

Titles and Abstracts

Short Course

William Meeker, Iowa State University

Statistical Methods for Product Life Analysis and Accelerated Testing

Reliability improvement and reliability assurance processes in manufacturing industries require data-driven reliability information for making business, product-design, and engineering decisions. This will be a hands-on workshop where participants will use the JMP 11 software for analyzing reliability data and test planning. The course will focus on concepts, examples, models, data analysis, and interpretation. Examples and exercises will include examples using product field (maintenance or warranty) data, accelerated life tests, and accelerated degradation tests. After completing this course, participants will be able to recognize and properly deal with different kinds of reliability data and properly interpret important reliability metrics. Topics will include the use of probability plots to identify appropriate distributional models (e.g. Weibull and lognormal distributions), estimating important quantities like distribution quantiles and failure probabilities, the analysis of data with multiple failure modes, the analysis of both destructive and repeated measures degradation data, and the analysis of recurrence data from a fleet of systems or a reliability growth program.

Keynote Presentation

Antonio Possolo, National Institute of Standards and Technology

Shape Metrology

Shape metrology is a multidisciplinary program at the National Institutes of Standards and Technology, involving statisticians, mathematicians, and scientists working in several different fields, aiming to develop measurement services for the shape of objects relevant to the execution of the agency's mission, which is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life. This presentation highlights several aspects of this program, to reveal the contributions that methods of applied statistics are making to it, including for the representation and measurement of shapes, in particular of polymeric scaffolds used for tissue engineering, and of mineral grains. We will also review how deformable templates are being used in chemical spectroscopy. And in both cases we will discuss probabilistic models that may be used to describe the variability of the corresponding shapes.

BIO: Antonio Possolo (Chief, Statistical Engineering Division, Information Technology Laboratory, NIST) holds a Ph.D. in Statistics from Yale University. Besides his current role in government, he has previous experience in industry (General Electric, Boeing), and in academia (Princeton University, University of Washington in Seattle, University of Lisbon). He is committed to the development and application of probabilistic and statistical methods that contribute to advances in science and technology, and in particular to measurement science.

Invited Presentations

Richard Davis, Columbia University

Applications of the Extremogram to Time Series and Spatial Processes

In this talk, we discuss the application of the extremogram to the modeling of heavy-tailed multivariate time series and spatial-temporal processes. Like the autocorrelation function in time series, the extremogram can be used in various phases of the modeling exercise for heavy-tailed/extremal dependence in temporal and spatial processes. First, the extremogram provides a measure of extremal dependence in the data as a function of the time (or spatial) lag. Plots of the extremogram must include confidence bands to assess significant extremal dependence and this can be achieved using permutation procedures and/or block bootstrap procedures. Second, the extremogram can provide an assessment about how well the estimated extremogram matches the population extremogram based on a fitted model. Finally, the extremogram of residuals from a fitted model can be examined to see if extremal dependence has been satisfactorily removed. We illustrate the use of these ideas with several examples.

David Hunter, Pennsylvania State University

Model-Based Clustering of Large Networks

A network clustering framework, based on finite mixture models, is described. It can be applied to discrete-valued networks with hundreds of thousands of nodes and billions of edge variables. Relative to other recent model-based clustering work for networks, we introduce a more flexible modeling framework, improve the variational-approximation estimation algorithm, discuss and implement standard error estimation via a parametric bootstrap approach, and apply these methods to much larger datasets than those seen elsewhere in the literature. The more flexible modeling framework is achieved through introducing novel parameterizations of the model, giving varying degrees of parsimony, using exponential family models whose structure may be exploited in various theoretical and algorithmic ways. The algorithms, which we show how to adapt to the more complicated optimization requirements introduced by the constraints imposed by the novel parameterizations we propose, are based on variational generalized EM algorithms, where the E-steps are augmented by a minorization-maximization (MM) idea. The bootstrapped standard error estimates are based on an efficient Monte Carlo network simulation idea. Last, we demonstrate the usefulness of the model-based clustering framework by applying it to a discrete-valued network with more than 131,000 nodes and 17 billion edge variables.

David Marchette, Naval Surface Warfare Center, Dahlgren Division

A Statistical Analysis of a Time Series of Twitter Graphs

In this talk I will describe a set of Twitter data that we have been collecting for nearly two years. Using the Twitter streaming API, we collect all tweets geolocated within a set of rectangles covering the main land-masses of the world, as well as tweets containing certain key phrases. We collect "all" geolocated tweets, in the sense that Twitter provides all the tweets that are geolocated within the rectangle, provided the volume does not exceed a fixed limit. These tweets define a "reference" digraph -- each screen name is a vertex and there is an edge from s to t if a tweet from s refers to t : @s:"@t u wanna go to lunch?". These reference digraphs can be computed on time intervals to produce a time series of graphs. These graphs tend to have power law degree distributions, and I will describe the graphs and discuss some thoughts on how one might model these graphs. Using the graphs, I will discuss methods for inferring node attributes, such as the geolocation of a user whose tweet is not geolocated, or detecting spoofed geolocations.

Shane Reese, Brigham Young University

Mixing Apples and Oranges: Complex System Reliability with Multi-Modal Testing

We describe a hierarchical model for assessing the reliability of multi-component systems. Novel features of this model are the natural manner in which failure time data collected at either the component or subcomponent level is aggregated into the posterior distribution, and pooling of failure information between similar components. Prior information is allowed to enter the

model in the form of actual point estimates of reliability at nodes, or in the form of prior groupings. Censored data at all levels of the system are incorporated in a natural way through the likelihood specification. We demonstrate a fully Bayesian approach with incorporation of diverse data types, including discrete, continuous, and censored data at multiple data collection levels (components versus systems). The framework introduced includes accommodation for many commonly encountered system structures, including series, parallel, k out of n, and more complex Bayesian networks. We illustrate the methodology on an actively developed DoD system.

Special Session 1: Predictive Analytics

Organizer: Tom Donnelly, SAS Institute

Robert Gramacy, University of Chicago

Practical Large-Scale Computer Model Calibration to Real Data

As computational horsepower becomes ever cheaper, practitioners increasingly augment physical and field experiments with (potentially vast amounts) of computer simulation output derived from mathematical models. Computer simulations can be biased, but often have free parameters (or "knobs") describing unknown physical quantities of the mathematical system that can be tuned to adjust their behavior. Computer model calibration is the exercise of simultaneously adjusting those knobs, while building an emulator for computer model output that closely matches the real data, and then estimating the (ideally minimized) bias, so that the fitted model(s) can be used to make predictions for novel configurations in the field. The prevailing statistical apparatus used for calibration involves jointly modeling field and simulation data with coupled Gaussian process models (GPs) and inference via Markov chain Monte Carlo (MCMC). Although pleasing technically, there are practical challenges: (1) GPs and MCMC are cumbersome with large simulation data; and (2) the joint modeling framework is not very modular, meaning that bespoke implementation is usually required. In this talk, we pare down the canonical approach and demonstrate how effective calibration can be performed in modern large-data simulation contexts by patching together off-the-shelf R libraries for approximate GP inference and blackbox optimization. In addition to some pedagogical synthetic examples, we show a real-data calibration from a radiative shock experiment. This is joint work with Derek Bingham at Simon Frasier University.

Chris Gotwalt, SAS Institute

Interactive Model Building Using JMP

In this presentation we illustrate tools in JMP for building and interpreting models on data from designed experiments as well as observational data sets. In particular, we demonstrate the Generalized Regression platform, the Variable Importance module of the Profiler, and take a sneak peek of some interactive model building features coming in JMP 12.

Andrew Fast and John Elder, Elder Research

Ensemble Methods in Data Mining

Ensemble methods are one of the true disruptive technologies in data mining and machine learning. The simple idea of combining multiple models into one usually leads to significant improvements in model performance. We highlight two recent developments in ensembles: Importance Sampling and Rule Ensembles and show how these methods are generalizations of classical ensemble methods including bagging, random forests, and boosting. Finally, we explain the paradox of how ensembles achieve greater accuracy on new data despite their apparently much greater complexity by showing the connection between ensemble methods and regularization techniques.

Special Session 2: Applications of Multivariate Heavy-Tailed Statistics

Organizer: Sidney Resnick, Cornell University

Gennady Samorodnitsky, Cornell University

Tail Inference: Where Does the Tail Begin?

The quality of estimation of heavy tail parameters, such as tail index in the univariate case, or the spectral measure in the multivariate case, depends crucially on the part of the sample included in the estimation. A simple approach involving sequential statistical testing is proposed for choosing this part of the sample. This method can be used both in the univariate and multivariate cases. It is computationally efficient, and can be easily automated. No visual inspection of the data is required. We establish consistency of the Hill estimator when used in conjunction with the proposed method, as well describe its asymptotic fluctuations. We compare our method to existing methods in univariate and multivariate tail estimation, and use it to analyze Danish fire insurance data.

John Nolan, American University

Computational Tools for Non-Gaussian Multivariate Distributions

In an increasing number of applications, there are multivariate data sets that are poorly modeled by a Gaussian distribution. This can be due to non-elliptical spread of the data. For this type of data, we describe the generalized spherical distributions, a class of distributions that are determined by a contour (the level curve of the density) and a radial decay function. An R software package GeneralizedSpherical implements a flexible family of such distributions. This gives a versatile tool for modeling non-standard distributions.

Don Towsley, University of Massachusetts

Sampling Heavy-Tailed Multivariate Distributions on Large Networks

Estimating characteristics of large graphs via sampling is a vital part of the study of complex networks. Current sampling methods based on (independent) random vertex and random walks (RWs) have been shown to be useful when the graph underlying the network is undirected. In particular, sampling based on random walks has been shown to be particularly effective for characterizing the tail of distributions strongly related to the degree distribution. However, many large networks are directed (e.g., Twitter, Wikipedia, Flickr) and the quantities of interest are often multivariate, in-degree and out-degree being prime examples. In this talk, we explore various random walk based sampling algorithms, paying particular attention to their effectiveness in characterizing the joint in-degree/out-degree distribution. In particular, we examine how the underlying graph supporting the RW affects the quality of the characterization. For example, should the walker ignore edge direction? Travel in/against the direction of the edges? We also examine how the dependence between in-/out-degrees affects the performance of RW-based estimation. Last, we will examine other characteristics such as reciprocity and clustering coefficient.

Sidney Resnick, Cornell University

Models with Hidden Regular Variation: Generation and Detection

We review definitions of multivariate regular variation (MRV) and hidden regular variation (HRV) for distributions of random vectors and then summarize methods for generating models exhibiting both properties. We also discuss diagnostic techniques that detect these properties in multivariate data and indicate when models exhibiting both MRV and HRV are plausible fits for the data. We illustrate our techniques on simulated data and also two real Internet data sets. (Joint work with Bikramjit Das, Singapore University of Technology and Design)

Special Contributed Session 1: Reliability, Rare-Events, and Censored Data

Organizer: Shuguang Song, Boeing Company

Russell W. Morris, Technical Fellow of The Boeing Company

Application of Failure Mode Distributions to Reliability Analysis of Combat Systems

Dragos D. Margineantu, Technical Fellow of The Boeing Company

Statistical Tests for Rare Event Predictors

Standard normal-distribution based tests are inappropriate for assessing the risk of predictions in the case of rare classes. This talk will motivate this statement and the need for a new tests for predictors employed high-risk and rare-event decisions. We will also propose new tests based on the bootstrap and will present and discuss an assessment of the newly proposed tests.

Ying Zhang, Indiana University

Tensor Spline-Based Sieve Nonparametric Maximum Likelihood Estimation for Bivariate Current Status Data

Special Contributed Session 2: Network Science

Organizer: David Jakubek

Stephen Russell, Naval Research Laboratory

Applications of the Law of First Digits in Cyber and Soft Biometrics

Benford's Law is derived from an observation that the frequencies of the first digit in naturally occurring measurements follow a power law distribution. Applications of Benford's Law have found success in several domains identifying fraudulent and aberrant measures including: financial auditing, elections, and economic data analysis. Independent of Benford's Law, research has shown that complex systems have a tendency to produce measured observations that follow the power law, and this also applies to measurements of human behavior. This would suggest that Benford's Law would have utility in applications concerning observations with embedded human behavior. Recent attention on cyber and network-based soft biometrics has driven a great deal of research on modeling the human behavior in networked environments. However despite the dependence on computational machinery, few of these models explicitly isolate machine-generated noise from human signatures. As a means of large data preprocessing, Benford's Law may provide a heuristic method for checking consistency and localization prior to downstream analytics or prediction. This talk presents background on Benford's Law and early experimentation applying it to cyber and soft biometric data.

Elizabeth Bowman, Army Research Laboratory

Statistical Approaches in Mining Unstructured Text for Contextual Understanding and Event Prediction

The Office of the Secretary of Defense (OSD) Data to Decisions (D2D) program is a multi-year program that supports rapid maturation of technologies to support contextual understanding and event prediction for military operations. Information is required from a variety of sources and diverse formats to support awareness of cultures, attitudes, events, and relationships in an area of interest. Text analytics is the organizing domain of this program and refers to the linguistic, statistical, and machine learning techniques that model and structure the information content of text sources for exploratory data analysis, research, and investigation. The contextual

understanding component of this program supports the discovery of events, stated values and beliefs that motivate behaviors of interest, discovery of topics and concepts in a shared community, analysis of semantic relationships and associated strength, community structure and clusters of social networks, and semantic analysis and trending of support expressed in text toward topics or persons. The event prediction aspect of D2D seeks to advance capabilities for extracting events from unstructured text with the following aspects: identification of proxy features of a network, temporal trend extraction and characterization, visual and semantic analysis at scale, key actor and supported relationship associations, and detection of bridging nodes that can uncover hidden sub-networks and determine the flow of resources within a network. Natural Language Processing (NLP) methods for text analysis rely upon statistical and language/rule-based methods. Some hybrid approaches have been shown to be effective with deep NLP and semantic disambiguation. This talk will cover a range of statistical models used by D2D performers and will address issues arising in the technical development process.

Robert Bonneau, OASD (R&E)

Risk Analytics for Complex Information Networks

The talk will provide an overview of complex information systems including quantifying, managing, and designing heterogeneous networked systems. Methods of measuring and assessing the performance of network, software, and hardware infrastructures such as cloud architectures will be discussed including techniques of sparse approximation in systems measurement, and algebraic and topological statistical metrics for performance. Strategies of quantifying risk over different geometric and statistical classes of distributed systems will be examined as well as methods of tracking and coding dynamic information flows.

Chris Arney and Kate Coronges, USMA and Army Research Office

Social Networks as Basic Science for Understanding Human Dimensions of Military Operations

Over the last decade, the Department of Defense has begun to recognize the importance of human dimensions in military operations. Modeling, predicting and evaluating social processes has become critical for DoD planning and long-term goals, especially in the climate of the full spectrum of modern operations, e.g. humanitarian aid, disaster relief, foreign nation stability and security, intelligence gathering, community building, and technological infrastructure. Beyond understanding the social context of our operational landscape, we also need to understand the social processes of military teams and units that carry out these specialized and diverse missions. Research that investigates the cognitive and social dynamics that lead to wise decision making and peak performance is critical for predicting, evaluating and building successful units. The Social & Cognitive Networks program at the Army Research Office (ARO) has been initiated to direct and facilitate basic social science research to address the role of human beliefs and behaviors in group level phenomena, with a focus on team processes. A related program at ARO in Social Informatics lays the mathematical and statistical foundation to construct, analyze, and solve the social and network models for operational and doctrinal improvement. This presentation outlines the elements of these two research programs and their utilization of statistical analysis.

Clinical Session 1: Reliability

John L. Eltinge, U.S. Bureau of Labor Statistics

Prospective Application of Component and System Reliability Concepts and Methods to Analysis of Survey Participation

In the analysis of survey participation, two phenomena have some similarities to patterns studied in work with component and system reliability. For purposes of this discussion, we define “component and system reliability” broadly to involve a trajectory of events (e.g., failure and possible recovery of specified system components or of an overall system) and associated measurements of underlying component or system characteristics. The first phenomenon is the initial decision of a selected sample unit to participate in the survey. Predictors of this decision may include underlying characteristics of the unit (e.g., the size and industrial classification of a selected business; or the demographic characteristics of a selected household). In addition, the sample unit’s decision may occur after a series of attempts by the survey organization to contact the unit and to persuade it to participate. Survey methodologists wish to understand as much as possible about the ways in which the timing of the decision to participate, and related negotiations (the “trajectory of events” of interest here) are linked with the abovementioned unit characteristics, and with measurements recorded during the negotiations (e.g., indicators of the degree of reluctance to participate). The second phenomenon is the decision of a selected sample unit to stop participating in the survey after an initial decision to cooperate. For example, in a single-period survey with a total of A sections in a questionnaire, the selected unit may respond to the first B sections, but then decline to respond to the remaining A-B sections. To take another example, in a panel survey (i.e., a survey in which a sample unit is asked to provide responses in each of P periods), the unit may participate for the first K periods, but choose not to participate for the remaining P – K periods. The decision to stop participating may be associated with the unit characteristics mentioned in the preceding paragraph, or with experiences that arise during earlier parts of the survey collection (e.g., questions that are perceived to be sensitive or burdensome). For these cases, the “trajectory of events” may include the number of sections (or periods) of survey participation before the unit stops responding; measures of the quality of the responses received; and indications of question sensitivity or burden for the unit. For both phenomena, statistical issues include development of models for the abovementioned trajectories. For the second phenomenon, statistical issues also include efforts to characterize and model a prospective “recovery pattern” that results from efforts to persuade the “dropout” unit to resume responding.

Contributed Session 1

Terril Hurst, Jarom Ballantyne, Allan Mense, Raytheon Missile Systems

Building Requirements-Flow Models Using Bayesian Networks and Designed Simulation Experiments

System-level requirements must usually be decomposed into several lower-level, “derived” requirements before system- and component-level design work can proceed. In the past, various ad-hoc graphical methods have been used to visualize and articulate the set of derived requirements. Ambiguous references are often made to “probabilities” (i.e., joint, conditional, or marginal probabilities). These methods and references have proven useful to guide design decisions. However, when the time arrives to verify compliance with the requirements or to troubleshoot a non-compliant case, issues often arise. Bayesian networks offer a more rigorous method of constructing a related set of derived requirements. Initially developed during the 1980s within the artificial intelligence community, Bayesian networks have proven to be a useful tool for creating a logically consistent probability model of a set of verifiable requirements. The model consists of (1) a directed acyclic graph, (2) a set of fully defined states for each node in the graph, and (3) a conditional probability table (CPT) for each of the nodes. Probability estimates in each CPT are first made by mining data from designed simulation experiments on prior, similar systems; the CPT entries are then altered as deemed necessary to satisfy top-level requirements for the system under current design. This paper describes the method for constructing and evaluating Bayesian networks. Central to the construction and evaluation of Bayesian networks are designed simulation experiments. Examples are given to illustrate the

use of simulation experiments in constructing and reasoning about the flow between requirements within a Bayesian network. Major benefits of using Bayesian networks are reported, including the ability to analyze design margin, to allocate subsystem tolerances, and to estimate the achievable upper bound on system performance for a proposed subsystem design improvement.

David Collins, Aparna Huzurbazar, Los Alamos National Laboratory

Petri Net Models of Adversarial Scenarios in Safety and Security

Adversarial scenarios of interest to the defense and intelligence communities, such as attacks on guarded facilities, involve multiple autonomous actors operating concurrently and interactively. These scenarios cannot be modeled realistically with methods such as stochastic game theory, Markov processes, event graphs, or Bayesian networks, which assume sequential actions, serialized sample paths, or situations static in time. Petri nets, originally developed to model parallelism and concurrency in computer architectures, offer a powerful graphic tool for eliciting scenarios from experts, as well as a basis for simulating scenario outcomes. In this talk we describe how generalized stochastic Petri nets can be used for deriving statistical properties of dynamic scenarios involving any number of concurrent actors. We illustrate with an application to site security, implemented using an object-oriented framework for stochastic Petri net simulation developed using the statistical computing language R.

Eunho Yang, Yulia Baker, Pradeep Ravikumar, Genevera Allen, Zhandong Liu, University of Texas at Austin, Rice University, and Baylor College of Medicine

Mixed Graphical Models via Exponential Families

Markov Random Fields, or undirected graphical models are widely used to model high-dimensional multivariate data. Classical instances of these models, such as Gaussian Graphical and Ising Models, assume all variables arise from the same distribution. Complex data in modern settings however, from high-throughput genomics and social networking for example, often contain discrete, count, and continuous variables all measured on the same set of samples. Statistical modeling of such mixed or heterogeneous data presents an important and pressing problem in modern data analysis. Towards this, we develop a novel class of mixed graphical models by specifying that each node-conditional distribution is a member of a possibly different univariate exponential family. We discuss several instances of our class of models, and propose scalable M-estimators for recovering the underlying network structure. Simulations as well as an application to learning mixed genomic networks from next generation sequencing and mutation data demonstrate the versatility of our methods.

Contributed Session 2

Terrance D. Savitsky, Daniell Toth, and Michail Sverchkov, Bureau of Labor Statistics

Bayesian Estimation Under Informative Sampling

The Bayesian hierarchical framework provides readily estimable models for capturing complex dependence structures expressed within a population. Yet, there remains an open question about how to perform Bayesian estimation when the observed sample data are acquired under an informative design. Typically recommended methods require parameterizing the sampling design or conditional expectations of inclusion into the model, which may conflict with desired inference or disrupt estimation. We propose two new approaches that are implemented with common, nearly automated procedures employing weights that encode the sampling design (and response propensities) using first order inclusion and response probabilities for observed units. The first approach conducts an inverse inclusion probability-weighted resampling of the observed data to produce a set of pseudo-populations on which we estimate our model parameters. The process performs a Monte Carlo integration over the conditional population generating distribution, given the observed data, and accounts for uncertainty from both finite

population generation and estimation of parameters. The second approach develops a weighted adjustment to full conditional posterior distributions that incorporates sampling weights into specifications for hyperparameter statistics that depend on both the data and weights, directly, and also through non-sampled parameters. A key feature of our approaches is that they don't alter the population model parameterization. Our motivating application is composed of time-indexed, functional observations for reported employment count errors among a set of business establishments. The errors emanate from a dual reporting requirement for both the Quarterly Census of Employment and Wages (QCEW) and the Current Establishment Survey (CES). These data were collected under a stratified design. Our modeling performs unsupervised inference on the number of and memberships in clusters of parameters generating the latent time-indexed functions.

Bruce Baber (USAF, Eglin AFB, FL), Allan T. Mense (Raytheon Missile Systems), C. Shane Reese (Brigham Young University)

How Bayesian Reliability Analysis was Developed and Implemented for Production Decisions

Often in the development of complex discrete functioning systems the systems level testing is very limited at the point that significant decisions need to be made in the development process. One such point is typically the production decision to commit large resources to low rate initial production concurrent with the completion of developmental and operational testing. This situation driven by schedule and budgets introduces significant risk into the program and stresses technical management of these types of complex systems. However, often these types of systems are not designed from scratch, but utilize components and subsystems from previous programs that have extensive usage data in similar or identical environments. Most developments require extensive component and subsystem design verification testing and qualification testing across most of the environments that are expected to be encountered. A method is needed to utilize previous system development and production data and subsystem level test results combined with the systems level test data taken during development testing. This paper explores Bayesian methodology to combine different types of data into a mathematically useful result for evaluating system reliability for making these production decisions. The model presented is relatively simple, but allows combining expert opinion, previous system data, component and subsystem level testing with a limited amount of system level testing to develop a more comprehensive reliability case early in the system level test phase.

Randy Griffiths and Andy Thompson, U.S. Army Evaluation Center, Integrated Suitability Methodology Evaluation Division (ISMED) Directorate

Developing Prior Distributions to be used in Bayesian evaluation of Army Acquisition Cycle Systems

Recent interest in leveraging more test data for reliability evaluations has led to consideration of Bayesian analysis and planning for test and evaluation in Army acquisition. This paper identifies concerns with Bayesian analysis in the Army acquisition process and provides general methods for developing prior distributions that address the concerns identified. Three examples are provided to illustrate the benefits of using prior distributions in Army evaluations and how the general methods can be tailored for different test and evaluation cases.

Contributed Session 3

Shuguang Song, The Boeing Company

Multi-Criteria Decision Analysis on Aircraft Stringer Selection

Multi-Criteria Decision Analysis (MCDA) problems often involve multiple Decision Makers (DMs). In this paper, we present several decision analysis algorithms, considering both subjective and objective decision criteria with different strategies to account for uncertainty. We

address the uncertainty and availability of weights for decision criteria, and develop probability scoring for the criteria. We demonstrate an application of our method with a case study concerning aircraft stringer decisions.

Karla Hernandez and James Spall, The Johns Hopkins University Department of Applied Mathematics and Statistics and Applied Physics Laboratory

Convergence of Cyclic Stochastic Optimization and Generalizations

A common problem in applied mathematics is the minimization of a loss function $L(\theta)$ with respect to a parameter vector θ . In many practical applications, however, the loss function output is corrupted by noise; the loss function itself is unknown. As a consequence, traditional deterministic optimization algorithms are not applicable. Such situations arise in many problems, including so-called online training of neural networks, simulation-based optimization, and stochastic control. Instead, stochastic optimization methods such as stochastic gradient (SG) and simultaneous perturbation stochastic approximation (SPSA) can be used. Both methods are iterative in nature yielding a value $\hat{\theta}_k$ (the latest estimate for a minimizer of the loss function) at the k th iteration. In

this technical paper we discuss a cyclic version of stochastic optimization where only a strict subset of the parameters of θ are updated at a time while holding the remaining parameters fixed. There are many reasons we might be interested in such a setting. Briefly, such reasons include possible increase in stability, improved convergence rate, and the possibility to combine different methods for updating each parameter subset. In addition, it has also been observed that large differences in parameter magnitudes can have a significant impact on performance in methods like SG. Many methods such as the expectation maximization and k -means exhibit a similar alternating nature. Here we prove convergence of cyclic versions of SG and SPSA with a few generalizations. Future work would include analyzing the rate of convergence of such algorithms and an application to multiagent optimization.

Soumyo Moitra, CERT/Software Engineering Institute, Carnegie Mellon University

Modeling the Active and Idle Durations of Network Hosts

In this paper we analyze the distributions of active and idle durations of network hosts using flow data. Active periods are defined as those where there has been one or more flows. The analysis provides a particular perspective on network activity and is important for Situational Awareness. The distribution of the idle times is also important because it can help us estimate the probability of a host still being active after a period of idleness, analogous to survivability in reliability theory. The distribution can also be used to estimate the conditional probability of a host being active again within a time horizon given it has been idle for some length of time. We estimate these distributions and metrics from some public domain data. We discuss the implications for Situational Awareness and network inventory.

Contributed Session 4

David Hughes and Bill Thomas, Raytheon Missile Systems

Automatic Defect Searching and Categorization, Method and Application

This paper will outline a method for comparing parametric manufacturing tests statistically against one another and provide an overview of a software application that implements the described method. In the course of the normal manufacturing cycle discrete parts are tested under pre-defined test operations. With the time of the test and the serial number of the unit under test a unique parametric test result is created. This "test result" contains n number of individual measurements. Looking at the test result as its own unique data set comprising a collection of measurements, we can compare the shape of one test result to any number of

other test results to create a rank order list of similar test results. Normalizing the data points within the test results relative to their historic behavior and performing a Pearson correlation provides a method to compare the results. Given a high correlation coefficient between the failure/behavior mode of a test under review and another test we can suggest that using a similar corrective action to address the current failure/behavior mode will produce a similar result. Significantly, the normalization and correlation calculation for the entire set of test results being reviewed is completed in a single step.

Vince Pulido, Craig Lennon, Mary Anne Fields, Laura Barnes, University of Virginia and Vehicle Technology Directorate, Army Research Laboratory

Constructing a Movement Model for a Small Unit

The Autonomous Squad Member (ASM) project is a research effort to develop an intelligence that would allow a ground robot to support a dismounted unit with little human direction and intervention. Such an intelligence must make predictions of the future actions of the soldiers it supports, both in order to plan given current priorities, and in order to detect events which might signal a change in those priorities. To support this effort, we have developed a model of squad movement which allows the robot to estimate possible future positions of the squad, and the likely times of arrival at those positions, over a variety of possible unit priorities. Specifically, we apply the A* algorithm to bound the small unit's movement to a specified area based on an a priori map of the terrain, known mission constraints, and a weighted combination of costs, where the weight is derived from a possible mission context. By varying these weights, we can develop a diverse population of paths which represent the predicted position of the unit. We further model the time at which arrival at a given cell is likely to occur. The method by which we implement our model also allows for inference of mission contexts based on the actual path taken.

Matt Avery and Kelly McGinnity, Institute for Defense Analyses

Empirical Signal to Noise Ratios for Operational Tests

Statistical power is a common metric for assessing experimental designs. While this metric depends on many factors, one of the most critical is the expected effect size of relevant factors and the relative noise expected in the data. Together, these values are summarized as the signal-to-noise ratio (SNR). Software packages like JMP 10 and Design Expert use SNR as a critical component in power calculations, and by general "rule of thumb", values such as 0.5, 1, and 2 are used. However, it is not clear that these values represent the true spectrum of likely outcomes from operational test data. Due to the operational realism strived for in such testing, there are often many sources of uncontrolled variation, making it difficult to plan an appropriate test based on the SNR. In this talk, we summarize observed SNRs from a wide spectrum of operational tests and discuss how this data might be used for planning future tests using a case study approach.

Barry Bodt, Marshal Childers, Craig Lennon (U.S. Army Research Laboratory), Richard Camden (Engility Corporation), and Nicolas Hudson (Jet Propulsion Laboratory)

An Autonomous Robotic Trenching Experiment

The Robotics Collaborative Technology Alliance administered by the U.S. Army Research Laboratory includes the focus areas of Perception/Sensing and Dexterous Manipulation. In a periodic integrated research assessment of component technologies on a robotic platform, a robotic 3-degree-of-freedom arm with a 6-degree-of-freedom wrist and claw was mounted on a Dragon Runner robotic platform. Sensory information from a wrist force sensor and stereo camera pair, generating frame-to-frame visual odometry measurements, was used to estimate the state of robot-environment interaction. The task for the robot was to autonomously traverse a short (~20 ft.) course while maintaining ground penetration at a depth of 1-2 inches using arm

mounted trenching claws. The goal for the experiment was to assess performance of the autonomous trenching function over varied ground composition, topography, obstacle clutter, robot control, and arm motion. A straightforward experimental design is discussed, with encouraging findings reported toward the development of an autonomous trenching capability.

Contributed Session 5

Pamela Harris, MAJ Jarrod Shingleton, MAJ James Starling, and MAJ Christopher Thoma, USMA

Success Indicators in the USMA Advanced Core Mathematics Program

In this paper we study the relationship between Advanced Placement scores on the Calculus AB/BC exams and the success of cadets in the Advanced Core Mathematics Program (ACMP) at the United States Military Academy. Previously, to place into the ACMP, cadets were required to take and pass a summer validation exam. This created a large burden for instructors to administer and grade the placement test, but, more importantly, placed an additional requirement on new cadets' already full summer training schedules. The primary purpose of this study is to analyze whether cadets with satisfactory AP scores on the Calculus examinations should be offered admittance in the Advanced Core Mathematics Program (ACMP), without having to take the summer validation exam. We define success as achieving some form of A or B in the first course of the ACMP. Using tree-based methods and logistic regression we compared the predictive power of placement models based solely on AP scores with placement models based on AP and summer validation exam scores. Additionally, we considered models that included other metrics available at admission, such as ACT, SAT and USMA specific academic scores.

Jane Pinelis and Jared Huff, CNA Center for Naval Analyses

The Economy and Enlisted Retention in the Navy

At the end of their service obligations, sailors decide whether to reenlist in the Navy or to enter the civilian workforce. The Navy is a closed labor force: it relies solely on sailor retention to maintain its ranks and grow senior leadership. Retention forecasts affect planning of all personnel functions, like readiness, advancements, and recruiting. Generally, a decline in the economy is correlated with high retention and, during times of economic expansion, the Navy struggles to retain sailors. But, without knowing the functional form behind this relationship, drafting personnel policies or accurately budgeting for retention incentives is challenging. Using 20 years of data, we model retention as a function of the civilian economy and sailor attributes. Going beyond the unemployment rate to represent the economy, we use various statistical methods to account for multicollinearity of economic indicators, economic tipping points, and retention climate effects. Our model can be used to forecast Navy retention as a function of the civilian economy. We combine our results into useful and usable tripwires to help Navy leaders identify when retention policies need to be activated.

Vasanthan Raghavan, Qualcomm Flarion Technologies

Comparative Analysis of TAR, SEHM, and HMM Frameworks in Modeling the Activity Profile of Terrorist Groups

There has been ongoing interest in modeling the activity profile of terrorist groups over the last few decades. Pioneered by Enders and Sandler, initial work in terrorism modeling focused on time-series analysis techniques such as threshold auto-regression (TAR) models. More recent developments in this area have been along two directions. The first approach leverages a self-exciting hurdle model (SEHM) popularized in diverse applications such as seismology and gang warfare modeling. The second approach builds a hidden Markov model (HMM) framework to capture terrorist group dynamics. The focus of this work is on a comparative analysis of the

TAR, SEHM and HMM approaches to model the activity profile of different terrorist groups. While model comparisons have been done for individual terrorist groups (or across multiple groups spread over a region) in the literature, the focus here is on addressing the commonalities/divergences of the three frameworks across a large class of groups with specific proclivities (leftist, Islamist, ethno-chauvinistic, etc.). While all the three models assume that the current observation/activity in a terrorist group is dependent on the past history of the group, the models differ in how this dependence is realized. In the TAR model, the current observation is explicitly dependent on the past observations along with (possibly) the impact from other independent variables corresponding to certain geopolitical events/interventions. In the SEHM, the probability of a future attack is enhanced by the history of the group. The HMM combines both these facets by introducing a hidden state sequence. The state sequence depends explicitly on its most immediate past (one-step Markovian structure), whereas the probability of an attack is enhanced based on the state realization. Explanatory and predictive powers (of past and future attacks, respectively) of these three models are also studied and contrasted.

Contributed Session 6

Janice Hester and Laura Freeman, Institute for Defense Analyses

Applying Risk Analysis to Acceptance Testing of Combat Helmets

Acceptance testing of combat helmets presents multiple challenges that require statistically-sound solutions. For example, how should first article and lot acceptance tests treat multiple threats and measures of performance? How should these tests account for multiple helmet sizes and environmental treatments? How closely should first article testing requirements match historical or characterization test data? What government and manufacturer risks are acceptable during lot acceptance testing? Similar challenges arise when testing other components of Personal Protective Equipment and similar statistical approaches should be applied to all components. We explore these questions using operating characteristics curves and simulation studies.

Robert G. Easterling, Sandia National Laboratories (retired)

Statistical Issues in Combat Helmet Acceptance Testing

In 2007 the DoD Office of Test & Evaluation was assigned the responsibility for determining First-Article Test (FAT) plans for combat helmets. The legacy FAT penetration-resistance plan was to fire a specified projectile at five patterned locations on four helmets. This penetration-resistance test was passed if there were no penetrations in the 20 shots. In 2010 DOT&E issued a greatly expanded plan covering more helmet sizes and replications for a total of 48 helmets and 240 shots. The acceptance limit for this plan, pegged to a "90/90 standard," was 17 penetrations. Rep. Louise Slaughter (NY) challenged the new plan, concerned that it could lead to a reduced level of soldier protection. In response, DOT&E asked the National Academy of Sciences to form a panel and conduct a study of the dispute and attendant issues. This presentation will summarize the Committee's main statistical analyses, findings, and recommendations.

V. Bram Lillard and Laura Freeman, Institute for Defense Analyses

Science of Test: Improving the Efficiency and Effectiveness of DoD Test and Evaluation

The Director, Operational Test & Evaluation (DOT&E) provides oversight of DoD acquisition programs operational test and evaluation (OT&E). Among other responsibilities, the Director issues OT&E policy and guidance, certifies that operational testing is adequate, and provides an objective and rigorous assessment of operational effectiveness and suitability. By leveraging advanced methods from the statistical community we can ensure that T&E is as efficient as possible without degrading the effectiveness of T&E. In this talk I will describe several efforts that DOT&E is pursuing to ensure that we are making the most of test resources. I will provide

an overview of how we are using Design of Experiments to ensure testing is not only adequate, but as efficient as possible. Through case studies I will illustrate how DOE provides a rigorous, systematic approach designing test and an analytical trade space for determining how much testing is enough. I will also show how using statistical analysis techniques we can maximize the information obtained from test data and ensure the conclusions drawn are objective and robust.

Contributed Session 7

Hongda Zhang and Yuanzhang Li, Digital Systems Inc and Preventive Medicine Program, Walter Reed Army Institute of Research

Analysis of High-Dimensional Biomarker Data for Binary Outcomes

Identification of disease and high-risk populations provides a useful resource for studying common diseases and their component traits. Well-characterized human populations provide excellent opportunities for epidemiologists and clinical scientists to study the associations between biomarkers or genes and biological disease. Analysis of biomarkers frequently involves regression of high dimensional data. This is problematic when the number of observations is limited. The dependency among various biomarkers cannot be avoided and the collinearity makes unbiased and stable conclusions difficult. The goal of this study is to find a particular partition of the space \mathbf{X} , such that each subspace consists of all independent factors. We propose a three-step approach: 1. Decomposing the sample space; 2. Finding an orthogonal base including the most significant linear combination of biomarkers, which can be used to identify the cases most efficiently; and 3. General linear regression based on the vectors generated in step 2 to evaluate the multiple biomarker associations with the disease, and identify new cases. A new sum statistic and its distribution are developed to be used test the biomarker simultaneously for biomarker selection. We used the proposed approach to determine the potential association of serum biomarkers derived from a case-control study of schizophrenia. Numerical results demonstrate that the proposed Decomposition Gradient Regression (DGR) approach can accomplish significant dimension reduction. Information on specific biomarkers allows for potentially useful epidemiologic interpretation.

MAJ Wesley McCardle, Alongkot Ponlawat, Department of Entomology, Armed Forces Research Institute of Medical Sciences (USAMC-AFRIMS)

Modeling and Inference in Ecological Problems

Abundance (N) is a key descriptor of a central concept in ecology and epidemiology, the population. Host abundance is intrinsically linked to our knowledge of pathogen-host dynamics and the deterministic/stochastic mechanisms driving these systems. Unfortunately, individuals are always overlooked so N can never be directly observed. Classical mark-recapture methods to estimate N are well developed, however the extra-information comes at a physical and administrative cost due to individual identification of each animal. An alternative methodology uses counts from spatially and temporally replicated samples to derive estimates of N in a hierarchical modeling framework. This technique is applied to demonstrate the utility of hierarchical modeling in estimating the impact on an ecologically relevant parameter from the experimental control of the Yellow Fever mosquito (*Aedes aegypti*) using pyriproxyfen, a juvenile hormone analog.

Erin Hodgess, Visiting Scientist, National Geospatial Intelligence Agency

An R Plugin Package For Geospatial Analysis

We have produced a menu-driven package as a plugin package from the Rcmdr (Fox, 2005)

that performs geospatial analysis on spatial and spatio-temporal data sets. Kriging is used for both types of data. Then Google Earth maps are produced based on the kriging results. The spatio-temporal Google maps show the results over time. These maps can be displayed on such devices as iPads or smartphones.