Galleria		Refreshment break

10:15 - 11:45am

Room TU	FriAM-InvPapSess 1	Change Detection in Functional Sequences - I
Room VW	FriAM-InvPapSess 2	Sequential Estimation for Dependent Data - I
Room YZ	FriAM-InvPapSess 3	Sequential Methodologies and High-Dimensional Data Analysis

11:45am - 1:15pm

Banquet Area		Lunch

1:15 - 2:45pm

Room TU	FriPM-InvPapSess 1	Recent Advances in Sequential Methodologies with Applications - II
Room VW	FriPM-InvPapSess 2	Recent Results in Sequential analysis and Change-Point Analysis
Room YZ	FriPM-InvPapSess 3	Methodologies for High-Dimensional Data Analysis - I

2:45 - 3:15pm

Concourse		Refreshment break

3:15 - 4:45pm

Room TU	FriPM-InvPapSess 4	Sequential Methods in the Hands of Young Researchers - II
Room VW	FriPM-InvPapSess 5	Sequential Change Point Detection

7 - 10pm

Banquet Area		Banquet

Saturday, July 20, 2013

7 - 8:30am

Banquet Area		Breakfast

8:30 - 9:45am

Masters Hall	Plenary Session 3	Effective Methodologies for High-Dimensional Data

9:45 - 10:15am

Pecan Galleria		Refreshment break

10:15 - 11:45am

Room TU	SatAM-InvPapSess 1	Applications of Sequential Analysis
Room VW	SatAM-InvPapSess 2	Change Detection in Functional Sequences - II
Room YZ	SatAM-InvPapSess 3	Sensor Exploitation

11:45am - 1:15pm

Banquet Area		Lunch

1:15 - 2:45pm

Room TU	SatPM-InvPapSess 1	Recent Advances in Sequential Change Detection
Room VW	SatPM-InvPapSess 2	Sequential Inference, Change-Point Detection and Clinical Trials

2:45 - 3:15pm

Concurse		Refreshment break

6 - 7pm

Pecan Galleria		Social Hour

Sunday, July 21, 2013

7 - 8:15am

Banquet Area		Breakfast

8:15 - 9:45am

Room TU	SunAM-InvPapSess 1	Change Point Detection in Skew Distributions and Related Topics
Room VW	SunAM-InvPapSess 2	Sequential Inference

9:45 - 10:15am

Concourse		Refreshment break

10:15 - 10:45am

Masters Hall		Abraham Wald Prize Ceremony

10:45 - 12:00noon

Masters Hall	Plenary Session 4	Nonparametric Monitoring of Time Series

12:00 - 12:30pm

Masters Hall		Closing Remarks

12:30pm

Connector		Boxed Lunch

Fourth International Workshop in Sequential Methodologies

Program Schedule

Invited Sessions and Speakers

Thursday, July 18, 2013, 8:00am - 8:30am

Opening Ceremony

Thursday, July 18, 2013, 8:30am - 9:45am

Plenary Lecture 1
Chair: T. N. Sriram, University of Georgia-Athens, USA
Alexander Tartakovsky, University of Southern California, USA: *Sequential Hypothesis Tests: Historical Overview and Recent Results*

Thursday, July 18, 2013, 10:15am - 11:45am

ThuAM-InvPapSess 1: Adaptive Designs in Clinical Trials
Organizer: William F. Rosenberger, George Mason University, USA
Chair: Anastasia Ivanova, University of North Carolina, USA
Qing Liu, Janssen Pharmaceuticals, USA: *Fisherian Evidential Approach to Sequential Clinical Trials*
Vladimir Dragalin, AptivSolutions, USA: *Adaptive Clinical Trials with Population Enrichment Design*
Alex Sverdlov, Novartis, USA: *Novel Response-Adaptive Designs for Clinical Trials with Time-to-Event Outcomes*

ThuAM-InvPapSess 2: Change Detection in Time Series
Organizers: Edit Gombay and Abdulkadir Hussein, University of Alberta, Canada
Chair:Ansgar Steland, RWTH Aachen University, Germany
Steve Coad, Queen Mary, University of London, UK: *Estimation of Parameters of the Absolute Autoregressive Model*
Abdulkadir Hussein, University of Windsor, Canada: *Issues and Remedies in Surveillance of Binary Outcomes*
Edit Gombay, University of Alberta, Canada: *Change Detection for Binary Time Series*

ThuAM-InvPapSess 3: Multiple Comparisons in Sequential Experiments
Organizer: Michael Baron, University of Texas at Dallas, USA
Chair: Michael Baron, University of Texas at Dallas, USA
Venugopal V. Veeravalli, University of Illinois at Urbana-Champaign, USA: *Controlled Sensing for Multihypothesis Testing*
Shyamal K. De, Binghamton University, New York, USA: *Sequential Multiple Testing Controlling Generalized Error Rates*
Kartlos Kachiashvili, Georgian Technical University, Republic of Georgia: *Sequential Analysis methods of Bayesian Type for Testing Hypotheses*

Thursday, July 18, 2013, 1:15pm - 2:45pm

ThuPM-InvPapSess 1: Design of experiments
Organizers: Yajun Mei and Abhyuday Mandal, Georgia Institute of Technology & University of Georgia, USA
Chair: Yajun Mei, Georgia Institute of Technology, USA
Xin Wang & Richard W. Lu, Georgia Institute of Technology, USA: *Layers of Experiments with Adaptive Combined Design*
Abhyuday Mandal, University of Georgia, USA: *Optimal Designs for Two-Level Factorial Experiments with Binary Response*
Ying Hung & Huijuan Li, Rutgers University, USA: *Adaptive Latin Hypercube Designs for Computer Experiments*

ThuPM-InvPapSess 2: Applications of Change-Point Detection
Organizer: Cheng-Der Fuh, National Central University, Taiwan
Chair: Vasanthan Raghavan, University of Southern California, USA
Robert Lund, Clemson University, USA: *Multiple Change Point Detection*
Yao Xie, Duke University & Georgia Institute of Technology, USA: *Detecting Change-Point in Signal Correlation*
Yuan Wang, Georgia Institute of Technology, USA: *Efficient Sequential Monitoring of Multiple Data Streams via Shrinkage*

ThuPM-InvPapSess 3: Sequential Methods in the Hands of Young Researchers - I
Organizer: Nitis Mukhopadhyay, University of Connecticut-Storrs, USA
Chair: Nitis Mukhopadhyay, University of Connecticut-Storrs, USA
Kazuyoshi Yata, University of Tsukuba, Ibaraki, Japan: *Asymptotic Normality for Inference on Multi-Sample, High-Dimensional Mean Vectors under Mild Conditions*
Bhargab Chattopadhyay, University of Texas-Dallas, Texas, USA: *Two-Stage Fixed-Width Confidence Interval of Nonparamteric Regression parameters Using Nadaraya-Watson Estimator*
Bruno Buonaguidi, Bocconi University, Milan, Italy: *Recent Developments on Sequential Testing for Levy Processes*

Thursday, July 18, 2013, 3:15pm - 4:45pm

ThuPM-InvPapSess 4: Optimal Stopping and Sequential Statistics
Organizer: Albrecht Irle, University of Kiel, Germany
Chair: Alexander Tartakovsky, University of Southern California, USA
Sören Christensen, University of Kiel, Germany: *Representations of Excessive Functions and Their Application to Optimal Stopping Problems*
Hans Rudolf Lerche, University of Freiburg, Germany: *Overshoot and Optimality in Sequential Testing*
Alex Novikov, University of Technology Sydney, Australia: *Bayesian Sequential Estimation of a Drift of Fractional Brownian Motion*

ThuPM-InvPapSess 5: Recent Advances in Sequential Methodologies with Applications - I
Organizer: Tumulesh K. S. Solanky, University of New Orleans, Louisiana, USA
Chair: Elena M. Buzaianu, University of North Florida, USA
Hokwon Cho, University of Nevada, Las Vegas, USA: *Statistical Inference of a Measure of Reduction for Two Binomial Variates*
Elena M. Buzaianu, University of North Florida, USA: *A Two-Stage Selection and Testing Procedure for Comparing Several Treatments with a Control*
Joshua McDonald & David Goldsman, Georgia Institute of Technology, USA: *Conditional Probability of Correct Selection after Procedure Termination*

Friday, July 19, 2013, 8:30am - 9:45am

Plenary Lecture 2
Chair: Hans Rudolf Lerche, University of Freiburg, Germany
Shelemyahu Zacks, SUNY Binghamton, New York, USA: *Exact Distributions of Stopping Times in Two-Stage and Sequential Sampling*

Friday, July 19, 2013, 10:15am - 11:45am

FriAM-InvPapSess 1: Change Detection in Functional Sequences - I
Organizer: Eric Chicken, Florida State University, USA
Chair: Alex Novikov, University of Technology Sydney, Australia
Peihua Qiu, University of Florida, USA: *Some Recent Research On Nonparametric Profile Monitoring*
Vasanthan Raghavan, University of Southern California, USA: *Multi-Sensor Change Detection with Change Propagation*

FriAM-InvPapSess 2: Sequential Estimation for Dependent Data - I
Organizers: Leonid Galtchouk, University of Strasbourg, France & S. Pergamenchtchikov, University of Rouen, France
Chair: Igor Nikiforov, Université de Technologie de Troyes, France
Ouerdia Arkoun, University of Rouen, France: *Sequential Robust Efficient Adaptive Estimation for Nonparametric Autoregressive Models*
Serguei Pergamenchtchikov, University of Rouen, France: *Minimax Sequential Kernel Estimators for Nonparametric Diffusion Processes*
Yaser Samadi, University of Georgia, USA: *Sequential Fixed-Width Confidence Interval Based on Bhattacharyya-Hellinger Distance: The Nonparametric Case*

FriAM-InvPapSess 3: Sequential Methodologies and High-Dimensional Data Analysis
Organizer: T. N. Sriram, University of Georgia, USA
Chair: Wenbo Wu, University of Georgia, USA
Moshe Pollak, The Hebrew University of Jerusalem, Israel: On Reaching Nirvana (a.k.a. Steady State)
Yu Liu, University of New Orleans, USA: Performance Analysis of Sequential Probability Ratio Test
Umashanger Thayasivam, Rowan University, USA: Unsupervised Anomaly Detection for High Dimensional Data

Friday, July 19, 2013, 1:15pm - 2:45pm

FriPM-InvPapSess 1: Recent Advances in Sequential Methodologies with Applications - II
Organizer: Tumulesh K. S. Solanky, University of New Orleans, USA
Chair: Tumulesh K. S. Solanky, University of New Orleans, USA
Tumulesh K. S. Solanky, University of New Orleans, USA: *A Note on Partitioning Exponential Populations*
Nitis Mukhopadhyay, University of Connecticut-Storrs, USA: *On Determination of an Appropriate Pilot Sample Size*
Tung-Lung Wu, University of Connecticut-Storrs, USA: *A Sequential Procedure for Multiple Window Scan Statistics*

FriPM-InvPapSess 2: Recent Results in Sequential analysis and Change-Point Analysis
Organizers: Shelly Zacks & Aleksey Polunchenko, Binghamton University, USA
Chair: Sören Christensen, University of Kiel, Germany
Marlo Brown, Niagara University, USA: *Detection of Changes of Multiple Poisson Processes Monitored at Discrete Time Points Where the Arrival Rates Are Unknown*
Wenyu Du, Binghamton University, USA: *An Accurate Method to Study the Shiryaev-Roberts Detection Procedure's Run-Length Distribution*
Yifan Xu, Binghamton University, USA: *First Crossing Times of Compound Poisson Processes with Two Linear Boundaries - Applications in SPRT and Queuing*

FriPM-InvPapSess 3: Methodologies for High-Dimensional Data Analysis - I
Organizer: T. N. Sriram, University of Georgia, USA
Chair: Makoto Aoshima, Institute of Mathematics, University of Tsukuba, Japan
Haileab Hilafu, University of Georgia, USA: *Sequential Sufficient Dimension Reduction for Large p Small n Problems*
Wenbo Wu, University of Georgia, USA: *Stable Estimation in Dimension Reduction by Sub-Sampling with Random Weights*
Wenhui Sheng, University of Georgia, USA: *Sufficient Dimension Reduction Via Distance Covariance*

Friday, July 19, 2013, 3:15pm - 4:45pm

FriPM-InvPapSess 4: Sequential Methods in the Hands of Young Researchers - II
Organizer: Nitis Mukhopadhyay, University of Connecticut-Storrs, USA
Chair: Debanjan Bhattacharjee, Utah Valley University, Orem, Utah
Aleksey Polunchenko, Binghamton University, USA: *A Bird's View on Computational Quickest Change-Point Detection*
Sankha Muthu Poruthotage, University of Connecticut-Storrs, USA: *Multiple Crossing Sequential Fixed-Size Confidence Region Methodologies for Normal Mean Vector*
Swarnali Banerjee, University of Connecticut-Storrs, USA: *Sequential Negative Binomial Problems with Applications in Statistical Ecology*

FriPM-InvPapSess 5: Sequential Change Point Detection
Organizer: Igor Nikiforov, Université de Technologie de Troyes, France
Chair: Edit Gombay, University of Alberta, Canada
Boris Darkhovsky, Russian Academy of Sciences & **Alexandra Piryatinska**, San Francisco State University, USA: *Quickest Detection Via -Complexity of Continuous Functions*
Yasin Yilmaz, Columbia University, USA & **George V. Moustakides**, University of Patras, Greece: *Sequential Joint Detection and Estimation*
Michael Baron, Univ. Texas at Dallas, USA: *Change-Point Detection in Multiple Channels*

Saturday, July 20, 2013, 8:30am - 9:45am

Plenary Lecture 3
Chair: Serguei Pergamenchtchikov, University of Rouen, France
Makoto Aoshima, Institute of Mathematics, University of Tsukuba, Japan: *Effective Methodologies for High-Dimensional Data*

Saturday, July 20, 2013, 10:15am - 11:45am

SatAM-InvPapSess 1: Applications of Sequential Analysis
Organizer: Steve Coad, Queen Mary, University of London, UK
Chair: Vladimir Dragalin, AptivSolutions, USA
Robert Keener, University of Michigan, USA: *The Modified Keifer-Weiss Problem, Revisited*
Anastasia Ivanova, University of North Carolina, USA: *Treatment Selection with the Sequential Parallel Comparison Design*
Chih-Chi Hu, Columbia University, USA: *On the Efficiency of Nonparametric Variance Estimation in Sequential Dose-Finding*

SatAM-InvPapSess 2: Change Detection in Functional Sequences - II
Organizer: Eric Chicken, Florida State University, USA
Chair: Shelemyahu Zacks, SUNY Binghamton, New York, USA
Eric Chicken, Florida State University, USA: *Change Points in Nonstationary Density Estimation*
Shing Chang, Kansas State University, USA: *Real-Time Detection of Wave Profile Changes*
Kamran Paynabar, Georgia Institute of Technology, USA: *Process Monitoring and Fault Diagnosis Using Multichannel Profiles*

SatAM-InvPapSess 3: Sensor Exploitation
Organizers: Mark Koch, Sandia National Laboratories, USA
Chair: Marlo Brown, Niagara University, USA
Annabel Prause, RWTH Aachen University, Germany: *Sequential Detection of Three Dimensional Signals under Dependent Noise*
Igor Nikiforov, Université de Technologie de Troyes, France: *Sequential detection of transient changes*
Qian Xie, Florida State University, USA: *Metric-Based Multiple Image Registration*

Saturday, July 20, 2013, 1:15pm - 2:45pm

SatPM-InvPapSess 1: Recent Advances in Sequential Change Detection
Organizer & **Chair:**: Georgios Fellouris, University of Southern California & University of Illinois at Urbana Champaign
George V. Moustakides, University of Patras, Greece: *Multiple Optimality Properties of the Shewhart Test*
Hongzhong Zhang, Columbia University, USA: *Robustness of the N-CUSUM Stopping Rule*
Grigory Sokolov, University of Southern California, USA: *Unstructured Sequential Change Detection in Sensor Networks*

SatPM-InvPapSess 2: Sequential Inference, Change-Point Detection and Clinical Trials

Organizer & **Chair:** Bhargab Chattopadhyay, University of Texas-Dallas, USA

Dong Xi, Northwestern University, Illinois, USA: *Allocating Recycled Significance Levels in Group Sequential Procedures for Multiple Endpoints*

Tian Zhao, University of Texas-Dallas, Texas, USA: *Multiple Testing in Group Sequential Clinical Trials*

Tiansong Wang, University of Texas-Dallas, Texas, USA: *Change-Point Detection with Multiple Sensors*

Sunday, July 21, 2013, 8:15am - 9:45am

SunAM-InvPapSess 1: Change Point Detection in Skew Distributions and Related Topics

Organizer: Wei Ning, Bowling Green State University, USA

Chair: George V. Moustakides, University of Patras, Greece

Wei Ning, Bowling Green State University, USA: *Information Approach for the Change Point Detection in the Skew Normal Distribution and Its Applications*

Abeer Hasan, Bowling Green State University, USA: *A Computational Based Methodology for the Change Point Problem Under the Non-central Skew t Distribution*

Haiyan Su, Montclair State University, USA: *Empirical Likelihood Inference for Two-sample Comparison with Censored Data*

SunAM-InvPapSess 2: Sequential Inference

Organizer & **Chair:** Venugopal V. Veeravalli, University of Illinois at Urbana-Champaign, USA

Georgios Fellouris, University of Southern California, USA: *Multichannel Sequential Hypothesis Testing*

Taposh Banerjee, University of Illinois Urbana-Champaign, USA: *Data-Efficient Quickest Change Detection*

Jun Geng & Lifeng Lai, Worcester Polytechnic Institute, USA: *Quickest Change Point Detection with Stochastic Energy Constraints*

Sunday, July 21, 2013, 10:15am - 12:00noon

Abraham Wald Prize Ceremony

Chair: Nitis Mukhopadhyay, University of Connecticut-Storrs, USA

10:15-10:45 am: Presentation of 2013 Abraham Wald Prize in Sequential Analysis

Plenary Lecture 4

Ansgar Steland, RWTH Aachen University, Germany: *Nonparametric Monitoring of Time Series*

Title and Abstracts

Effective Methodologies for High-Dimensional Data

Makoto Aoshima
Institute of Mathematics, University of Tsukuba, Japan

High-dimensional data situations occur in many areas of modern science. A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. We call such a data set HDLSS data. Aoshima and Yata (2011a,b) developed a variety of inference for HDLSS data such as given-bandwidth confidence region, two-sample test, testing the equality of covariance matrices, correlation test, classification, regression, and variable selection. The keys are non-Gaussian, HDLSS asymptotics, geometric representations, cross-data-matrix methodology and sample size determination to ensure prespecified accuracy for the inference. Hall et al. (2005) and Jung and Marron (2009) gave geometric representations of HDLSS data under a Gaussian-type assumption. As for non-Gaussian HDLSS data, Yata and Aoshima (2012) found a completely different geometric representation. Jung and Marron (2009) considered the conventional PCA for HDLSS data under a Gaussian-type assumption. As for non-Gaussian HDLSS data, Yata and Aoshima (2009) showed that the conventional PCA cannot give consistent estimates for eigenvalues, eigenvectors and PC scores. Yata and Aoshima (2010) created a new PCA called the *cross-data-matrix (CDM) methodology* that offers consistent estimates of those quantities for non-Gaussian HDLSS data. The cross-data-matrix methodology is also an effective tool to construct a statistic of inference for HDLSS data at a reasonable computational cost. Recently, Yata and Aoshima (2013) created the *extended cross-data-matrix (ECDM) methodology* that offers an optimal unbiased estimate in inference for HDLSS data and they applied the ECDM methodology to the test of correlations given by Aoshima and Yata (2011a). A review of research on high-dimensional data analysis is given by Aoshima and Yata (2013a). In this talk, I would like to give an overview of recent developments on high-dimensional data analysis. I will introduce effective methodologies to handle HDLSS data sets. In this talk, the Gaussian assumption is not assumed and also the equality of covariance matrices is not assumed. I would like to show the possibility of a variety of inference that can ensure prespecified accuracy by utilizing geometric characteristics of HDLSS data. I will introduce a new classification procedure called the *misclassification rate adjusted classifier*, developed by Aoshima and Yata (2013b), that can ensure accuracy in misclassification rates for multiclass classification. Finally, I will give some examples of discriminant analysis and cluster analysis for HDLSS data by using microarray data sets.

Sequential Robust Efficient Adaptive Estimation for Nonparametric Autoregressive Models

Ouerdia Arkoun
LMRS, University of Rouen, France
Collaborators/co-authors: Serge Pergamenchtchikov

We construct efficient robust truncated sequential estimators for the pointwise estimation problem in nonparametric autoregression models with smooth coefficients. For gaussian models we propose an adaptive procedure based on the conctructed sequential estimators. The minimax adaptive convergence rate is obtained. It turns out that in this case this adaptive rate is the same as for the regression gaussian models.

Sequential Negative Binomial Problems with Applications in Statistical Ecology

Swarnali Banerjee
Department of Statistics, University of Connecticut, USA
Collaborators/co-authors: Nitis Mukhopadhyay

Count data is abundant in entomology, more generally in statistical ecology. Anscombe (1949, Biometrics) emphasized Negative Binomial (NB) modeling for overdispersed count data. The undeniable status of a NB model is well-documented in agriculture, insect infestation, soil and weed science, and others. Young and Young (1998, Kluwer) shows the importance of a NB model under integrated pest management (IPM) and EPA. Most of the existing methodologies, both non-sequential and sequential, were reviewed by Mukhopadhyay and Banerjee (2012, under submission). Willson and Folks (1983) developed a fixed-size confidence interval methodology for μ, the mean of NB distribution for a known value of the known value of the thatch parameter k. Motivated by this methodology, we propose a novel method of sequential fixed-size confidence interval estimation. The procedure enjoys asymptotic consistency and first order asymptotic efficiency and its applications in ecology are widespread and useful. The new methodology was compared with that of Willson and Folks (1983) using theory, simulations, and a real dataset from insect infestation with very encouraging results. We extended this idea of estimation to other distributions with positive parameters like the Poisson distribution. Here too, we could prove similar theoretical results and simulations and real data analyses showed encouraging results.

Data-efficient Quickest Change Detection

Taposh Banerjee
Department of ECE, University of Illinois at Urbana-Champaign, USA
Collaborators/co-authors: Venugopal V. Veeravalli

In the classical problem of quickest change detection a decision maker observes a sequence of random variables over time. At some point of time, unknown to the decision maker, the distribution of the random variables changes. The decision maker has to detect this change in distribution as quickly as possible subject to a constraint on the false alarm rate. In many engineering applications where the quickest change detection algorithms may be employed, the change happens rarely and there is a cost associated with acquiring observations or acquiring data. Motivated by this, we study the classical quickest change detection problem with an additional constraint on the number of observations used before the change point, with the cost of observations taken after the change point already penalized through the delay. We propose Bayesian and minimax problem formulations for data-efficient quickest change detection and obtain solutions that are asymptotically optimal, as the false alarm rate goes to zero. We also discuss data-efficient quickest change detection in sensor networks.

Change-point Detection in Multiple Channels

Michael Baron
Department of Mathematical Sciences, University of Texas at Dallas, USA

Bayesian multichannel change-point detection problem is studied. The loss function penalizes for false alarms and detection delays, and the penalty increases with each missed change-point. For wide classes of stochastic processes, with or without nuisance parameters, asymptotically pointwise optimal (APO) rules are obtained, translating the classical concept of Bickel and Yahav to the sequential change-point detection. These APO rules are attractive because of their simple analytic form and straightforward computation.

Detection of Changes of Multiple Poisson Processes Monitored at Discrete Time Points Where the Arrival Rates Are Unknown

Marlo Brown
Department of Mathematics, Niagara University, USA

We observe a Poisson process in several categories where the arrival rates in each category change at some unknown integer. For some of these categories the arrival rates increase, while in other categories the arrival rates decrease. The point at which the process changes may be different for each category. We assume both the arrival rates for each category as well as the change-point are unknown. We develop procedures for detecting when a change has occurred in at least one of these categories. We provide some numerical results to illustrate the effectiveness of the detection procedures.

Recent Developments on Sequential Testing for Lévy Processes

Bruno Buonaguidi
Department of Decision Sciences, Bocconi University, Milan, Italy
Co-authors: Pietro Muliere

Let $X = (X_t)_{t \geq 0}$ be a Lévy process, that is, a stochastic process with independent and stationary increments. It is well known that its distributional properties are described by its Lévy-Khintchine triplet $g = \{\gamma, \sigma^2, v(\cdot)\}$, where $\gamma \in \mathbb{R}$, $\sigma^2 > 0$ and $v(\cdot)$ is the so called Lévy measure. Our presentation is focused on the following problem: the triplet of X, $g_\vartheta = \{\gamma_\vartheta, \sigma^2, v_\vartheta(\cdot)\}$, depends on the unobservable random variable ϑ, which takes at time $t = 0$ the values 1, with probability π, and 0, with probability $1 - \pi$, where $\pi \in [0, 1]$ is fixed and $g_i = \{\gamma_i, \sigma^2, v_i(\cdot)\}$, $i = 0, 1$, are known. Through the sequential observation of X, we want to test the two hypotheses $H_0 : \vartheta = 0$ and $H_1 : \vartheta = 1$, by means of a stopping rule which minimizes the sum of the expected cost of the observation process and the expected loss due to a final wrong choice. This problem can be reduced to an optimal stopping problem and, as such, it requires to find the value function and the boundaries, which separate the stopping and continuation region. This task is accomplished through the formulation of a suitable free-boundary problem, involving an integro-differential equation and some boundary conditions, whose presence is justified by the variational principles of the smooth and/or continuous fit. Exact and numerical solutions to this problem are obtained for an important class of Lévy processes.

A Two-Stage Selection and Testing Procedure for Comparing Several Treatments with a Control

Elena M. Buzaianu
Department of Mathematics and Statistics, University of North Florida, USA
Collaborators/co-authors: Pinyuen Chen

We propose a two-stage selection and testing procedure for finding the best among several experimental treatments, provided that it is better than a control. The treatments are assumed to follow normal distributions with unknown means and unknown variances. The comparison is relative to the population means and "better than the control" means that the mean of an experimental treatment is larger than that of the control. Our procedure is based on the selection and testing procedure proposed by Thall et al.(1988). At the first stage, ranking and selection techniques are used to detect the most promising experimental treatment. At the second stage, testing hypothesis is used to check whether the chosen population is better than the control. If none of the experimental populations seems to be promising, the procedure allows for an early termination of the experiment.

Real-Time Detection of Wave Profile Changes

Shing I Chang
Quality Engineering Laboratory, Department of Industrial and Manufacturing Systems Engineering, Kansas State University, USA
Collaborators/co-authors: Behnam Tavakkol and Tzong-Ru Tsai

A statistical process control (SPC) framework is proposed to detect potential changes of a wave profile on a real-time basis. In regular profile monitoring, change detection takes place when a complete profile is generated. In this study, the detection of a potential profile change takes place before the entire information on the profile of interest is fully available. The main research goal is to make a correct process decision as soon as possible. A real-world example of condensation-water-temperature profile monitoring was used to demonstrate the proposed framework. A simulation study was also conducted. The simulation results confirm that the proposed framework is capable of detecting profile changes without having to wait for the entire profile to be generated.

Two-Stage Fixed-Width Confidence Interval of Nonparamteric Regression Parameters Using Nadaraya-Watson Estimator

Bhargab Chattopadhyay
Department of Mathematical Sciences, University of Texas at Dallas, USA

Let us begin by assuming availability of a sequence (X_1, Y_1), (X_2, Y_2),... of independent observations, where, for each $i \geq 1$, X_i is the i^{th} order non-random predictor variable and Y_i is the response variable. to begin with let us assume that we have an equally spaced design on [0,1]. Thus we can define, a nonparamteric regression model as: $\mathbf{Y} = m(\mathbf{X}) + \sigma\varepsilon$where, $E(\varepsilon) = 0$ and

$V(\varepsilon) = I$. The letters in bold signifies a vector. Hence, $E(\mathbf{Y}|\mathbf{X}) = m(\mathbf{X})$ and $V(\mathbf{Y}|\mathbf{X}) = \sigma^2 I$. Our main aim is to construct a fixed-width confidence interval J for $m(x)$ at $x = x_0$ such that, $P(m(x_0) \in J) \geq 1-\alpha$.In principle, confidence intervals can be obtained from asymptotic normality for an estimator of $m(x)$, say, $\widehat{m}(x)$. But for constructing fixed-width confidence intervals, no fixed sample size methodology can be applied. Only a multi-stage or a sequential procedure will solve the problem. Using simulation, Dharmasena et al. (ANZIAM, 2008) proposed a two-stage procedure using a stopping rule for this problem and then compared the emperical coverage probabilities for procedures based on Nadaraja-Watson estimator and Local linear estimator of the regression parameter. In this presentation, a different two-stage procedure for constructing a fixed-width confidence interval for the nonparametric regression parameter using Nadaraya-Watson estimator will be talked about along with several properties satisfied by our method. A comparative study of the two methods will also be presented.

Change Points in Nonstationary Density Estimation

Eric Chicken
Department of Statistics, Florida State University, USA
Collaborators/co-authors: Rachel Becvarik

We consider a sequence of observations where each observed point is a sample of data from some unknown distribution. A statistic is proposed which monitors nonstationary densities based on these observations through the use of transformed wavelet coefficients of the quantiles. This method is completely nonparametric, designed for no particular distributional assumptions. The observed samples are treated functionally by using their quantile-quantile relations. Thresholding of these functional quantile relation of samples reduces the dimension of the problem, allowing for high power in detecting changes over time in the underlying distribution of the observed samples.

Statistical Inference of a Measure of Reduction for Two Binomial Variates

Hokwon Cho
Department of Mathematical Sciences, University of Nevada, Las Vegas, USA

We consider a measure of reduction for two sequences of independent binomial proportions. In particular, a measure of reduction (MOR) ρ is studied for two sample proportions with sequential sampling based on a modified MLE. First, we study the desirable properties of the estimator: the asymptotic unbiasedness and its the variance. Then, from the asymptotic normality of the measure ρ we can establish approximate confidence intervals for ρ. The Monte Carlo experiment is performed for the various scenarios of two sets of samples as well as to examine its finite sample behavior. Two examples are provided to illustrate the use of the measure, and to extend to the hypothesis testing for further statistical inference.

Representations of Excessive Functions and Their Application to Optimal Stopping Problems

Sören Christensen
Mathematical Insititute, Christian-Albrechts-University Kiel, Germany
Collaborators/co-authors: Paavo Salminen

Two approaches for solving sequential decision problems are presented. Both are based on representation results for excessive functions of Markov processes. In the first approach, we represent these functions as expected suprema up to an exponential time. This leads to generalizations of recent findings for Lévy processes obtained essentially via the Wiener-Hopf factorization to general strong Markov processes. In the second approach, the Riesz integral representation is utilized to solve sequential decision problems without the machinery of local time-space-calculus on manifolds. This approach is illustrated by treating the famous optimal investment problem in arbitrary dimension for geometric Brownian motions and certain classes of Lévy processes without making use of the smooth/continuous fit principle directly.

Estimation of Parameters of the Absolute Autoregressive Model

Steve Coad
School of Mathematical Sciences, Queen Mary, University of London, UK

The absolute autoregressive model is considered in which the errors follow a normal distribution. Interest lies in estimating the parameters of the model. Approximations are obtained for the bias of the maximum likelihood estimators and corrected confidence intervals are constructed. The derivations involve using the moments of the stationary distribution for the responses, which is an example of a skew-normal distribution. Simulation results are used to assess the accuracy of the approximations.

Sequential Multiple Testing Controlling Generalized Error Rates

Shyamal Krishna De
Department of Mathematical Sciences, Binghamton University, USA
Collaborators/co-authors: Michael Baron

A number of sequential experiments such as sequential clinical trials with multiple endpoints, multichannel change-point detection problems, etc., involve multiple hypothesis testing. For instance, in a clinical trial, patients are typically collected sequentially to answer multiple questions about safety and efficacy of a new treatment based on multiple endpoints. That is, a family of hypotheses is tested simultaneously, based on these endpoints. A separate decision, accept or reject, is expected on each individual hypothesis. The goal of this work is to develop optimal stopping rules and decision rules such that desired error rates are controlled at pre-specified levels and the expected sample size is as low as possible. In case of simultaneous testing of a large number of hypotheses (e.g., in microarray experiments), a few Type I and Type II errors can

often be tolerated. It is then desirable to control the generalized familywise error rates GFWER-I and GFWER-II defined as the probabilities of making at least $k(\geq 2)$ Type I errors and at least $m(\geq 2)$ Type II errors respectively. A number of tests are proposed in the literature on multiple testing that control GFWER-I when the sample size is fixed. In this work, we extend the ideas of Lehmann and Romano from fixed-sample multiple testing to sequential experiments. Our proposed procedures control both GFWER I and II under any combination of true null and alternative hypotheses whereas fixed-sample multiple tests cannot control both GFWER I and II simultaneously. Moreover, the proposed sequential schemes require less expected sample size compared to the Lehmann-Romano scheme. Also, a new curtailed approach is proposed that leads to a further reduction of the expected sample size.

Adaptive Clinical Trials with Population Enrichment Design

Vladimir Dragalin
Innovation Center, Aptiv Solutions, USA

There is a growing interest among regulators and sponsors in using precision medicine approaches that allow for targeted patients to receive maximum benefit from the correct dose of a specific drug. Population enrichment designs offer a specific adaptive trial methodology to study the effect of experimental treatments in various sub-populations of patients under investigation. Instead of limiting the enrollment only to the enriched population, these designs enable the data-driven selection of one or more pre-specified subpopulations at an interim analysis and the confirmatory proof of efficacy in the selected subset at the end of the trial. Sample size reassessment and other adaptive design changes can be performed as well. Strong control of the familywise Type I error rate is guaranteed by combining closure principle and p-value combination tests. In this presentation, the general methodology and designing issues when planning such a design will be described. Criteria for assessing the operating characteristics of these designs will be given, and the application will be illustrated by examples.

An Accurate Method to Study the Shiryaev–Roberts Detection Procedure's Run-Length Distribution

Wenyu Du
Department of Mathematical Sciences, State University of New York at Binghamton, USA
Collaborators/co-authors: Aleksey Polunchenko, Grigory Sokolov

Change-of-measure is a powerful technique ubiquitously used in Statistics and Probability. Particularly, it enabled the proof of a number of exact and asymptotic optimality results in the field of sequential change-point detection. We put the technique to a novel use: to develop a numerical method to study the in-control Run-Length distribution of the Shiryaev–Roberts (SR) detection procedure. The latter is a capable competitor of the popular Cumulative Sum (CUSUM) chart and Exponentially Weighted Moving Average (EWMA) scheme. Specifically, the numerical method is based on the integral-equation approach and the use of the change-of-measure ploy is to enable a higher rate of convergence (exact rate is provided). A tight bound on the method's error is supplied. The method is not restricted neither to a particular data distribution nor to a particular initial "head start" value of the SR detection statistic. Using a martingale property of the latter, it is also shown that the method's high accuracy is preserved even when the partition is rough. To conclude, we offer a case study to demonstrate the method at work. Specifically, assuming Gaussian observations, we employ the method to tabulate the SR's Run-Length's in-control average (i.e., the ARL), its standard deviation, and several quantiles. We also remark on extending the method's idea to other performance measures as well as to other control charts.

Multichannel Sequential Hypothesis Testing

Georgios Fellouris
Department of Mathematics, University of Southern California, USA
Collaborators/co-authors: Alexander Tartakovsky

In this talk, we will consider the problem of signal detection when observations are sequentially acquired at different sensors. We will examine different formulations depending on the presence of bandwidth and energy constraints, the sparsity of the signal and the dynamics of the observed processes. For each formulation, we will present efficient sequential tests with strong asymptotic optimality properties.

Quickest Change Point Detection with Stochastic Energy Constraints

Jun Geng
Department of Electrical and Computer Engineering,
Worcester Polytechnic Institute, USA
Collaborators/co-authors: Lifeng Lai

In this talk, we discuss the design and analysis of optimal detection scheme for sensors that are deployed to monitor the change in the environment and are powered by the energy harvested from the environment. In this type of applications, detection delay is of paramount importance. We model this problem as quickest change detection problem with a stochastic energy constraint. In particular, a wireless sensor powered by renewable energy takes observations from a random sequence, whose distribution will change at a certain unknown time. Such a change implies events of interest. The energy in the sensor is consumed by taking observations and is replenished randomly. The sensor cannot take observations if there is no energy left in the battery. Our goal is to design a power allocation scheme and a detection strategy to minimize the worst case detection delay, which is the difference between the time when an alarm is raised and the time when the change occurs. Two types of average run length (ARL) constraint, namely an algorithm level ARL constraint and a system level ARL constraint, are considered. We propose a low complexity scheme in which the energy allocation rule is to spend energy to take observations as long as the battery is not empty and the detection scheme is the Cumulative Sum test. We show that this scheme is optimal for the formulation with the algorithm level ARL constraint and is asymptotically optimal for the formulations with the system level ARL constraint.

Change Detection for Binary Time Series

Edit Gombay
Department of Mathematical and Statistical Sciences, University of Alberta, Canada
Collaborators/co-authors: Abdulkadir Hussein, Stefan Steiner

Binary time series is a model of great practical importance as it is often used in measuring health care performance, in financial markets, in industrial processes, and in climate studies. We survey recent theoretical developments concerning logistic and other regression models that allow $AR(p)$-type dependence structure and the presence of covariates. Conditions are set for the Maximum Partial Likelihood Estimator's existence and its convergence to the true values at the optimal rate. The performance of the score vector of the partial likelihood function is analysed and used for change detection. Its usefulness will be demonstrated with data on 30-day mortality rates after cardiac surgery.

A Computational Based Methodology for the Change Point Problem Under the Non-central Skew t Distribution

Abeer Hasan
Department of Statistics, Bowling Green State University, USA
Collaborators/co-authors: Wei Ning, Arjun Gupta

Skew Normal distributions have been very popular in research over the past three decades. The demand for skew distributions with tails longer than normal motivated scholars to investigate variations of the skew t distributions that can be suitable to fit long tailed data. There is a special interest in studying change point problems while modeling data from applications in the health and financial fields. In this talk a new skew t distribution family will be presented. Different settings of the change point problem for this distribution will be discussed and an information based approach to solve these problems will be discussed. Some applications will be presented to illustrate the use of the theoretical approach in practice.

Sequential Sufficient Dimension Reduction for Large p Small n Problems

Haileab Hilafu
Department of Statistics, University of Georgia, USA
Collaborators/co-authors: Xiangrong Yin

In this talk, we propose a new and simple framework for dimension reduction in the large p small n setting. The framework decomposes the data into pieces and undertakes reduction of the data in a sequential manner–enabling existing approaches for $n > p$ to be adapted to the $n < p$ problems. Estimating large covariance matrix, a very difficult task, is avoided. We propose two separate paths to implement the framework. The paths provide sufficient procedures for identifying informative variables. The efficacy of the proposed paths is demonstrated through extensive empirical studies and real data application.

On the Efficiency of Nonparametric Variance Estimation in Sequential Dose-Finding

Chih-Chi Hu
Department of Biostatistics, Columbia University, USA
Collaborators/co-authors: Ying Kuen K. Cheung

Typically, phase I trials are designed to determine the maximum tolerated dose, defined as the maximum test dose that causes a toxicity with a target probability. In this talk, we formulate dose finding as a quantile estimation problem and focus on situations where toxicity is defined by dichotomizing a continuous outcome, for which a correct specification of the variance function of the outcomes is important. This is especially true for sequential study where the variance assumption directly involves in the generation of the design points and hence sensitivity analysis may not be performed after the data are collected. In this light, there is a strong reason for avoiding parametric assumptions on the variance function, although this may incur efficiency loss. We investigate how much information one may retrieve by making additional parametric assumptions on the variance in the context of a sequential least squares recursion. By asymptotic comparison and simulation study, we demonstrate that assuming homoscedasticity achieves only a modest efficiency gain when compared to nonparametric variance estimation: when homoscedasticity in truth holds, the latter is at worst 88% as efficient as the former in the limiting case, and often achieves well over 90% efficiency for most practical situations.

Adaptive Latin Hypercube Designs for Computer Experiments

Ying Hung
Department of Statistics, Rutgers University, USA
Collaborators/co-authors: Huijuan Li

Most of the existing designs in computer experiments are chosen in advance, which results in insufficient information. To overcome this problem, a new class of adaptive designs is introduced. It is model-free and space-filling; therefore, it is attractive for complex computer experiments particularly in the early stage where model fitting tends to be unreliable. The construction procedure and design-unbiased estimators are discussed. Moreover, a refinement of the proposed design is introduced which provides better control over sample size and avoids replicates in the final sample. An application to a simulation study of the piezoelectric energy harvester shows that the proposed adaptive designs are more efficient than the existing non-adaptive designs in reducing the estimation variance.

Issues and Remedies in Surveillance of Binary Outcomes

A. Hussein
Department of Mathematics & Statistics, University of Windsor, Canada

Surveillance of Public health has recently gained momentum due the availability of real-time administrative data sets as well as increased threats of epidemics. In this talk, we will describe the use of risk-adjusted methods in the statistical surveillance of binary outcomes in public health. We point out issues in these methods and propose some remedies.

Treatment Selection with the Sequential Parallel Comparison Design

Anastasia Ivanova
Department of Biostatistics, UNC at Chapel Hill, USA

The sequential parallel comparison design (SPCD) is a two-stage design recommended for trials with possibly high placebo response. Drug-placebo comparison is followed by placebo non-responders being re-randomized between drug and placebo. Such two-stage strategy might be beneficial when combined with dose selection in a Phase 2 trial. We compare various methods of data analysis in such a trial and give recommendations on what method to use.

Sequential Analysis Methods of Bayesian Type for Testing Hypotheses

K. J. Kachiashvili
I. Vekua Institute of Applied Mathematics of the Tbilisi State University,
Georgian Technical University, Georgia

New methods of sequential analysis of Bayesian type of testing hypotheses are offered. The methods are developed on the basis of the specific statement of the Bayesian approach of testing hypotheses as constrained optimization problem instead of usual unconstrained problem. Unconstrained Bayesian statement of hypotheses testing minimizes the risk function which contains two types of errors: for incorrect rejection of hypotheses when they are correct and for incorrect acceptance of hypotheses when they are erroneous. We consider constrained statement of Bayesian approach which consists in the upper restriction of the probability of the error of one kind and the minimization of the probability of the error of the second kind. Such statement of the problem leads to the specificity of hypotheses acceptance regions. Particularly, in this case among of hypotheses acceptance sub-regions in the observation space we have the regions of the suspicion on the validity of several (more than one) tested hypotheses and the region of impossibility of acceptance of the tested hypotheses. Using these properties there are developed new sequential methods of testing hypotheses which give consistent, reliable and optimum results in the sense of the chosen criterion. The results of research of the properties of these methods are given. The examples of testing of hypotheses for the case of the sequential independent sample from the multidimensional normal law of probability distribution with correlated components are cited. On the basis of theses examples there are compared the Wald and the Berger and offered new sequential methods among them. The positive and negative sides of these approaches are considered on the basis of computed examples.

The Modified Keifer-Weiss Problem, Revisited

Robert W. Keener
Department of Statistics, University of Michigan, USA

The expected sample size τ for sequential probability ratio test of θ_1 versus θ_2 may be unacceptably large if the true parameter θ lies between these values. To ameliorate this concern, Keifer and Weiss suggest designing a test to minimize $\sup_\theta E_\theta \tau$ with error probabilities controlled at θ_1 and

θ_2. In symmetric cases, the modified problem of minimizing $E_{\theta_0}\tau$ with the same bounds for error probabilities provides a solution. Lorden and others show that a combination of two one-sided sequential probability ratio tests, called a 2-SPRT, provides a near-optimal solution to the modified Keifer-Weiss problem as the error rates for the test tend to zero. In this talk, we will consider effective strategies for the modified problem as the difference $\theta_2 - \theta_1$ tends to zero with the error rates fixed, a limit that leads to Brownian motion approximations.

Fisherian Evidential Approach to Sequential Clinical Trials

Qing Liu
Janssen Research and Development, L.L.C. of Johnson & Johnson, USA

D. R. Cox (2004) states that "*Indeed, I believe that many statisticians approaching statistics from a broadly frequentist perspective are uneasy at notions such as 'spending error rates', perhaps because these treat notions of error rates as more than just hypothetical concepts used for calibrating measures of uncertainty against performance in idealized situations. While in some situations there may be compelling quasi-political arguments, as well as cost considerations, pointing against too frequent an analysis, in principle it is hard to see an argument at a completely fundamental level.*" For clinical trials in life-threatening conditions, a principle of fundamental importance is to close monitor the trial so that accumulating data can be used to optimize treatment or to stop the trial upon overwhelming statistical evidence in favor of a new treatment. We propose a Fisherian evidential approach by which the statistical evidence is evaluated with the occurrence of each clinical event. Upon observing overwhelming statistical evidence, clinical decisions could be made to stop the trial or to switch all patients to the new treatment. The evidential approach relies on the O'Brien-Fleming or the more general Wang-Tsiatis boundary which is a function of the information or number of events. This permits multiple endpoints for which the maximum information levels cannot be specified for all endpoints at the final analysis. We show how the Fisherian evidential approach controls the multiple type 1 error rates. More importantly, we show that the Fisherian evidential approach does not increase the cost.

Performance Analysis of Sequential Probability Ratio Test

Yu Liu
Department of Electrical Engineering, University of New Orleans, USA
Collaborators/co-authors: X. Rong Li

The sequential probability ratio test (SPRT) is a fundamental tool for sequential analysis. It forms the basis of numerous sequential techniques for different applications, for example, the truncated SPRT and Page's cumulative sum test (CUSUM). The performance of SPRT is characterized by two important functions—operating characteristic (OC) and average sample number (ASN), while CUSUM's performance is revealed by the average run length (ARL) function. These functions have been studied extensively under the assumption of independent and identically distributed log-likelihood ratios (LLRs) with constant bounds, which is too stringent for many applications. In this paper, inductive integral equations governing these functions are developed under very general settings—the bounds can be time-varying and the LLRs are assumed independent but non-stationary. These inductive equations provide a theoretical foundation for performance analysis. Unfortunately, they have non-unique solutions in the general case except for the truncated SPRT. Numerical solutions for some frequently encountered special cases are developed and are compared with the results of Monte Carlo simulations.

Layers of Experiments with Adaptive Combined Design

Richard W. Lu

School of Industrial and Systems Engineering, Georgia Institute of Technology, USA

Collaborators/co-authors: Lu, J.-C., Grover, M. A., Kim, S., Casciato, M., Hess, D. W., Kim, H., Wang, X., Wang, F.-F., Woo, H

This presentation proposes a methodology for batch sequential design of experiments (DOEs) with focused region changing adaptively for exploring optima in nano-fabrication processes. Due to limitations in experimental budget and equipment availability, demanding engineering-tolerance-requirement in process performance, and complicated nonlinear process response surface, the traditional DOEs such as fractional factorial designs, optimal designs and space-filling designs and also the commonly used optimum-searching strategy such as the response surface method (RSM) are not suitable in many nano-fabrication studies. This presentation first proposes a Layer-of-Experiment (LoE) concept to place experimental resources into local regions with high probabilities of covering process optima. Then, depending on how much confidence on the model assumed in the optimal design at various design stages (called layers), our Adaptive Combined Design (ACD) adaptively chooses a balanced design between the space-filling design for exploring process model uncertainties and the optimal design for best-estimating model parameters. The proposed LoE/ACD methodology is applied to optimize an elevated pressure/temperature carbon dioxide (epet-CO_2) assisted deposition process for silver nanoparticles (AgNPs) successfully.

Multiple Changepoint Detection

Robert Lund

Department of Mathematical Sciences, Clemson University, USA

This talk presents a method to estimate the number of changepoint times and their locations in time-ordered data sequences. A penalized likelihood objective function is developed from minimum description length information theory principles. Optimizing the objective function yields estimates of the changepoint numbers and location time(s). Our model penalty is based on the types of model parameters and where the changepoint(s) lie, but not the total number of model parameters (such as classical AIC and BIC penalties). Specifically, changepoints that occur relatively closely are penalized more heavily. Our methods allow for autocorrelation in the observations and general mean shifts at each changepoint time. A genetic algorithm, which is an intelligent random walk search, is developed to rapidly optimize the penalized likelihood. Much of the discussion pertains to why a sequential/dynamic algorithm is difficult to develop, but desirable. Several applications to climatic time series are given.

Optimal Designs for Two-level Factorial Experiments with Binary Response

Abhyuday Mandal
Department of Statistics, University of Georgia, USA
Collaborators/co-authors: Jie Yang, Dibyen Majumdar

We consider the problem of obtaining locally D-optimal designs for factorial experiments with binary response and k qualitative factors each at two levels. We obtain a characterization for a design to be locally D-optimal. We develop efficient numerical techniques to search for locally D-optimal designs, investigate the properties of fractional factorial designs and study the robustness, with respect to unknown parameters, of locally D-optimal designs. Using prior distribution on the parameters, we investigate EW D-optimal designs, that are designs which maximize the determinant of the expected information matrix. These designs are much easier to find and still highly efficient compared to Bayesian D-optimal designs, as well as quite robust.

Conditional Probability of Correct Selection after Procedure Termination

Joshua McDonald
School of Industrial and Systems Engineering, Georgia Institute of Technology, USA
Collaborators/co-authors: Anton Kleywegt, David Goldsman, Craig Tovey, Eric Tollefson

We consider the problem of estimating the probability of correctly selecting the best system after a selection procedure concludes sampling. In particular, we provide exact conditional probabilities for a number of multinomial selection procedures, and both exact and simulation-based estimated conditional probabilities for certain normal selection procedures. We find that the conditional probabilities are often less than the nominal probability of correct selection; and we suggest ways to deal with this issue.

Multiple Optimality Properties of the Shewhart Test

George V. Moustakides
Department of Electrical and Computer Engineering, University of Patras, Greece

By adopting the probability maximizing approach as alternative to the minimization of the average detection delay used in most sequential change detection formulations, we propose modified versions of the Shiryaev, Lorden and Pollak performance measures. We demonstrate that for all these alternative formulations, the optimum sequential change detection scheme turns out to be the simple Shewhart rule. Interestingly we can solve, in a straightforward manner, problems which under the classical formulations have been open for many years. This includes optimum change detection with time varying observations or with multiple post-change probability measures.

On Determination of an Appropriate Pilot Sample Size

Nitis Mukhopadhyay
Department of Statistics, University of Connecticut, Storrs, USA

Any kind of multistage sampling strategy requires a practitioner to begin with an initial dataset with a pilot size m, an appropriately chosen number. Under purely sequential sampling, a choice of m may not be very difficult to come up with as long as m is reasonably small. It is so because here one moves forward step-by-step adaptively. Under multi-stage sampling, on the other hand, especially under two-stage sampling, the choice of an appropriate pilot size m happens to be extremely crucial. In such situations, understandably m should not be too large or too small. Our common wisdom may dictate that, but what choice of m qualifies to be labeled not "too large" or "too small", especially when the "optimal fixed sample size" remains unknown? We shall explore some concrete ideas based on, for example, (i) large-sample approximations, (ii) Fisher information, and (iii) available practical information. Illustrations will follow.

Sequential Detection of Transient Changes

Igor Nikiforov
Université de Technologie de Troyes, France
Collaborators/co-authors: Blaise Kévin Guépié, Lionel Fillatre

In many applications, like radar and sonar, industrial monitoring, safety critical surveillance, it is desirable to reliably detect situations where the sequentially observed data $\{X_t\}_{t\geq 1}$ contains, in addition to noise, a suddenly arrived at time t_0 signal of short duration L. Moreover, in safety critical applications, an acceptable detection delay is limited by a prescribed value L even if this signal is of long duration. In contrast to the traditional abrupt change detection, where the post-change period is assumed to be infinitely long, the detection at time T of such a short signal (transient change) has to be done before it disappears. Latent detection, i.e. detection which occurs after signal disappearance or with a delay greater than the prescribed value L (i.e. the event $\{T - t_0 + 1 > L\}$), is considered as a missed detection in such a case. The problem of sequential transient change detection is considered. A minimax type criterion is used : its goal is to minimize the "worst case" probability of missed detection $\inf_{T \in C_\alpha} \left\{ \mathbb{P}_{\mathtt{md}} = \sup_{t_0 \geq L} \mathbb{P}_{t_0}(T - t_0 + 1 > L \mid T \geq t_0) \right\}$ subject to the constraint on the "worst case" probability of false alarm $C_\alpha = \left\{ T \; : \mathbb{P}_{\mathtt{fa}} = \sup_{\ell \geq L} \mathbb{P}_0(\ell \leq T < \ell + m_\alpha) \leq \alpha \right\}$. A suboptimal algorithm is proposed and studied. The theoretical findings are applied to an important practical problem : drinking water quality monitoring.

Information Approach for the Change Point Detection in the Skew Normal Distribution and Its Applications

Wei Ning
Department of Mathematics and Statistics, Bowling Green State University, USA
Collaborators/co-authors: Grace Ngunkeng

The skew normal distribution is an extension of the normal distribution allowing the presence of skewness. Such a distribution can be used to model the skewed data in many applied fields such as finance, economics and medical research. The skew distribution family has been studied extensively since it was introduced by Azzalini (1985) for the first time. In this talk, we will propose an information approach method based on the Schwarz information criterion (SIC) to detect the change points in the parameters. Simulations are conducted to illustrate the performance of such a test procedure. Power comparisons between the skew normal distribution and the normal distribution under the different scenarios with the existence of the skewness are also showed through the simulations to indicate the advantage of using this flexible distribution family. Such a method is applied to some Latin American emerging market data sets to illustrate the testing procedure.

Bayesian Sequential Estimation of a Drift of Fractional Brownian Motion

Alex Novikov

University of Technology, Sydney, Australia, and Steklov Mathematical Institute, Moscow, Russia

Collaborators/co-authors: A. N. Shiryaev

We solve explicitly a Bayesian sequential estimation problem for the drift parameter μ of a fractional Brownian motion under the assumptions that a prior density of μ is Gaussian and that a penalty function is quadratic or Dirac-delta. The optimal stopping time for this case is deterministic.

Process Monitoring and Fault Diagnosis Using Multichannel Profiles

Kamran Paynabar

School of Industrial and Systems Engineering, Georgia Institute of Technology, USA

Collaborators/co-authors: Changliang Zou, Peihua Qiu

Analysis of profile data for the purpose of process monitoring and fault diagnosis has been extensively studied in the statistical process control literature. Most of existing methods deal with cases, in which the production process is characterized by single profiles. In many industrial practices, however, the sensing system is so designed that it records more than one profile during each operation cycle, which are known as multichannel or multivariate profiles. In this research, we develop a new phase-I process monitoring and diagnosis method for multichannel profile data. In the proposed method, first, we use multiple functional principal components analysis to reduce the dimensions of multichannel profile data, while considering the cross-correlation among profile channels. Next, using the extracted PC-scores, a multivariate change-point model is developed to detect possible changes in the historical data. Furthermore, a diagnosis method is proposed to determine which profile channels have changed. The performance of the proposed method is evaluated and compared with other methods by using simulations and a case study in a multi-operation forging process.

Minimax Sequential Kernel Estimators for Nonparametric Diffusion Processes

S. Pergamenchtchikov

Laboratory of Mathematics, LMRS, University of Rouen, France

Collaborators/co-authors: Galtchouk, L., University of Strasbourg, France

A sequential procedure for estimating the drift function of a diffusion process is constructed. The asymptotic properties such as the asymptotic minimax of the procedure and the asymptotic normality of observation time are established. In this talk we show that the sequential kernel estimators may have the advantages as compared with nonsequential estimators in the case of nonasymptotic risks. The example is given which illustrates this situation. It is shown that nonsequential estimators have the infinite nonasymptotic risk whereas the risk is finite under the sequential estimation schemes.

Quickest Detection Via ϵ-complexity of Continuous Functions

Alexandra Piryatinska
Department of Statistics, San Francisco State University, USA
Collaborators/co-authors: Boris Darkhovsky

Let $X = \{x(t)\}$ be an observed sequence. Suppose that the sequence generation mechanism is changed at an unknown moment t^* (a change-point). No prior information about the generation mechanisms before, and after t^* is given. These mechanisms can be stochastic, deterministic, or mixed. Our goal is to estimate t^* as soon as possible following its occurrence, based on the on-line sequence of observations. The detection method is based on a new concept, *the ϵ-complexity of continuous functions*, introduced by the authors earlier. The ϵ-complexity is an "intrinsic" characteristic of the function, and does not depend on the function generation mechanism. One of our findings is that, for the class of *Hölder* functions defined on an interval, ϵ-complexity can be characterized by a pair of real numbers . We call these numbers *complexity coefficients*. Sequential calculation of the complexity coefficients produces diagnostic sequences which are then utilized in the change-point detection procedure which relies on the assumption that a change in the generation mechanism leads to a change in the mean of the diagnostic sequence. Finally, detection of a change in the mean can be accomplished by the two-sided non-parametric CUSUM procedure. The simulation results are also provided.

On Reaching Nirvana (a.k.a. Steady State)

Moshe Pollak
Department of Statistics, The Hebrew University of Jerusalem, Israel

In many contexts one observes a stochastic process with the goal of learning steady-state characteristics. This talk addresses the question of how to declare with confidence that steady-state has been reached. The focus of the talk is in defining a reasonable stopping time and proposing relevant operating characteristics.

A Bird's View on Computational Quickest Change-Point Detection

Aleksey S. Polunchenko
Department of Mathematical Sciences
State University of New York (SUNY) at Binghamton, USA
Collaborators/co-authors: Wenyu Du, Sven Knoth, Grigory Sokolov

Quickest change-point detection, while considered a branch of statistics, can hardly be made independent from numerical analysis or allied areas studying computational methods. These methods enable accurate evaluation of sequential detection procedures' performance, a major problem essential for application of change-point detection in practice. Notwithstanding its importance, the computational side of change-point detection has been largely neglected. The literature on the subject is sporadic, scattered across many unrelated journals, and is long due for a critical assessment. Important questions such as convergence rates, error bounds and computational complexity are often poorly discussed or simply left out. The aim of this work is to fill in these gaps by providing an overview of the state-of-the-art in computational quickest change-point detection.

The overview is a critical synthesis of the main approaches to the problem of evaluating the performance of a detection procedure: Markov chain and integral (renewal) equations. As the former is known to be a special case of the latter, we focus on the integral equations approach and build a bridge between change-point detection and the fairly-well developed theory of numerical methods to solve integral equations. This section of numerical analysis is pillared by four main approaches: Nyström, collocation, Galerkin, and degenerate kernels. We survey these methods as well as a relatively recent direction that exploits Markov Chain Monte Carlo (MCMC) techniques. For all these methods, we discuss the questions of accuracy, computational complexity, and ease of implementability. To conclude, we give practical recommendations as to the pros and cons of each approach.

Multiple Crossing Sequential Fixed-Size Confidence Region Methodologies for Normal Mean Vector

Sankha Muthu Poruthotage
Department of Statistics, University of Connecticut, USA
Collaborators/co-authors: Nitis Mukhopadhyay

The purely sequential procedure proposed by Mukhopadhyay and Al-Mousawi (1986) can be used to construct a confidence region for the mean vector of $N_p(\mu, \sigma^2 H)$ with known $H_{p\times p}$ but unknown σ^2. The main advantage of this procedure is that the size of the confidence region and the confidence coefficient $(= 1 - \alpha)$ could be fixed in advance. Even though this procedure has efficiency, and asymptotic consistency properties it does not have the consistency property. In this article, a purely sequential sampling procedure is proposed which allows for continuous sampling even after the sample size exceeds the corresponding boundary $(= \frac{aS^2}{d^2})$ multiple times. Hence the stopping rules of this kind are called multiple crossing stopping rules. Some analytical properties such as efficiency, and asymptotic consistency of multiple crossing stopping rules are discussed. Its ability to achieve the required coverage probability without significant over sampling is illustrated via a large scale simulation exercise, across a wide range of values of d and α. A truncation technique as well as a fine-tuned adjustment to the multiple crossing stopping rule is proposed which are intended to increase the usability of the proposed procedure. Then we demonstrate how this methodology can be effectively applied for constructing fixed-sized confidence regions for treatment effects of ANOVA models. Finally some real data illustrations are provided in order to demonstrate the actual implementation and possible applications of the proposed procedure.

Sequential Detection of Three Dimensional Signals Under Dependent Noise

Annabel Prause
Institute of Statistics, RWTH Aachen University, Germany
Collaborators/co-authors: Ansgar Steland

We study detection methods for multivariable signals under dependent noise. We mainly focus our attention on three dimensional signals, i.e. on signals in the space-time domain. Examples for such signals are multifaceted. They include geographic and climatic data as well as image data, that are observed over a fixed time horizon. We are interested in the reconstruction of these signals and, at the same time, in detecting changes from a given reference signal. Thus, we assume that we observe a finite block of noisy samples $\{y_i = y_{i_1,i_2,i_3} : i_1, i_2, i_3 \in \{1, \ldots, n\}\}$ that is obtained from the model $y_{i_1,i_2,i_3} = f(i_1\tau_1, i_2\tau_2, i_3\tau_3) + \epsilon_{i_1,i_2,i_3}$. Here, f is the unknown signal depending on time (i_1) and location (i_2 and i_2), $\{\epsilon_i = \epsilon_{i_1,i_2,i_3}\}$ is a zero mean noise random field and τ_j, $j = 1, 2, 3$, are the sampling periods. We assume that they fulfill $\tau_j \to 0$ and $n\tau_j \to \overline{\tau}_j$, $j = 1, 2, 3$ as $n \to \infty$. Our detector statistic is based on a sequential partial sum process, related to classical signal decomposition and reconstruction approaches applied to the sampled signal. In order to determine the critical value / control limit of our detector statistic, it is helpful to know its limiting distribution, preferably under weak assumptions on the dependence structure, thus allowing a wide applicability of the results. Hence we need to establish the weak convergence of the sequential partial sum process in an appropriately chosen function space. However, we are not only interested in the asymptotic behaviour of our statistic under the null hypothesis, where the signal corresponds to our reference signal, but also in its asymptotic behaviour under the alternative, which may include deterministic as well as random changes. Thus, we also establish the limiting distribution under these local alternatives.

Some Recent Research On Nonparametric Profile Monitoring

Peihua Qiu
Department of Biostatistics, University of Florida, USA
Collaborators/co-authors: Kamran Paynabar, Changliang Zou

Quality of a process is often characterized by the functional relationship between response variables and predictors. Profile monitoring is for checking the stability of this relationship over time. In the literature, most existing control charts are for monitoring parametric profiles, and they assume that within-profile observations are independent of each other. In reality, these model assumptions are often invalid. In this talk, I will present some of our recent research on nonparametric profile monitoring. By our proposed approaches, profiles could be nonparametric, within-profile observations could be correlated, and the response variable could be multivariate (i.e., the multichannel profile monitoring problem).

Multi-Sensor Change Detection with Change Propagation

Vasanthan Raghavan
Department of Mathematics, University of Southern California, USA

In multi-sensor applications such as critical infrastructure monitoring and source of epidemic detection, not all the sensors observe the change at the same instant when the change actually happens. Motivated by such scenarios, the focus of this talk is on changepoint detection with random change propagation. In prior work, we had considered the setting where the first sensor that observes the change is known. We relax this condition here and consider a scenario that is applicable for more realistic two- or three-dimensional geometries. We propose a low-complexity detection procedure that results in substantial improvement in performance over a naive strategy that wrongly assumes that the change is seen instantaneously across all the sensors.

Sequential Fixed-Width Confidence Interval Based on Bhattacharyya-Hellinger Distance: The Non-Parametric Case

S. Yaser Samadi
Department of Statistics, University of Georgia, USA
Collaborators/co-authors: T.N. Sriram

We consider the construction of sequential fixed-width confidence interval for the minimum Bhattacharyya Hellinger (B-H) functional, $T(g)$, when the random sample comes from an unknown probability mass function g. Defining the minimum B-H distance estimator of $T(g)$ as $T(\hat{f}_n)$, we show that our sequential procedure is asymptotically consistent and efficient, as the width of the interval approaches zero. Monte Carlo simulations show that our sequential procedure based on minimum B-H distance estimator performs as good as that based on the MLE, when g is Poisson(θ). However, when the samples come from a contaminated Poisson distribution, our sequential procedure based on minimum B-H distance estimator continues to perform well, while the sequential procedure based on the MLE performs poorly. We will indicate extensions of this to the continuous case.

Sufficient Dimension Reduction Via Distance Covariance

Wenhui Sheng
Department of Statistics, University of Georgia, USA
Collaborators/co-authors: Xiangrong Yin

We introduce a novel approach to sufficient dimension reduction problems through using distance covariance. Our method requires very mild conditions on X. It estimates the central subspace efficiently with continuous predictors, mixed continuous and discrete predictors, or discrete predictors alone. Our method keeps the model-free advantage and it works efficiently for all types of link functions, including linear or nonlinear link functions. Under regularity conditions, we prove that our estimate is root n consistent and asymptotically normal. We compare the performance of our method with some other dimension reduction methods by simulations and find that our method is very competitive and robust across a number of models. We also analyze a real data set concerning the identification of different sounds to demonstrate the efficacy of our method.

Unstructured Sequential Change Detection in Sensor Networks

Grigory Sokolov
Department of Mathematics, University of Southern California, USA
Collaborators/co-authors: Georgios Fellouris, Alexander Tartakovsky

The problem of sequential change-point detection in multi-sensor networks has recently gained increased attention. In this setup, an area is being monitored by a number of distributed sensors, which transmit their data to a fusion center under bandwidth and energy constrains. An abrupt change occurs and is captured by a subset of sensors. The goal is to detect this change at the fusion center as soon as as possible, based on the sequentially acquired information from the sensors. We will consider this problem in the case that the size of the set of affected sensors is not known in

advance. We will propose a decentralized detection rule that is asymptotically optimal under all possible post-change scenarios. Moreover, using extensive simulation experiments, we will show that the proposed scheme is as efficient as alternative detection rules that do not take into account bandwidth constraints.

A Note on Partitioning Exponential Populations

Tumulesh K. S. Solanky
Department of Mathematics, University of New Orleans, USA

We consider the problem of partitioning a set of K exponential populations with respect to a control population. For this problem some multistage methodologies are proposed and their theoretical properties are derived. Using the Monte Carlo simulation techniques, the small and moderate sample size performance of the proposed procedure are studied.

Nonparametric Monitoring of Time Series

Ansgar Steland
Institute of Statistics, RWTH Aachen University, Germany

Present day data observable as a data stream is typically affected by serial correlations, i.e., we are given time series, the dimensionality ranging from univariate to multivariate and high-dimensional. As a matter of fact, the construction of optimal procedures requires knowledge of the finite-dimensional distributions which are usually unknown to us. Thus, there is a need of nonparametric approaches which provide good performance over sufficiently large classes of distributions. The lecture aims at providing an overview of recent approaches for the nonparametric detection of changes in the structure of a time series. We will discuss methods and related asymptotic results for general classic problems as well as more specific problems arising in certain fields of applications such as financial statistics, signal processing and image processing. Theoretical results, computational issues as well as practical applications will be discussed. In order to investigate the statistical properties of such procedures, one has to rely on large sample theory for dependent time series. A powerful technique is to represent a detector as a functional of the sequential partial sum process. Then we can apply invariance principles in order to derive approximations of the stopping times for large (nonparametric) classes of time series. This approach also provides the interesting general result that the distribution of the maximally selected CUSUM statistic can be obtained as a simple functional of the Brownian bridge process, as long as the time series satisfies an invariance principle in the sense of Donsker. For more involved detectors, that approach frequently provides explicit limit processes which give rise to concrete simulation procedures. We shall illustrate this by a new procedure aiming at the detection of signals arising in (communication) engineering, which is related to the Shannon-Whittaker sampling theorem. Further, one may rely on general consistency results of resampling procedures such as sequential subsampling, a method that works under very weak assumptions. We shall illustrate its use by applying it to a purely sequential method based on estimated characteristic functions, where the limit processes are rather untractable. A calibrated subsampling procedure provides accurate and quite robust approximations that allowed us to apply it to the monitoring of SO_2 data. Lastly, we shall discuss some recent results on detection procedures for highdimensional functional data such as measurment curves or images.

Empirical Likelihood Inference for Two-Sample Comparison with Censored Data

Haiyan Su
Department of Mathematical Sciences, Montclair State University, USA
Collaborators/co-authors: Mai Zhou, Hua Liang

Two-sample comparison problems are often encountered in practical projects and have been widely studied in the literature. Owing to practical demands, the research for this topic under special settings such as a semi-parametric framework have also attracted great attentions. In this study, we develop a new empirical likelihood-based inference under more general framework by using the hazard formulation of censored data for two sample semi-parametric hybrid models. We demonstrate that our empirical likelihood statistic converges to a standard chi-squared distribution under the null hypothesis. We further illustrate the use of the proposed test by testing the ROC curve with censored data, among others. Numerical performance of the proposed method is also examined.

Novel Response-Adaptive Designs for Clinical Trials with Time-to-Event Outcomes

Alex Sverdlov
IIS Translational Sciences, Novartis Pharmaceuticals Corporation, U.S.A.
Collaborators/co-authors: Yevgen Ryeznik, Weng Kee Wong

Many clinical trials use time-to-event (TTE) outcomes as primary measures of safety or efficacy. Efficient designs for such trials are needed, but finding such designs in practice may be challenging due to uncertainty about the model for event times, patient staggered entry, delayed responses and censored observations. In this presentation we will describe three types of our recently developed response-adaptive designs for clinical trials with TTE outcomes. The first type is response-adaptive randomization (RAR) designs for multi-armed comparative survival trials. In such trials the goal is to allocate more patients to treatments exhibiting longer survival times while maintaining important statistical properties, such as estimation efficiency, power and type I error, and the randomized nature of treatment assignments. We present several optimal RAR designs developed under assumptions of censored exponential and Weibull survival outcomes. These designs outperform traditional balanced randomization designs both in terms of statistical efficiency and ethical criteria. A user-friendly software `RARtool` to facilitate implementation of the new designs in practice will be presented. The second type is covariate-adjusted response-adaptive (CARA) randomization designs, which are applicable in survival trials with important patient covariates correlated with survival outcomes. A conventional approach for such trials is covariate-adaptive randomization to balance the distribution of covariates across treatment arms. Our proposed CARA designs generate treatment assignments according to estimated covariate-specific treatment differences, to reduce the number of events in the study without sacrificing important statistical properties of the design. We illustrate an application of the proposed methodology by redesigning two recent survival trials from the literature. The third type is dose finding trials with TTE outcomes where the dose is measured on a continuous scale. We assume a quadratic dose-response relationship for log-transformed Weibull event times that are subject to right censoring by a fixed or random censoring time. We investigate D-optimal design for the most precise estimation of the dose-response curve and c-optimal design for the most precise estimation of the dose at which the downturn occurs. Multi-stage adaptive designs to approximate optimal designs are constructed and robustness of the designs to model misspecification is assessed.

Sequential Hypothesis Tests: Historical Overview and Recent Results

Alexander Tartakovsky
Department of Mathematics, University of Southern California, Los Angeles, California, USA

Sequential hypothesis testing theory has a long history beginning with a famous *Sequential Probability Ratio Test* (SPRT) invented by Wald in the mid 1940s. We will provide a historical overview of sequential tests of simple and composite hypotheses as well as focus on recent advances and challenges, considering both Bayesian and frequentist problems. We will begin with conventional i.i.d. models and then tackle general non-i.i.d. stochastic models. Multiple decision problems are much more difficult than two-decision problems. In the case of finitely many simple hypotheses, we will show that matrix combinations of SPRTs are asymptotically optimal in minimizing the expected sample sizes for all hypotheses or more generally positive moments of the stopping time distribution when probabilities of errors are small. These results will be extended to multiple composite hypotheses and adaptive as well as mixture SPRT-like combinations both for i.i.d. and non-i.i.d. models. The general results will be illustrated by several examples that include Markov, hidden Markov and more general models. Certain applied problems will be also discussed.

Unsupervised Anomaly Detection for High-dimensional Data

Thayasivam Umashanger
Department of Mathematics, Rowan University, USA

Application and development of specialized statistical machine learning techniques is gaining increasing attention in the anomalies/intrusion detection on high-dimensional data. Most current approaches to detect anomalies in network traffic adopt supervised or unsupervised machine learning techniques, such as information theory neural networks support vector machines genetic algorithms, and many more. Typically, detection of anomalies in network traffic face many issues, namely, generation of numerous false positives due to slight deviation in statistical behavior between training and test data sets, unavailability of anomaly-free data for training, and attackers avoiding detection by training a system over a period of time to accept attack activity as benign outliers. In addition, supervised machine learning based anomaly detection requires availability of labeled training data which is an expensive and impractical operation. In this talk, an unsupervised statistical technique for detecting anomalies in network traffic based on multivariate Gaussian Mixture Model(GMM) using the integrated square error or L2 estimation (L_2E-GMM) is proposed. The method does not require labeled training data to detect anomalies. Furthermore, unlike the standard expectation maximization (EM) algorithm, L2E does not require careful initialization and is efficient in achieving computational speed. The competitive performance of this method is illustrated with simulated and real data. The L_2E technique is compared with supervised learning, Support Vector Machines (SVM), with respect to the detection accuracy.

Controlled Sensing for Multihypothesis Testing

Venugopal V. Veeravalli
ECE Department and Coordinated Science Lab,
University of Illinois at Urbana-Champaign, USA
Collaborators/co-authors: Sirin Nitinawarat, George Atia

The problem of multiple hypothesis testing with observation control is considered in both fixed sample size and sequential settings. In the fixed sample size setting, for binary hypothesis testing, it is shown that the optimal exponent for the maximal error probability corresponds to the maximum Chernoff information over the choice of controls. It is also shown that a pure stationary open-loop control policy is asymptotically optimal within the larger class of all causal control policies. For multihypothesis testing in the fixed sample size setting, lower and upper bounds on the optimal error exponent are derived. It is also shown through an example with three hypotheses that the optimal causal control policy can be strictly better than the optimal open-loop control policy. In the sequential setting, a test based on earlier work by Chernoff for binary hypothesis testing, is shown to be first-order asymptotically optimal for multihypothesis testing in a strong sense, using the notion of decision making risk in place of the overall probability of error. Another test is also designed to meet hard risk constrains while retaining asymptotic optimality. The role of past information and randomization in designing optimal control policies is discussed.

Efficient Sequential Monitoring of Multiple Data Streams Via Shrinkage

Yuan Wang
School of Industrial and Systems Engineering, Georgia Institute of Technology, USA
Collaborators/co-authors: Yajun Mei

In this talk, we consider the sequential change-point detection problem where one monitors observations from multiple independent data streams but the post-change distributions involve unknown parameters. One intuitively appealing scheme is to estimate the post-change parameters from past observations and then raise a global alarm based on Shiryaev-Roberts-type procedure. Lorden and Pollak (*Ann. Statist.*, 2005) established the first-order asymptotic optimality properties of such a scheme. Meanwhile, it is well-known from off-line statistics that shrinkage can improve the performance of estimators in the multi-dimensional scenario, see James and Stein (1961). This motivates us to investigate the impact of James-Stein-type shrinkage on the scheme proposed in Lorden and Pollak (2005). We will report theoretical analysis on the second-order asymptotic properties of the new scheme, and present numerical simulation results.

Change-Point Detection with Multiple Sensors

Tiansong Wang
Department of Mathematical Sciences, The University of Texas at Dallas, USA
Collaborators/co-authors: Michael Baron

Sequential change-point detection schemes are developed that are based on simultaneously observed multiple data sequences. It is assumed that a number of sensors simultaneously collect and report data. When a significant event occurs, the distribution of data in one or several sequences changes,

and the goal is to detect such a change as soon as possible after it occurs, subject to the rates of false alarms and incorrect conclusions (Type III errors). Detection schemes are derived by the Bonferroni approach as well as the Holm stepwise idea that has been recently implemented for multiple testing in sequential experiments. Performance of the proposed procedures is evaluated, in particular, as a function of the number of sensors.

A Sequential Procedure for Multiple Window Scan Statistics

Tung-Lung Wu
Department of Statistics, University of Connecticut, USA
Collaborators/co-authors: Joseph Glaz

The literature on the scan statistics is substantial and growing rapidly due to the widespread applications in many areas, such as genetics and quality control. The distributions of discrete, continuous and conditional multiple window scan statistics are studied. In addition, it has been of practical interest on how to select the window size of a scan statistic. An adaptive procedure for multiple window scan statistics is introduced. The idea of the procedure is to select the data-dependent window sizes sequentially. We compare the power of our adaptive scan statistic with standard scan statistics for discrete case. Numerical results and an application for disease clusters detection are given to illustrate our procedure.

Stable Estimation in Dimension Reduction by Sub-Sampling with Random Weights

Wenbo Wu
Department of Statistics, University of Georgia, USA
Collaborators/co-authors: Xiangrong Yin

Many dimension reduction methods form dimension reduction matrices, which can be transformed into a regression problem and then adopt shrinkage estimators to produce more accurate estimates, such as sparse eigen-decomposition estimation (SED, Zhu et al. 2010) and sparse sufficient dimension reduction (SSDR, Li 2007). However, penalized estimates can be instable due to the sensitiveness of the selection of tuning parameters. We propose a stable approach to alleviate the influence of tuning parameters. The proposed method provides consistent results in testing dimension of central subspace and only selects the correct directions in central subspace. We also propose a thresholding procedure to obtain more accurate estimates for sparse and non-sparse basis directions of central subspace. Simulation studies and a data analysis demonstrate the efficacy of our proposed stable procedures.

A General Method for Allocating Recycled Significance Levels in Group Sequential Procedures for Multiple Endpoints

Dong Xi
Department of Statistics, Northwestern University, USA
Collaborators/co-authors: Ajit C.. Tamhane

Bretz et al. (2009) and Burman et al. (2009) independently proposed a graphical approach for testing hypotheses on multiple endpoints by recycling significance levels from rejected hypotheses

to unrejected ones. Maurer and Bretz (2013) extended this graphical approach to constructing group sequential procedures (GSPs). Ye et al. (2013) studied the same extension in a special setting. This extension requires a method to allocate the recycled significance level from rejected hypotheses to the analyses of the GSPs for unrejected hypotheses. Ye et al. (2013) studied two methods, one of which (also used by Maurer and Bretz, 2013) allocates the recycled significance level to all analyses while the other allocates it only to the final analysis. Both these methods have some drawbacks. We propose a general method that allocates the recycled significance level to analyses $k \geq r$ where r is prespecified. One can choose r to minimize the expected sample size for a given power requirement. Alternatively, r can be chosen to provide adequate sample size for obtaining sufficient data on safety and any secondary outcomes. Simulation studies to compare powers of alternative methods are carried out and illustrative examples are given.

Detecting Change-Point in Signal Correlation

Yao Xie

Duke Univerity and Georgia Institute of Technology, USA

Detecting a change-point in correlation of a sequence of data is an important aspect of many engineering applications such as detecting anomaly of power lines and detecting spectrum opportunity in wireless communications. Here in particular, we consider detecting a change-point, where before the change-point the data are i.i.d. white noise before, and after the change-point the data can be modeled as an auto-regressive moving average (ARMA) process. We assume the order of the ARMA process is known, but the model coefficients are unknown. We develop a new Rao's score detector, which avoids the computational burdensome inverse of the covariance matrix needed by the maximum likelihood detector. We also obtain analytic approximation for the performance metrics of the score detector. This is joint work with David Siegmund at Stanford University.

Metric-Based Multiple Image Registration

Qian Xie

Department of Statistics, Florida State University, USA

Collaborators/co-authors: Sebastian Kurtek, Eric Klassen, Anuj Srivastava

Registering pairs or groups of images is a widely-studied problem that has seen a variety of solutions in recent years. Most of these solutions are variational, using energy functions that fail to satisfy two most basic and desired properties in registration: (1) invariance under identical warping: since the registration between any two images is unchanged under identical domain warping, the objective function evaluating registrations should also remain unchanged; (2) inverse consistency: the optimal registration of image A to B should be the same as that of image B to A. We present a novel registration approach that uses the L^2-norm, between certain functions derived from images, as an objective function for registering images. This framework satisfies the invariance and the symmetry properties. Additionally, our framework induces a metric in the space of equivalence classes of images which enables us to perform joint registration of multiple images, using the concept of a mean image. We demonstrate this framework using examples from a variety of image types and compare performances with some recent methods.

First Crossing Times of Compound Poisson Processes with Two Linear Boundaries - Applications in SPRT and Queueing

Yifan Xu
Department of Mathematical Sciences, State University of New York at Binghamton, USA

Compound Poisson processes (CPP) are important in queueing theory, insurance risk modeling and SPRT. We consider CPPs with positive absolutely continuous jumps, and derive exact distributions of first crossing times of such CPPs with both upper and lower piece-wise linear boundaries. We also give a numerical approximation method to reduce the calculation complexity in the case of parallel boundaries. In sequential probability ratio tests of distributions from exponential families, with Poissonized sampling procedure, the first crossing time of CPP is used to calculate the exact power function of the test. In queueing theory, the crossing times reveal time dependent behaviors of M/G/1 queues with workload (virtual waiting time in the first-come-first-served policy) restriction. To be more specific, the first upper crossing time is the first time that the workload of the queueing system exceeds a predetermined threshold; The lower crossing time is the busy period where the workload stays below the threshold.

Asymptotic Normality for Inference on Multi-Sample, High-Dimensional Mean Vectors under Mild Conditions

Kazuyoshi Yata
Institute of Mathematics, University of Tsukuba, Japan

In this talk, we consider the asymptotic normality for inference on multi-sample and high-dimensional mean vectors. We relax the conditions to verify the asymptotic normality of concerned statistic when the data dimension is very high. We verify the asymptotic normality under some mild conditions. With the help of the asymptotic normality, we show that concerned statistics can ensure consistency for inferences on multi-sample and high-dimensional mean vectors. We give sample size determination so that those inferences can ensure prespecified accuracy. We demonstrate how to apply the asymptotic normality to various inferences on multi-sample and high-dimensional mean vectors by using microarray data.

Sequential Joint Detection and Estimation

Yasin Yilmaz
Electrical Engineering Department, Columbia University, USA
Collaborators/co-authors: George V. Moustakides, Xiaodong Wang

We consider the problem of simultaneous detection and estimation under a sequential framework. In particular we are interested in sequential tests that distinguish between the null and the alternative hypotheses and every time the decision is in favor of the alternative hypothesis they provide an estimate of a random parameter. As we demonstrate with our analysis treating the two subproblems separately with the corresponding optimal strategies does not result in the best possible performance. To enjoy optimality one needs to take into account the optimum estimator during the hypothesis testing phase.

Exact Distributions of Stopping Times in Sequential Analysis

Shelemyahu Zacks
Department of Mathematical Sciences, SUNY Binghamton, USA

The talk consists of three parts. The first part develops the exact distribution of the stopping time in a truncated SPRT, for testing hypotheses about the mean of exponential distribution. We derive the exact formulae of the expected stopping time and the operating characteristic function of the procedure. In part two, we discuss the exact distribution of the stopping time, in sequential estimation of the mean of exponential distribution. We show the expected value and MSE of the estimator at stopping. In the third part, we develop the exact distribution of the stopping time in sequential estimation of the log-odds of Bernoulli trials.

Robustness of the N-CUSUM Stopping Rule in Change-Point Detection

Hongzhong Zhang
Statistics Department, Columbia University, USA
Collaborators/co-authors: Neofytos Rodosthenous , Olympia Hadjiliadis.

This work considers the problem of quickest detection of signals in a system of N sensors coupled by *arbitrary* correlated noises, which receive continuous sequential observations from the environment. It is assumed that the signals are time invariant, but with different strength and their onset times may differ from sensor to sensor. The objective is the optimal detection of the first time at which any sensor in the system receives a signal. The problem is formulated as a stochastic optimization problem in which an extended Lorden's criterion is used as a measure of detection delay, with a constraint on the mean time to the first false alarm. The case in which the sensors employ their own cumulative sum (CUSUM) strategies is considered, and it is proved that the minimum of N CUSUMs is asymptotically optimal as the mean time to the first false alarm increases without bound. Implications of this asymptotic optimality result to the efficiency of the decentralized versus the centralized system of observations are further discussed.

Multiple Testing in Group Sequential Clinical Trials

Tian Zhao
Department of Mathematical Sciences, University of Texas at Dallas, Texas, USA
Collaborators/co-authors: Michael Baron

Statistical procedures are developed for the simultaneous testing of multiple hypotheses in group sequential experiments. Proposed tests control both the Type I and Type II familywise error rates. The problem is different from the standard SPRT because of the restriction on the maximum allowed number of sampled groups, which is a typical situation in sequential clinical trials. Multiple testing procedures can be derived on basis of Pocock, O'Brien-Fleming, and other classical tests. Further optimization is available that yields lower expected sampling costs of a sequential experiment than one can obtain by the Bonferroni approach.

List of Participants

Name	Affiliation	Email
Aoshima, Makoto	University of Tsukuba	aoshima@math.tsukuba.ac.jp
Arkoun, Ouerdia	University de Rouen	ouerdia.arkoun@gmail.com
Banerjee, Taposh	Coordinated Science Laboratory	banerje5@illinois.edu
Banerjee, Swarnali	University of Connecticut	swarnali009@gmail.com
Baron, Michael	University of Texas at Dallas	mbaron@utdallas.edu
Bhattacharjee, Debanjan	Utah Valley University	debanjan.bhattacharjee@gmail.com
Brown, Marlo	Niagara University	mbrown@niagara.edu
Buonaguidi, Bruno	Bocconi University	bruno.buonaguidi@phd.unibocconi.it
Buzaianu, Elena	University of North Florida	e.buzaianu@unf.edu
Chakraborty, Adrijo	University of Georgia	adrijoc@uga.edu
Chang, Shing	Kansas State University	changs@ksu.edu
Chattopadhyay, Bhargab	University of Texas at Dallas	bhargab@utdallas.edu
Chicken, Eric	Florida State University	chicken@stat.fsu.edu
Cho, Hokwon	University of Nevada, Las Vegas	cho@unlv.nevada.edu
Chou, Shih-Hsiugn	Kansas State University	cls3415@ksu.edu
Christensen, Sören	Christan-Albrechts-University	christensen@math.uni-kiel.de
Coad, Steve	Queen Mary, University of London	d.s.coad@qmul.ac.uk
Darkhovsky, Boris	Institute for Systems Analysis	darbor2004@mail.ru
De, Shyamal Krishna	Binghamton University	sde@binghamton.edu
Dragalin, Vlad	Aptiv Solutions	vladimir.dragalin@aptivsolutions.com
Du, Wenyu	State University of New York	duwenyu2010@gmail.com
Fellouris, Georgios	Univ. of Southern California	Fellouri@usc.edu
Geng, Jun	Worcester Polytechnic Institute	jgeng@wpi.edu
Gombay, Edit	University of Alberta	egombay@ualberta.ca
Hasan, Abeer	Bowling Green State Univ	statlover0.05@hotmail.com
Hilafu, Haileab	University of Georgia	htesfe@uga.edu
Hu, Linwei	University of Georgia	pangaea@uga.edu
Hu, Chih-Chi	Columbia University	ch2547@columbia.edu
Hussein, Abdulkadir	University of Windsor	ahussein@uwindsor.ca
Ivanova, Anastasia	UNC at Chapel Hill	aivanova@bios.unc.edu
Kachiashvili, Kartlos	Georgian Technological Univ.	kartlos55@yahoo.com
Keener, Bob	University of Michigan	keener@umich.edu
Koch, Mark	Sandia National Laboratories	mwkoch@sandia.gov
Lai, Lifeng	Worcester Polytechnic Institute	llai@wpi.edu
Lerche, Hans	Dept of Mathematical Stochastics	lerche@stochastik.uni-freiburg.de
Li, Yan	Georgia Institute of Technology	liyan6750@126.com
Li, Huijuan	Rutgers University	hli@stat.rutgers.edu
Liu, Yu	University of New Orleans	lyu2@uno.edu
Lu, Richard	Georgia Institute of Technology	richard.wh.lu@gmail.com
Lund, Robert	Clemson University	Lund@clemson.edu
Malinovsky, Yaakov	University of Maryland (UMBC)	yaakov.malinovsky@gmail.com

Name	Affiliation	Email
Mandal, Abhyuday	University of Georgia	amandal@stat.uga.edu
Martin, Jacob	University of Georgia	jmarsn1215@gmail.com
McDonald, Josh	Georgia Institute of Technology	jmcdonald17@gatech.edu
Mei, Yajun	Georgia Institute of Technology	ymei@isye.gatech.edu
Moustakides, George	University of Patras	moustaki@upatras.gr
Mukhopadhyay, Nitis	University of Connecticut	nitis.mukhopadhyay@uconn.edu
Nikiforov, Igor	Université de Technologie de Troyes	igor.nikiforov@utt.fr
Ning, Wei	Bowling Green State University	wning@bgsu.edu
Novikov, Alex	University of Technology, Sydney	alex.novikov@uts.edu.au
Paynabar, Kamran	Georgia Institute of Technology	kamran.paynabar@isye.gatech.edu
Pergamenchtchikov, Serguei	University of Rouen	Serge.Pergamenchtchikov@univ-rouen.fr
Piryatinska, Alexandra	San Francisco State University	alpiryat@sfsu.edu
Pollak, Moshe	Hebrew University of Jerusalem	msmp@mscc.huji.ac.il
Polunchenko, Aleksey	Binghamton University	polunchenko@gmail.com
Poruthotage, Sankha	University of Connecticut	sankha.perera@uconn.edu
Prause, Annabel	RWTH Aachen University	prause@stochastik.rwth-aachen.de
Qiu, Peihua	University of Florida	qiuxx008@umn.edu
Raghavan, Vasanth	Univ of Southern California	vasanth_295@yahoo.com
Samadi, Yaser	University of Georgia	ysamadi@uga.edu
Sauls, Jonathan	Michelin	jonathan.sauls@us.michelin.com
Sheng, Wenhui	University of Georgia	wsheng@uga.edu
Sokolov, Grigory	University of Southern California	gsokolov@usc.edu
Solanky, Tumulesh	University of New Orleans	tsolanky@uno.edu
Sriram, T. N.	University of Georgia	tn@uga.edu
Steland, Ansgar	RWTH Aachen University	steland@stochastik.rwth-aachen.de
Su, Haiyan	Montclair State University	suh@mail.montclair.edu
Sverdlov, Alex	Novartis	alex.sverdlov@novartis.com
Tan, Xijue	University of Georgia	buddist@uga.edu
Tartakovsky, Alex	University of Southern California	tartakov@usc.edu
Thayasivam, Umashanger	Rowan University	thayasivam@rowan.edu
Veeravalli, Venu	Coordinated Science Laboratory	vvv@illinois.edu
Wang, Guannan	University of Georgia	guannan@uga.edu
Wang, Yuan	Georgia Institute of Technology	ywang489@gatech.edu
Wang, Tiansong	University of Texas at Dallas	txw107020@utdallas.edu
Wu, Wenbo	University of Georgia	wuwenbo@uga.edu
Wu, Tung-Lung	University of Connecticut	tung-lung.wu@uconn.edu
Xi, Don	Northwestern University	dong.xi@u.northwestern.edu
Xie, Yao	Duke University	su06ee@gmail.com
Xie, Qian	Florida State University	qxie@stat.fsu.edu
Xu, Yifan	State University of New York	yxu@math.binghamton.edu

Name	Affiliation	Email
Yata, Kazuyoshi	University of Tsukuba	yata@math.tsukuba.ac.jp
Yilmaz, Yasin	Columbia University	yasiny@gmail.com
Zacks, Shelley	Binghamton University	shelly@math.binghamton.edu
Zhang, Hongzhong	Columbia University	hz2244@columbia.edu
Zhao, Tian	University of Texas at Dallas	txz102020@utdallas.edu