



The Florida Chapter of the
American Statistical Association
Annual Meeting
February 3 and 4, 2012
Dept. of Math & Stat
1 UNF Drive, Jacksonville, FL. 32224



ABSTRACTS

Andrew Rosalsky

Department of Statistics University of Florida, Gainesville, FL

A Refinement of the Kolmogorov-Marcinkiewicz-Zygmund Strong Law of Large Numbers

For the partial sums formed from a sequence of i.i.d. random variables having a finite absolute p 'th moment for some p in $(0,2)$, we extend the recent and striking discovery of Hechner and Heinkel (Journal of Theoretical Probability (2010)) concerning "Complete Moment Convergence" from the case $1 < p < 2$ to the two cases where $0 < p < 1$ and $p = 1$. Moreover, for $0 < p < 2$, we obtain "Almost Sure Convergence" analogues of these "Complete Moment Convergence" results and these "Almost Sure Convergence" analogues may be regarded as being refinements of the celebrated Kolmogorov-Marcinkiewicz-Zygmund Strong Law of Large Numbers (Fundamenta Mathematica (1937)). Some extensions of our findings to partial sums of Banach space valued random elements will also be discussed but only rather briefly; the emphasis will be on real line results. This talk is based on work joint with Deli Li (Thunder Bay, Ontario, Canada) and Yongcheng Qi (Duluth, Minnesota).

Nancy L. Geller, NIH

2011 President, American Statistical Association (ASA)

How do I know if a recommended treatment works?

We are often faced with the decision of whether to take a recommended treatment or follow a recommended prevention regimen. We discuss designs of medical studies and compare other study designs to a randomized clinical trial (RCT), demonstrating the advantages of the latter. Examples of some RCTs of treatments widely believed to work are discussed, including hormone replacement therapy for prevention of heart disease in post-menopausal women, whether someone with a vertebral fracture should undergo a procedure that surgeons recommend and whether an over-the-counter lozenge helps avoid colds. How to draw conclusions from studies such as these is emphasized.

E. M. Buzaianu¹ P. Chen² T.-J. Wu³

Subset Selection Procedures to Identify EM Fields Following Log-normal Distributions

¹ Department of Mathematics and Statistics, University of North Florida, Jacksonville, , ²Department of Mathematics, Syracuse University, Syracuse, New York, ³Department of Statistics, National Cheng-Kung University, Tainan, Taiwan, Republic of China

We use statistical ranking and selection methodology to characterize multiple targets in electromagnetic (EM) fields. Previous research concluded that the observed EM field power fluxes and cable powers follow either a chi-square distribution with two degrees of freedom or a log-normal distribution. That is, such an EM field can be characterized by either an exponential distribution with mean μ or a log-normal distribution with parameters μ and σ . We consider the case when the distribution is log-normal. We assume that there are k clutter types and the distributions of the observed clutters are lognormal distributions with common σ . We would like to select all clutter returns that are matching or closed to a control clutter having a desired characteristic (signature) μ_0 . We propose subset selection procedures to such identify such EM fields, assuming there exists at least one good target clutter. We illustrate the properties of our procedures by numerical examples and simulation examples. The primary application of this research lies in the area of electromagnetic vulnerability (EMV) although the statistical theory developed here is as general as any other mathematical tools and the results can be applied to any electromagnetic problems in radar, sonar, and navigation systems.

Subhash C. Bagui¹, Sikha S. Bagui², and Rohan Hemasinha¹

¹Department of Mathematics and Statistics, ² Department of Computer Science, The University of West Florida, Pensacola, FL

Nonrigorous Proofs of Stirling's Formula

In this article, we present a new method of deriving Stirling's formula together with two existing methods. The new method uses a result on convergence of moments and Poisson distribution. The other two methods use changes of variables in the Gamma function and the Normal approximation of a Poisson probability, respectively. The proofs are elementary and could be useful in teaching. A numerical error analysis to gauge the accuracy of the approximation is also given.

Key Words: Central limit theorem, Gamma function, Factorial function, Beta function, Normal probability, Poisson probability, and Stirling's formula.

Eric Chicken
Department of Statistics Florida State University
Tallahassee, FL

Changepoint detection for high dimensional sequential data

A semiparametric method to simultaneously monitor sequences of profiles for changes in the functional structure and the magnitude of the error is proposed. This changepoint method makes no use of likelihood ratios and is essentially distribution free. The estimator makes use of the wavelet dimension reduction property and applies a unique weighting scheme to the wavelet coefficients of the observed profile data to maximize differences from one profile to the next. Simulation studies show the proposed estimator detects profile changes with power comparable to parametric likelihood estimators when only part of the in-control model is violated. When the entire in-control model is violated, the simulation results indicate a significant increase in the power and reduces false discovery rates.

Marianna Pensky
University of Central Florida, Orlando, FL

Laplace deconvolution with noisy observations and its application to analysis of DCE-CT data.

The Dynamic Contrast Enhanced Computed Tomography (DCE-CT) experiments follow the diffusion of a bolus of a contrast agent injected into a vein. The data obtained as a result of those experiments lead to solution of a noisy version of Laplace convolution equation. Although Fourier deconvolution has been studied by a number of authors, to the best of our knowledge, Laplace deconvolution has not been studied in statistical literature. The present paper essentially provides the first comprehensive statistical treatment of Laplace deconvolution problem. The main contribution of the paper is the explicit construction of the Laplace deconvolution estimator.

We show that the original Laplace deconvolution problem can be essentially reduced to estimating regression function and its derivatives on the interval of growing length.

Kate Fisher and Ji-Hyun Lee
Department of Biostatistics, Moffitt Cancer Center

Racial Disparities in Delay of Diagnosis and Treatment of Breast Cancer

Cancer is the second leading cause of death for most racial and ethnic minorities, and in the United States minority populations bear a heavier burden of cancer mortality than non-minority group. Specifically, although breast cancer is diagnosed less frequently in African American women than white women, African American women are more likely to die from the cancer. Such disparity has been attempted to understand with several factors, and delay in diagnosis and treatment has been one of explanations. The central theme of this study is to examine whether timely access to diagnosis and treatment is different across races. We will also examine whether such delays are associated with cancer specific mortality. The study will be nationwide in scope and will use Medicare insurance claims to quantify the delays and will assess cancer outcomes using the SEER tumor registry. Understanding the relationship between timeliness of evaluation once a cancer related abnormality is identified and cancer outcomes is critically important in developing effective national strategies for achieving cancer control goals and preparing for the significant changes that are occurring to our nation's health care system.

Edgard Maboudou and Jeremiah Perez Department of Statistics
University Of Central Florida

Hypothesis Testing for the Mean vector in High Dimension

We consider a test for the mean vector of independent and identically distributed multivariate normal random vectors where the dimension p is larger than or equal to the number of observations n . We propose a new test statistic based on a non-singular estimate of the covariance matrix to perform the testing. As the distribution of the statistic proposed is hard to derive, conclusions about the test are obtained by using sequential Monte Carlo procedure.

Hafiz M. R. Khan
Department of Biostatistics, Florida International University
Miami, FL

Predictive inference based on censored data with extremes

In this paper, we derive the likelihood function for the parameters given a type II censored sample which follows the two-parameter half-normal model. A Bayesian framework has been utilized in conjunction with an informative prior to derive the posterior density function for the parameters and the predictive density for a single future response. A simulation study is used to illustrate the predictive results for a future response. A real data set represents the stress-rupture life of kevlar 49/epoxy strands, which are subjected to constant sustained pressure at the 90% stress level until all have failed, is utilized to illustrate the results. The summary results of the posterior parameters are given by using the Markov chain Monte Carlo method

James U. Gleaton, M. Mahbubur Rahman,
University of North Florida

Asymptotic Properties of MLE's for Distributions Generated from a 2-Parameter Inverse Gaussian Distribution by a Generalized Log-Logistic Transformation

A generalized log-logistic (GLL) family of lifetime distributions is one in which any pair of distributions are related through a GLL transformation, for some (non-negative) value of the transformation parameter (the odds function of the second distribution is the $-$ th power of the odds function of the first distribution). We consider GLL families generated from a 2-parameter inverse Gaussian distribution (GLLIG families). It is shown that, for $\kappa \geq 3$, the Maximum Likelihood Estimators (MLE's) for the parameters of the generated, or composite, distribution have the properties of strong consistency and asymptotic normality and efficiency.

Keywords Generalized log-logistic families; asymptotic properties; 2-parameter inverse Gaussian

Mark Soskin and Bradley Braun
University of Central Florida, College of Business

Innovation in Effective Delivery of Streaming Video Instruction to the Mass-Section Business Statistics Course

Undergraduate business statistics has traditionally been among the least effective courses in the Business School curriculum. Learning outcome performance continues to be resistant to reforms in course design, textbooks, statistical software, and pedagogy. Business stat learning is further burdened by large class size, as universities treat business courses as “cash cows.” Technology has recently enabled higher education to realize cost savings from streaming video, breaking the constraints on class size of lecture hall and parking lot capacity while gaining access to more distant enrollment regions. This paper describes implementation methods to effectively deliver business statistics to 1200 student via streaming video. Now in the second year as course designers and instructors, the authors identify and tackle the challenges and opportunities that streaming video present to cost-strapped universities today. Among the approaches adopted for this unique delivery mode include an innovative in-house text, Blackboard case projects, testing center examinations, PowerPoint show, and lecture strategy.

Anthony Ruble,
University of West Florida, Pensacola, Florida

Support Vector Machines (SVM): It's All About the Kernel.

In using Support Vector Machines (SVM) for classification, poor results may be caused by inappropriate use of kernels. Error rates are examined for different kernels in the use of SVM to classify Breast Cancer. The popular Wisconsin Breast Cancer data sets are used. We analyze techniques to find the best kernel values. We hope to show that by using some simple kernel selection methods better kernel values are obtained. These better values result in better classification results, improving the role Support Vector Machines having in classifying Breast Cancer.

Zhengwu Zhang
Florida State University

Blurring-Invariant Riemannian Metrics for Signals and Images

We propose a novel Riemannian framework for comparing signals and images in a manner that is invariant to their levels of blur. This framework uses a log-Fourier representation of signals/images in which the set of all possible Gaussian blurs of a signal. With Riemannian metrics we defined, we demonstrate this framework using a number of experimental results involving 1D signals and 2D images.

Paul Hill
Florida State University

Bootstrap Prediction Bands for Non-Parametric Function Signals in a Complex System

Methods employed in the construction of prediction bands for continuous curves require a different approach to those used for a data point. In many cases, the underlying function is unknown and thus a distribution-free approach which preserves sufficient coverage for the signal in its entirety is necessary in the signal analysis. Three methods for the formation of $(1-\alpha)100\%$ prediction bands are presented and their performances are compared through the coverage probabilities obtained.

These techniques are applied to constructing prediction bands for spring discharge in a successful manner giving good coverage in each case. Spring discharge measured over time can be considered as a continuous signal and the ability to predict the future signals of spring discharge is useful for monitoring flow and other issues related to the spring.

There has been common use of the gamma distribution in the simulation of rainfall. Bootstrapping the rainfall in the proposed manner, allows for adequately creating new samples over different periods of time as well as specific rain events such as hurricanes or drought. This non-parametric approach to the input rainfall augurs well for the non- parametric nature of the output signal.

Cen Wu, Gengxin Li and Yuehua Cui
Department of Statistics and Probability, Michigan State University, East
Lansing, Michigan

Functional mapping of dynamic traits with robust t-distribution

Functional mapping has been a powerful tool in mapping quantitative trait loci (QTL) underlying dynamic traits of agricultural or biomedical interest. In functional mapping, multivariate normality is often assumed for the underlying data distribution, partially due to the ease of parameter estimation. When QTLs are present, heavy tails are often observed due to the mixture distribution of the data at the QTL locus. Thus, the normality assumption is often violated, especially under a small sample size. Departure from normality has negative effect on testing power and inference for QTL identification. In this work, we relaxed the normality assumption and proposed a robust multivariate t-distribution mapping framework for QTL identification in functional mapping. Simulation studies show increased mapping power and precision with the t distribution than that of a normal distribution. The utility of the method is demonstrated through a real data analysis.

Dimitris Vovoras¹, Keshav P. Pokhrel¹ Dr. Chris P. Tsokos¹

1. Department of Mathematics and Statistics, University of South Florida,
2. Spinal and Skull Base Oncology, H. Lee Moffitt Cancer Center, USA

APPLICATIONS OF JOIN POINT REGRESSION MODELS TO SURVIVAL RATES OF BRAIN TUMOR PATIENTS IN THE USA

To investigate the relationship between medical improvements and the survival experienced by the patient population, it would be useful to find out when and how much the cancer treatment breakthroughs and early diagnosis have significantly improved the prognosis of brain cancer patients. A join point model facilitates the identification of trends with significant change-points in survival; the main goal of such a model would be to find out when cancer survival starts exhibiting a pattern of improvement. The model will be applied to grouped relative survival data for major cancer sites from the 'Surveillance, epidemiology and end results' program of the National Cancer Institute.

Key words: brain tumor, Cox-PH, joinpoint, SEER, spline, survival

Hao W. Zheng¹, Babette A. Brumback^{1*}, Xiaomin Lu¹, Erin D. Bouldin², Michael B. Cannell², and Elena M. Andresen³

¹Department of Biostatistics ²Department of Epidemiology University of Florida Gainesville, FL , ³Disability and Health Research Group Child Development and Rehabilitation Center Oregon Health and Sciences University Portland, OR

Doubly Robust Testing and Estimation of Model-Adjusted Effect-Measure Modification with Complex Survey Data

Model-based standardization enables adjustment for confounding of a population-averaged exposure effect on an outcome. It requires either a model for the probability of the exposure conditional on the confounders (an exposure model), or a model for the expectation of the outcome conditional on the exposure and the confounders (an outcome model). The methodology can also be applied to estimate averaged exposure effects within categories of an effect modifier, and to test whether these effects differ or not. Recently, we extended that methodology for use with complex survey data, in order to estimate the effects of disability status on cost barriers to health care within three age categories, and to test for differences. We applied the methodology to data from the 2007 Florida Behavioral Risk Factor Surveillance System (BRFSS) Survey. The exposure modeling and outcome modeling approaches yielded two contrasting sets of results. In the present paper, we develop and apply to the BRFSS example two doubly robust approaches to testing and estimating effect modification with complex survey data; these approaches require that only one of these two models be correctly specified. Furthermore, assuming that at least one of the models is correctly specified, we can use the doubly robust approaches to develop and apply goodness-of-fit tests for the exposure and outcome models. We compare the exposure modeling, outcome modeling, and doubly robust approaches in terms of a simulation study and the BRFSS example.

KEY WORDS: Confounding; Model-Based Standardization; Heterogeneity; Interaction; Complex Survey Data; Causal inference; Doubly Robust.

Kelly McGinnity ,Eric Chicken
Florida State University, Tallahassee, FL

Nonparametric Cross-Validated Wavelet Thresholding for Non-Gaussian Errors

Wavelet thresholding generally assumes independent, identically distributed normal errors when estimating functions in a nonparametric regression setting. VisuShrink and SureShrink are just two of the many common thresholding methods based on this assumption. When the errors are not normally distributed, however, few methods have been proposed. In this paper, a distribution-free method for thresholding wavelet coefficients in nonparametric regression is described. Unlike some other non-normal error thresholding methods, the proposed method does not assume the form of the non-normal distribution is known. A simulation study shows the efficiency of the proposed method on a variety of non-normal errors, including comparisons to existing wavelet threshold estimators.

Jingyong Su, Sebastian Kurtek, Eric Klassen, Anuj Srivastava
Florida State University

A Novel Framework for Joint Registration, Comparison and Averaging of Paths on Nonlinear Manifolds

We address the problem of registration, comparison and averaging of temporal paths on nonlinear Riemannian manifolds such as spheres and shape manifolds. Past methods for registration of trajectories on manifolds, such as those used in temporal alignment of human activity data, used quantities that are not proper distances (they are not even symmetric). Without a proper distance, it is difficult to define average paths (or templates) or setup a classification solution. An important property needed here is that the chosen quantity should be a distance and it should be invariant to identical time-warpings (or re-parameterizations) of the paths. We present a novel framework that defines a proper (warping-invariant) distance between trajectories and is used in alignment, comparison and averaging. The chosen distance is based on a mathematical representation of trajectories, termed square-root velocity field, and the L2 metric on such vector fields. We illustrate this framework using examples of temporal alignment and averaging of trajectories on a simple manifold (S²) and a more complicated manifold (shape manifold of curves). We also present results from an application of this framework to activity classification where each activity is viewed as a path on a shape manifold.

Rida Benhaddou and Marianna Pensky
Department of Mathematics, University of Central Florida

Adaptive Nonparametric Empirical Bayes Estimation Via Wavelet Series

Empirical Bayes (EB) methods are estimation techniques in which the prior distribution, in the standard Bayesian sense, is estimated from the data. They are powerful tools in particular when data are generated by repeated execution of the same type of experiment.

The present paper proposes generalization of the linear empirical Bayes estimation method which takes advantage of the flexibility of the wavelet techniques. We present an empirical Bayes estimator as a wavelet series expansion and estimate coefficients by minimizing the prior risk of the estimator. Although wavelet series have been used previously for EB estimation, the method suggested in the paper is completely novel since the EB estimator as a whole is represented as a wavelet series rather than its components. Moreover, the method exploits de-correlating property of wavelets which is not instrumental for the former wavelet-based EB techniques.

As a result, estimation of wavelet coefficients requires solution of a well-posed sparse system of linear equations. The dimension of the system depends on the size of wavelet support and smoothness of the Bayes estimator. The number of nonzero diagonals of the matrix of the system is equal to the length of the support of the father wavelet, and, furthermore, this matrix tends to identity matrix as the scale tends to zero. The paper also proposes an adaptive choice of the resolution level using Lepskii (1997) method. The technique proposed in the paper is motivated by computational efficiency, nevertheless, it provides asymptotically optimal EB estimators posterior risks of which tend to zero at an optimal rate.

Rachel Becvarik
Florida State University, Tallahassee, FL

Change Points in Non-stationary High Dimensional Data

It has been shown likelihood ratio tests successfully monitor for changes in profiles involving high dimensional nonlinear data. These methods focus on using a traditional flat line upper control limit (UCL) based on average run length (ARL). The current methods do not take into consideration either the error or power associated with the test or the underlying distribution of the ARL. Additionally, if the statistic is known to be increasing over time, the flat UCL does not adapt to the increase. This paper will focus on a method to find the most powerful UCL for an increasing statistic at a specified α value.

Russell Romanella
Director Safety and Mission Assurance, NASA KSC
NASA Kennedy Space Center, Florida

NASA - Exploration at a Crossroad

Mr. Romanella explores the current Human and Robotic Exploration Missions within NASA, where we are and where we are heading with the retirement of Space Shuttle and Space Station programs as well as exploration of the solar system including Mars, Saturn, the asteroid belt, and NASA plans to return to the Moon and on to Mars



**Building 58 Student Union Ballroom 3703
University of North Florida**