Fifteenth Army Conference on Applied Statistics

A Conference of the Interface Foundation of North America

Arizona State University, Tempe Arizona

20-23 October, 2009

Hosted by ASU

Cooperating Organizations:

Raytheon Missile Systems
U.S. Army Research Office

15th ACAS
Abstracts

General Session I

Combinatorial Stochastic Processes and Bayesian Nonparametrics
Michael I. Jordan [University of California, Berkeley]

Computer Science has historically been strong on data structures and weak on inference from data, whereas Statistics has historically been weak on data structures and strong on inference from data. One way to draw on the strengths of both disciplines is to pursue the study of ``inferential methods for data structures''; i.e., methods that update probability distributions on recursively-defined objects such as trees, graphs, grammars and function calls. This is accommodated in the world of Bayesian nonparametrics, where prior and posterior distributions are allowed to be general stochastic processes. Both statistical and computational considerations lead one to certain classes of stochastic processes, and these tend to have interesting connections to combinatorics. I will focus on Bayesian nonparametric modeling based on Dirichlet processes and completely random processes, giving examples of how recursions based on these processes lead to useful models in several applied problem domains, including computational vision, statistical genetics and protein structural modeling.

Special Session I

Text Data Mining for Better Understanding of the Science and Technology Landscape
Jeffrey L. Solka, Nick Tucey, and Avory Bryant [Naval Surface Warfare Center]

This talk will provide a brief overview of text data mining along with an illustrate of the application of some of the text data mining techniques to a small corpus. Particular attention will be paid to the fruitful interaction between clustering and visualization. Various visualization schemes that allow one to interrogate cluster structure within a lower dimensional space will be illustrated. These visualization schemes provide a convenient mechanism to perform exploratory data analysis on document collections in order to reveal interesting structures within the documents meta tags.

Feature Selection for Complex Systems
George Runger [Arizona State University], Eugene Tuv [Intel Corporation]

A critical element of models for modern data sets is to select a compact subset of important variables, to better interpret, act, and predict. Modern data sets often contain a large number of variables, mixed with both numerical and categorical features, dirty and redundant data, and may contain interactive effects that require more complex models. This is a challenge for traditional variable selection. We discuss tree-based ensembles with combined serial and parallel elements that can handle the characteristics of modern data, and that provide statistical significance for selected variables and information for redundancy, and still compute quickly. Examples illustrate the method.

Approaches to Text Mining that Preserve Semantic Content
Yasmin Said [George Mason University]

Abstract unavailable.

Contributed Session I

Graph Theoretical Methods for Cluster Verification
Bernard Harris [University of Wisconsin-Madison]

To motivate this investigation, consider the following scenario. If n observations of random data in the unit square are plotted, frequently one may conclude that the data show several clusters of values. A natural question is whether these apparent clusters are "real" or simply a consequence of random variation. In some simulation experiments conducted by the author and his associates, one may easily conclude that clusters are present by looking at the plotted values. However, the data is in fact merely bivariate plots of random numbers. The results discussed in this report provide some techniques for deciding the above question.

Realizations of n independent, identically distributed k-dimensional random vectors are given. A metric ρ and a threshold parameter, $\delta > 0$ are selected.. The realizations are to be interpreted as vertices of a graph and two vertices are adjacent if the distance (ρ) between them is less than δ. Subject to some regularity assumptions on the distribution, the distributions of complete subgraphs, degrees of vertices and isolated vertices are determined. The asymptotic behavior of these distributions is studied. The efficacy of these various criteria is investigated. These methods are also useful in the study of mixtures.

Some preliminary ideas about applying such methods to data taking values in Hilbert spaces are also discussed.

Image Fusion and the Method of Paired Comparisons: Experiments, Evaluation, and Open Questions
Murray H. Loew [George Washington University]

The fusion of images is important in many DoD applications. Sensors that respond to different parts of the electromagnetic spectrum acquire images that can provide complementary information. The combination of those images thus may enable higher-accuracy detection and/or identification of objects. In general, the methods of fusion, and the uses of the fused imagery, will depend on whether they are to be interpreted by a human observer or processed by a machine. It is useful, therefore, to rank fusion methods according to how well they assist a human observer in a decision task. This work applies and compares two methods for ranking.

Two images (medium- and long-wave infrared), acquired for each of a number of outdoor scenes, were fused by each of nine methods. For each scene, a set of observers assessed each of the 36 possible pairwise combinations of fused images, choosing from each pair the one that was best for target identification. The goal then was to use that set of preferences (the responses) to rank the fusion methods (the stimuli) for their effectiveness in the identification task.

A classical and widely-used technique for ranking these "discriminal processes" is Thurstone's Law of Comparative Judgment and its implementation as the Thurstone-Mosteller (TM) Method of Paired Comparisons. To make meaningful statements about preferences, one should have a measure of uncertainty for each rank. The TM method, however, cannot readily provide such a measure. An alternative, the Bradley-Terry (BT) method, permits both a test of the hypothesis of no preference, and the calculation of confidence intervals for the ranks of the fusion methods.

We present results from a multi-observer, multi-view trial, evaluated using TM and BT. The methods yield similar results (evident, as we demonstrate, in both theory and practice). But the additional information provided by BT can have a substantial impact on the implementation of fusion in real systems. There could be substantial differences between fusion methods – in computation time or other resource requirements – that may not be exploited in the absence of those insights. We also look briefly at extensions (how to deal with ties; speed-up using incomplete designs) and some open questions (the effect of image type on choice of fusion method; how to characterize type).

Contributed Session II

Comparison of Prediction Accuracy of Surrogate Models Developed Using Nested Latin Hypercube Designs - "How Many Simulation Experiments Do I Need to Run?"
Tom Donnelly [SAS Institute Inc.]

For 20 years space-filling Latin Hypercube Designs have been used for developing fast-running surrogate models of long-running non-stochastic simulations. When simulations are computationally intensive, as can be the case with Computational Fluid Dynamics (CFD) models, it is still desired to run as few simulations as possible to generate Gaussian Process (Kriging) surrogate models to answer questions about the simulated process. Difficulties include knowing how many simulations are required to achieve the desired accuracy over an adequate volume of the factor space. This presentation will demonstrate an approach using recently developed Nested Latin Hypercube Designs (NLHD) to show how to assess when a sufficiently accurate surrogate model has been obtained. Evaluation of prediction accuracy for checkpoint trials is used to indicate when the running of additional NLHD blocks will yield limited improvement to the surrogate model.

Regression-Based Inverse Distance Weighting with Applications to Computer Experiments
V. Roshan Joseph, Lulu Kang [Georgia Institute of Technology]

Inverse distance weighting is a simple method for multivariate interpolation but has poor prediction accuracy. In this article we show that the prediction accuracy can be substantially improved by integrating it with the usual regression methods. This new predictor is quite flexible, computationally efficient, and works well in problems having high dimensions and/or large data sets. We also develop a heuristic method for constructing confidence intervals for prediction.

Special Session II

An Accumulate-toward-the-Mode Approach to Confidence Intervals and Hypothesis Tests and an Application to Binomially Distributed Data
Dwayne Nuzman [Army Materiel Systems Analysis Activity (AMSAA)]

Analyses of binomially distributed data usually depend on the normal approximation to the binomial or, for small sample sizes, the binomial cumulative distribution function. The first approach is only good for an unspecified “sufficiently large” sample size and does not reflect the asymmetry of the binomial distribution for probabilities other than 0.5. The second approach leads to suboptimal designs. Neither approach extends easily to handle multivariate binomial data. This paper describes a new accumulate-toward-the-mode approach to hypothesis tests and confidence intervals and the application of this approach to binomially distributed data. The

usual cumulative distribution function, which accumulates probability from left to right, is replaced by an accumulate-toward-the-mode distribution function which accumulates probability from areas of lower probability to areas of higher probability. This approach, when applied to asymmetrically distributed data, produces more powerful hypothesis tests and more accurate interval estimates. This in turn brings better information to decision makers from the evaluation of modeling and simulation as well as test results. In addition, this approach easily extends to analysis of multivariate distributions, providing decision makers with better information on the relative performance of alternatives. One example of this is a recently completed survivability study in which the Integrated Casualty Estimation Model was used to examine the vulnerability of soldiers with various levels of protection to fragments. The accumulate-toward-the-mode methodology was used for the analysis of these results. This briefing highlights the mathematical method as well as the study results. Results of the study have been provided to the Natick Soldier Center in support of their evaluation of a proposed change to the currently fielded helmet.

Confidence Interval Methodology for Ratio Means (CIM4RM)
John Nierwinski, Jr. [Army Materiel Systems Analysis Activity (AMSAA)]

The U.S. Army and many other government and private organizations need to evaluate ratio means to help them make informed life cycle management decisions. A ratio mean is the ratio of the mean of two random variables, X and Y, whose corresponding terms are paired. An example of a ratio mean that the Army evaluates is the maintenance ratio (MR). The Army collects MR data from a sample of vehicles within a selected system and applies CIM4RM to the sample data to construct approximate confidence intervals (CI's) for the MR.

CIM4RM is the combined effort of bootstrapping and creativity in constructing CI's around ratio means. CIM4RM was tested in many ratio mean applications and shown to be consistently valid. Prior to the development of CIM4RM, no documented tool existed that could produce consistently valid ratio mean CI's for various scenarios.

The Army is currently using this methodology to evaluate ratio means for its fielded ground and aviation systems. The Office of Inspector General for Health and Human Services is utilizing this methodology for reporting ratio mean CI's to the U.S. Congress. Other existing applications include: performance evaluations for Army test systems, evaluations of an Improvised Explosive Device detection demonstration, and hypothesis test development for many applications that compare two ratio means.

Although current applications are government centric, there are countless other areas in private industry (e.g. banking, automotive) where CIM4RM can be used for improving decision analysis.

Attack Pattern Analysis
MAJ Matthew Benigni [United States Military Academy], Reinhard Furrer [Colorado School of Mines]

Improvised Explosive Devices (IEDs) are the number one killer of coalition combat forces in the Iraq and Afghanistan Theaters of Operation. A unique characteristic of this terrain is that attacks happen almost exclusively on roads. This allows us to reduce location to one dimension and consider historical attacks to quantify periodic, spatio-temporal clusters. The end result for a set of specified routes, is a set of inhomogeneous, bivariate rate functions that aid the patrol leader in his or her route selection and/or intelligence preparation of the battlefield.

Contributed Session III

The Joint Signature of Systems with Shared Components
Francisco J. Samaniego [University of California, Davis]

The representation of a system's lifetime distribution as a function of the system's "signature" has proven to be a very useful tool in studying the reliability of coherent systems. Such representations allow one to separate the influence of a system's structure on that reliability from the influence of the stochastic behavior of the components. The goal of the present study is to consider extensions of the concept of system signatures to bivariate situations in which pairs of systems share some components and thus have dependent lifetimes. The problem explored here is motivated by examples of sharing of components in the design of selected computer networks. A "server" connected to several "slave" computers is a prototypical example. Here, we obtain representations for the joint distribution of pairs of coherent systems with shared components under the assumption all components have i.i.d. lifetimes. The expression derived for the joint distribution depends on a pair of matrices S and S*, each of which has "total mass" 1. The pair (S, S*) is referred to as the joint signature, and, under the assumption i.i.d component lifetimes, is independent of the underlying component distribution. Given two pairs of such joint systems, we study various forms of stochastic ordering among the systems' joint lifetimes. Specifically, we provide a variety of conditions on the joint signatures of the two pairs of systems, which are sufficient to ensure that the two joint lifetimes satisfy specific bivariate stochastic orderings. This work was done in collaboration with J. Navarro and N. Balakrishnan.

Overview of Modeling and Analyzing Complex System Reliability using the Los Alamos Developed Software tool, SRFYDO
Christine M. Anderson-Cook [Los Alamos National Laboratory]

The Statistical Sciences Group at Los Alamos has developed an approach to modeling system reliability leveraging information from a variety of different sources of data. Using components, testset and full-system tests an improved estimate of system reliability is generally possible. This talk presents an overview of the underlying statistical model for the analysis, discusses model assumptions, as well as demonstrates the usage of the tool. SRFYDO is a Python application using Excel spreadsheets for inputs, and is available to any government agency. It is appropriate for series systems with possibly several versions of system with some common components, multiple types of data and models reliability as a function of age and up to 2 other lifecycle covariates.

Bayesian Reliability Analysis: The Concept of Borrowing Power
David King, Charles A. Pace, and Allan T. Mense [Raytheon Missile Systems], Gary Glover and Benny Phillips [Tinker Air Force Base]

In this paper we introduce a novel Bayesian model for reliability of a missile which incorporates factors such as stress, missile phase and component. This particular model is similar in some ways to the classical Beta-Binomial model, except for the fact that a prior distribution is based upon the inverse logit of normally distributed variable whose mean is dependent upon the stress of test.

In the case that the data set includes data from two types of stresses we show that the Bayesian model borrows power when estimating the reliability for one type of testing compared with

another type. In addition we show that the model incorporates time to failure by including the variable missile phase in the model.

Bayesian Reliability Analysis: A Tool for the Analyst

Allan T. Mense, David King [Raytheon Missile Systems], Benny Phillips, Garry Glover [Tinker AFB]

For the past two years engineers at Raytheon Missile Systems and government engineers at Tinker AFB have been working to develop a Reliability approach for use on complex systems that have many phases of operation (time sequences) and which are tested under a variety of environments (modalities). A tool has been developed called RBRT and its capabilities and limitations will be outlined and discussed.

Field data from past tests are primarily binary (pass/fail) in nature and so suggest some type of logit transform to obtain an estimate of reliability (probability of a success). However logit analysis alone did not solve the variation in test modality problem. The work of Reese et al [Reese, 2007] suggested the use of a Bayesian approach using hyper-distributions for the regression parameters in a logistics regression. We have extended this work and combined it with multi-phased binary test data to produce an estimate of fleet reliability.

The design and development of this tool will be discussed including issues involving cross verification and the use of Bayesian Chi-squared procedures as a substitute for cross verification.

The tool can handle 45 components operating over 4 phases, and in several test modalities. It was written in R by Dr. David King and is being translated into MATLAB.

Reference: Reese, C. S., Hamada, M. and Robinson, D. (2005). Assessing system reliability by combining multilevel data from different test modalities. *Qual. Technol. Quant. Manag.* **2** 177--188.

Contributed Session IV

An Approach to Corroborating the Impact of Recruit Quality and Recruiting Mission on Resource Requirements

Robert Clemence, Jeremy Heusner, Robert Love, Raissa Nourieva, and Meredith Sachs [Booz Allen Hamilton, Inc.]

The Army enlists and re-enlists over 290,000 soldiers each year in its active and reserve components to carry out its mission in the near-term and to develop the mid-level and senior-level non-commissioned leaders of the future. Of this number, roughly 184,000 are new recruits. The Army prefers to recruit high-quality young men and women because of their better performance and lower attrition. However, since the late 1990s, military recruiting has been more difficult and the cost of recruiting high-quality youth, until the recent economic turn down, continued to rise. The reasons for these difficulties and these increases in cost have been attributed to a number of factors, key among them being increased college enrollment by high school graduates, low rates of unemployment and the hazards of wartime service. In the past six years, total funding for recruiting activities, for bonus incentives, and for advertising have increased almost four fold, rising from $1.41 billion in fiscal year (FY) 2002 to $5.45 billion in FY 2008 while the recruiting and retention mission has changed very little, decreasing from 303,465 to 293,935 contracts during the same period. Over forty percent of the Army's FY 2008 recruiting and retention

expenses would have been unaffordable were it not for the additional $2.25 billion provided by Global War on Terrorism (GWOT) supplemental funds that year. This paper reports on an effort initiated by the Resources Division, Plans and Resources Directorate, G-1 that applies statistical techniques to historical enlisted contract data, economic data, and polling data to develop cost estimating relationships to corroborate its future enlisted incentive bonus funding requirements.

Assessing Management and Statistical Risks in the OCONUS Immersion Experiment with Variations of Cohen's *d* (effect size) as a Tool for Decision-making.
Ray Mirikitani [Defense Language Institute]

Administrators are occasionally faced with budget decisions for training programs funded by the DOD. The management choices are usually (1) to continue funding at the current levels (2) to increase/decrease funding or (3) to suspend the project and reallocate the resources. Inevitably, the first question that administrators ask project managers is, "Does that training program really work and, if so, to what degree?" In this paper, I discuss an application of Jacob Cohen's ideas concerning statistical power and effect size, in facilitating the decision-making process.

A primary concern of administrators regarding language immersion programs conducted at the DLIFLC is whether or not these programs have a measureable impact on students' end-of-class Defense Language Proficiency Test (DLPT) scores in Listening Comprehension (LC), Reading Comprehension (RC) and Speaking (SP). That is, *quantitatively*, does the needle move in the positive direction on the achievement scale? If so, how much does the needle move? During 2007-2008, an experiment was conducted at the Defense Language Institute Foreign Language Center (DLIFLC), Monterey, CA, to determine whether or not linguist-trainees (N=207) benefited from a 4-week in-country (OCONUS) language immersion and, if so, the extent to which higher test outcomes resulted. DLIFLC students in basic language courses (Arabic, Chinese and Korean) were randomly assigned to an immersion or a control group; each comprised of 7-10 students. The immersion students were sent to host sites in Amman, Beijing, Cairo and Seoul. Control students remained in the schoolhouse. The experiment-wise effect size was found to be statistically and practically significant, while maintaining acceptable control of Type I and Type II error rates. Implications for meta-analyses are discussed.

Mine Clearing Efficiency Analysis
Rebecca E. Wentz [Army Materiel Systems Analysis Activity (AMSAA)]

This paper investigates the mine clearing effectiveness of an Army vehicle that utilizes explosives and a plow to clear a lane of all mines. Previous testing was never done using both methods together to neutralize mines. Test results were obtained for each mine clearing method when used alone, and techniques from the Maximus Method were used to combine these results to calculate the probability of clearing a lane of all mines using the two methods together. The method for calculating the mine clearing probability and subsequent lower confidence bounds involves multiple parts and steps, and these parts and steps are explored and detailed in this paper.

General Session II

Graphical Data Explorations of Very Large Databases: the ASA Data Expo 2009
Heike Hofmann [Iowa State University]

How often do we get stuck in unwanted places when we fly? This is a graphical exploration of data on commercial US flights since 1987. Graphical tools of working with very large data sets

will be discussed. Another interesting aspect of the data is the airline carriers financial situation and its impact on efficiency and customer perceived satisfaction. To that effect we will incorporate data that is not being recorded by carriers, but derived by us from publicly available sources: ghost flights, i.e. aircraft that are being moved (presumably flown) between airports while being empty.

Contributed Session V

Robust Estimation of Mixtures of Heavy-tailed Distributions
A. M. Santos [IBM and UC-Denver], Karen Kafadar, [Indiana University]

We examine the robustness and efficiency of a variant of the EM algorithm for estimating parameters of mixtures of h-distributions. Specifically, we use the biweight in an EM-like algorithm to estimate location, scale, and proportion of each component of a mixture of h-distributions, a family of long-tailed distributions that is indexed by a tail-length parameter (h=0 corresponds to the Gaussian) which provides a measure by which robustness and efficiency can be assessed. We compare our results to those using the conventional EM-algorithm (that assumes Gaussian distributions) and with Scott's L2E (2001) approach. Finally, we describe the application that motivated this research and our plans for future work.

Information Exchange Between Complex Networks
Bruce J. West [U.S. Army Research Office]

Complex networks form one of the most challenging areas of modern research overarching all the traditional scientific disciplines. The transportation networks of planes, highways and railroads; the economic networks of global finance and stock markets; the social networks of terrorism, governments, businesses and religion; the physical networks of telephones, the internet, earthquakes and climate change and the biological networks of gene regulation, the human body, clusters of neurons and food webs, all share a number of apparently universal properties as the networks become increasingly complex. Ubiquitous aspects of complex networks are the appearance of non-stationary and non-ergodic statistical processes with inverse power-law statistical distributions. In this talk we examine how information is exchanged between complex networks, where a complex network is characterized by an inverse power-law distribution (B.J. West, E.L. Geneston and P. Grigolini, “Maximizing information exchange between complex networks”, Physics Reports 468, 1-99, 2008).

We present a network of interacting two-state nodes (agents) as a dynamic model of cooperative decision making. Each agent in isolation generates a Poisson process with rate g. When the number of nodes is finite, the decision making process becomes intermittent. The decision-time distribution density for finite N is determined to be characterized by inverse power-law behavior with index $\mu=1.5$ and is exponentially truncated. This exponential generates a strong ergodicity breakdown, one stronger than that previously determined to be generated by the inverse power law without truncation. We find that the condition of perfect consensus is recovered by means of a fat tail that becomes more and more extended with increasing number of agents. The condition of partial consensus can be transmitted from one complex network to another provided that both are in a cooperative (self-organized) condition. Moreover, no significant information transmission is possible if one of the two networks is not yet self-organized. Many of the known experimental properties of synchronized networks can be deduced from this simple dynamic model.

Optimizing Computer Networks for Robustness and Efficiency
Edward J. Wegman [George Mason University], Hadi Rezazad, Orchidtech [Technology and Management], Roger W. Shores [U.S. Census Bureau]

Computers in a network are vulnerable if the computer is too central, having many connections, or too isolated, having very few connections. High centrality computers, such as servers like amazon.com or ebay.com, are vulnerable to denial-of-service attacks, which not only disrupts their functionality, but also potentially reduces throughput of the entire network because much traffic is routed through the high centrality computers. On the other hand, computers with low centrality are vulnerable to Trojan Horses and hacker breakins because they are less noticed. Because of their lack of centrality, they contribute little to the throughput of the network. Such computers are also vulnerable to insider attacks. Thus, a general strategy is to even out the node degree within the network. We have developed a recursive algorithm to optimize the network. Given the number of nodes and the number of edges, it is possible to compute the node degrees of the nodes in the limiting network a priori. The limiting network minimizes the degree standard deviation. The limiting network is not necessarily unique, but forms a subclass of networks with many similar properties. If the limiting network has a Hamiltonian cycle, we can enumerate all possible members of the class of limiting networks using a combinatorial argument. In all of the optimized limiting networks that we have examined, we have always found them to have Hamiltonian cycles, but we have no theoretical proof that this is the case.

Contributed Session VI

Stochastic Modeling and Applications
G. S. Ladde and Ling Wu [University of South Florida]

By employing data set, a stochastic dynamic model is formulated and analyzed. The results are compared with existing models and its limitations. The model is statistically tested.

A Probabilistic Approach to Extracting Symbols from Blueprints
Richard Warren and Robert S. Woodley [21st Century Systems, Inc.]

Accurate 3D models of unknown building would enable first responders to more efficiently and effectively move through the building during a crisis. Unfortunately, creating these models by hand can be a time-consuming and labor intensive process. Often, CAD (Computer-Aided Design) files can be used to automatically generate much of the building; unfortunately, for many buildings these files simply do not exist. In many cases, we may only have access to scans of the original hand-drawn blueprints.

Our goal is to extract as much information as possible from these blueprints, to minimize the time needed to create useful 3D models. To do this, we must extract a wide range of symbolic information from the drawing itself. However, there are several technical challenges that must be overcome. Blueprint symbols remain highly inconsistent. Even relatively standard symbols, like doors, can have a wide range of variations from drawing to drawing. This is especially true of the older, hand-drawn blueprints. Additionally, blueprints often contain clutter: guidelines, labels, and other stray marks. These can confuse many traditional symbol-detection algorithms. Finally, poor quality scans can introduce a large amount of noise into the image.

Established object detection algorithms, including Hu Moments and Template Matching [1] [2] [4] showed some initial success. We found these algorithms could often be tuned to produce

highly accurate results on a small set of blueprints. Unfortunately, they also proved to be far too brittle. Despite attempts to improve their performance, none of them produced a general solution that worked across a wide range of possible blueprints. Furthermore, many of these techniques could not compensate for the noise and clutter in the scanned blueprints.

Instead, we turned to a probabilistic approach to explicitly manage these underlying uncertainties. Here we used a two-step process. First, we create a function approximator that can extract a wide range of features for each pixel in the blueprint, and then calculate the probability that the pixel belongs to the desired symbol [5] [6]. This provides us with a separate probability estimate for each pixel. Noise and clutter may reduce the probability across a small portion of the symbol, but often the rest of the symbol remains unchanged. Next, we combine evidence from all pixels over a given line or curve segment. This helps both filter out the effect of noise and removes intermittent false alarms.

While our initial efforts focused on detecting doors, this algorithm represents a general approach that can be easily retrained to detect other, arbitrary symbols. Furthermore, our initial results showed robust results across a wide range of blueprint styles, including noisy, cluttered line drawings.

A Control-Stopping Differential Game for Stochastic Systems with Memory
Mou-Hsiung (Harry) Chang [U.S. Army Research Office]

We consider a stochastic differential game that consists of one controller (the minimizer) and one stopper (the maximizer). The state equation of the game is described by a stochastic functional differential equation with a bounded memory. The objective of the controller is choose an admissible control that minimizes the expected performance functional over a finite time horizon and the objective of the stopper is to seeks a stopping (game terminating) time that maximizes the same functional. In this paper, we show that the value of the game exists and is the unique viscosity solution of an infinite dimensional Hamilton-Jacobi-Issac equation.

Contributed Session VII

Statistical Issues in the Comparison of Multi-dimensional Profiles
Karen Kafadar [Indiana University]

Several common problems arise in a collection or large database of profiles: (1) Determine the number of features are needed to characterize a multi-dimensional profile; (2) Estimate the "false match" probability (and its uncertainty) without resorting to a comparison of all pairs of profiles; (3) Design a sequential sample to achieve (1) and (2) that accommodates an increasing database of profiles. These issues arise in fraud detection (identify behavior that differs substantially from a customer's typical profile), pattern comparisons (e.g., fingerprints, bite marks) and genetic studies (e.g., microarray experiments, spectra from proteomics experiments to identify proteins, DNA profiles). In this talk, I will describe scenarios in which these issues arise, and propose some possibilities for addressing them.

Conditional Logistic Modeling for Unbalanced Nested Case Control Studies with Bias Correction

Yuanzhang Li and David Cowan [Walter Reed Army Institute of Research], Robert Yolken [Johns Hopkins University], Natalya Weber and David Niebuhr [Walter Reed Army Institute of Research]

Bias is the incorrect assessment of the association between an exposure and an effect in the target population in which the estimated expectation does not equal the true value. This happens in epidemiological case-control studies, where a possible risk factor is compared between cases and controls to investigate the nature of the disease. The most common biases are those produced in the definition, selection of the study sample, data measurements and variable confounding. If the data is unbalanced, the estimation from a conditional logistic model is usually biased. Diverse analytical approaches have been proposed to evaluate the association between the disease and exposure. We use simulation to estimate the bias and show how the bias could be adjusted by data normalization, matching imputation and modeling with simulation.

Data Used: The U.S. military collects serum samples that are obtained at accession and approximately every two years thereafter as long as personnel remain on active duty. We identified subjects hospitalized and subsequently discharged from the military with a diagnosis of schizophrenia. Based on demographic and service characteristics, 155 female cases were matched to 3 controls, and, 700 male cases to one control each. For each study subject we obtained the first available and up to seven additional serum specimens per individual for cases and up to four specimens matched on the time of case drawing per individual for controls. The data are unbalanced between individuals and between groups.

METHODS: We developed statistical modeling strategies for case-control studies to evaluate the association of antibodies on risk of schizophrenia. Data normalization, match imputation, and weighted conditional logistic model with bias correction by simulation, was used for obtaining the potential 'unbiased' estimations for the exposure effect and their interaction with the initial sample. To study the initial antibody effect for identifying individuals at early stage, cases and their matched controls were grouped into categories by 50th and 75th percentile exposure level of the first sample of the cases.

RESULTS: For unbalanced data in a case control study, the estimation of the exposure effect from conditional logistic regression model is usually biased. The bias ratio could be as high as ten folds or higher depending on the data structure. The bias magnitude is usually reduced by a suitable data normalization and data match imputation. The bias is 'corrected' by simulation in the proposed modeling process to get the potential unbiased estimation. In addition, the distribution of bias is evaluated, which offer the valuable information to adjust the bias.

DISCUSSION: We have conducted seven agents for this nested case control study data, without adjusting biases, the conditional logistic model generated the biased estimations for the unbalanced data. Data normalization and match imputation can reduce the biases in general. After adjusting bias with simulation, we obtained a relatively unbiased estimation. The proposed approach can help investigators to find which agent could be used to identify the high risk population for a certain disease earlier. Without adjusting the bias, this estimation is usually biased. The proposed approach could be extended for any kind of case control studies. We selected normal as the bias distribution for the data used in this study; however the distribution may vary from data to data, hence different adjustment techniques should be considered. Finally, since biases arise from variety of sources, we study the biases generated from inconsistent data measurements, unbalanced data structure and confounding in the modeling.

Contributed Session VIII

General Orthogonal Solutions in Multivariate Regression
Jerry Alderman and William Carl Thomas [Raytheon Missile Systems]

A MATLAB based orthogonal multivariate regression model, GOS, will be introduced. The model uses qr decomposition to establish an orthogonal basis for regression. Some issues with the degeneracy of the qr decomposition will be discussed and resolved and it will be shown that the model is capable of resolving highly co-linear datasets. The model will be applied to the famous Longley data set and it will be shown that it reproduces the NIST standard solution for this dataset to 15 digits. An application to simulation data with 1 response, 66 input variables and 1000 sets of input and response data will also be discussed. Some emphasis will be placed on explaining how statistically significant model terms are identified and how model coefficients are mapped from the orthogonal data representation back to the original or natural set of input variables.

Variable Selection in Nonparametric Regression with Very Many Noise Variables
Wei-Yin Loh [University of Wisconsin-Madison]

The prediction accuracy of nonparametric regression models typically deteriorates quite rapidly as the number of noise variables is greatly increased. This talk highlights the severity of the problem and proposes a simple strategy, based on the GUIDE (Loh) regression tree algorithm, for slowing the deterioration rate by performing prior variable selection. The effectiveness of the method compared to the EARTH (Doksum et al.) algorithm on the resulting prediction accuracy of MARS (Friedman), Random forest (Breiman), and GUIDE itself is evaluated by simulation and with real data sets.

Contributed Session IX

Power Electronics System Compatibility Assessment Using Statistical Design Methods
Sydney Lofton and William C. Thomas [Raytheon Missile Systems]

Today’s electronic systems increasingly demand high performance power systems. Extreme Value Analysis (EVA) leads to over-design and is a significant cost driver. Statistical Design Methods (SDM) is an effective replacement for EVA, leading to a more optimal design solution. SDM is applied to evaluate a proposed power supply design and determine the required power rating, assess system compatibility, predict yield and quantify risk. A novel, easy method to simply define the required power is presented along with a user friendly design/analysis tool. The statistical analysis provides insight into the sensitivities and relative impact of the parameters affecting the response. Several benefits include identification of potential Key Performance Characteristics (KPC).

Key words: Statistical Design Methods (SDM), Extreme Value Analysis (EVA), Power System Design.

A Computer-Aided D-Optimal Design for Improving Ballistic Performance
David W. Webb & Brian M. Powers [U.S. Army Research Laboratory]

Many studies have shown that the ballistic performance of a ceramic material is affected by its surrounding structure. To quantify this effect, a numerical model of a ballistic impact on a ceramic tile supported by a metal backplate was created in the commercial finite element code Autodyn. All aspects of the model were held constant except for the material properties of the backplate. A bilinear elastic-plastic relationship was used for the material properties of the backplate. The material properties which served as factors for this study included Young's modulus, Poisson ration, failure strain, yield stress, and ultimate stress.

Because of sample size limitations and a constraint involving the stresses, a computer-assisted approach was utilized to select the design. The specific levels of the five factors were chosen via a genetic algorithm which evaluated candidate designs using a D-optimal criteria.

Two metrics were selected for evaluating the ballistic performance. First, the depth-of-penetration and mass-erosion of the projectile were tabulated over time and cross plotted. Then the area under this curve was calculated and used to evaluate the simulation results. Second, the momentum of the projectile at a fixed time was used to evaluate the simulations.

This presentation will focus on some of the challenges experienced in the design phase of the study, and the results of the ensuing analyses.

Contributed Session X

Length Bias in the Measurements of Carbon Nanotubes
Paul Kvam, [Georgia Tech]

Abstract unavailable

Laser Scanner Certification Study for Body Armor Back Face Deformation Measurement
Barbara J. Gillich, W. Scott Walton, Craig Miser, Adam Fournier, Jeffrey Hosto, Jeffrey Huber, Ward Boughers, and Craig Andres [Aberdeen Test Center]

Back face deformation for body armor testing is a critical measurement in the evaluation of a system's effectiveness against ballistic threats. The current methodology involves placing the body armor system against ballistic clay, impacting the body armor with a specified threat traveling at a specified velocity, and measuring the impression left in the clay from the body armor deforming to stop the threat. The maximum depth of the impression is used to evaluate a body armor system's ability to mitigate blunt force trauma to the wearer. Currently, a digital caliper system is used to measure the maximum depth of the back face deformation impression.

New technology, employing a laser scanning device and associated software, has become available that will increase the precision of this measurement as well as provide information on any point within the back face signature relative to the original undisturbed surface. Additionally the scanned information can be archived and retrieved long after testing has been completed. Although laser scanning technology is currently being used at U.S. military, government, and aerospace installations for other various metrology applications, this is the first known application of the laser scanner technique for back face deformation measurements. This study was

conducted to characterize the measurement uncertainty associated with both the digital caliper and the laser scanner techniques.

Special Session III

Beyond Correlation: Examining Causality with Agent-based Hypothesis Instantiation
Gary An, MD [Northwestern University]

The greatest challenge facing the biomedical research community is the effective translation of basic mechanistic knowledge into clinically effective therapeutics. In today's high throughput, data-rich environment technology has been able to augment one aspect of the scientific process via the identification of potential patterns of *correlation* within data. However, this represents only one portion of science; the missing essential step is the conversion of a correlative pattern to a hypothetical causal mechanism. The same issues regarding the magnitude of available data challenge traditional experimental methods in establishing causality; the solution lies in harnessing advances in computational power and methods to allow researchers to dynamically instantiate their hypotheses, providing a progressive means of moving from assessment of plausibility to verification. To date there has been sparse work on technological enhancement of the intuitive processes of hypothesis evaluation and verification. These "discovery" steps can be augmented via the instantiation of thought experiments in a "virtual sandbox" where researchers can both formally represent their hypotheses and instantiate them dynamically such that the behavioral consequences of their beliefs can be examined. This talk introduces an Agent-based modeling framework (ABMF) to concatenate ontological descriptions of components and functions, and facilitate the development of an executable layer of knowledge representation. The representation of a conceptual model in the ABMF provides a knowledge structure that can then be transformed into executable models. As an example, ABMF is used to represent an existing multi-scale model of systemic inflammation from the level of intracellular signaling pathways to clinically recognizable phenomenon. The intent is not to suggest that an agent-based framework is "the" format for an executable ontology layer; it is fully expected that other researches will develop similar types of tools with different modeling paradigms. However, ABMF represents a robust, evolvable approach, and will provide an essential translational step in the technological augmentation of the discovery process. Eventually, this type of architecture would allow the automated generation of models that can be picked through by the researcher, providing a truly evolutionary solution to the synthetic aspect of biomedical research.

Measures of Tipping Points, Robustness, and Path Dependence
Aaron Bramson [University of Michigan]

Abstract unavailable

Combining ABM and Behavioral Experiments
Marco Janssen [Arizona State University]

In this talk I will present recent work in combining agent-based models (ABMs) and behavioral experiments to study collective action problems. Our interest is to understand the conditions in which groups are able to self-organize in sharing common resources or provide public goods. One of the challenges to derive empirically grounded agent-based models is to collect relevant behavioral data at individual and group levels. We conduct field and laboratory experiments to investigate our theoretical frameworks and derive detailed observations. ABMs are one of the tools to analyze the data in more detail. Due to the emerging nature of social phenomena there is a methodological challenge to determine how well ABMs are good representations of the

empirical observations at different levels of scale. We will illustrate recent progress to address this challenge, especially using the concept of pattern oriented modeling.

General Session III

Modern Experimental Design Methods
Douglas C. Montgomery [Arizona State University]

Statistical experimental design methods have had significant impact on the scientific, business and industrial world. This impact has expanded considerably over the last 15 years, driven by a number of factors, including software innovations and the widespread adoption of six sigma as a process improvement strategy. This presentation surveys recent developments in the field that are anticipated to further stimulate the use of designed experiments, including the use of optimal designs, non-regular fractional factorials, and experiments for deterministic computer experiments. Examples from various fields including illustrated. Some personal experiences with designed experiments are included.

Contributed Session XI

Predicting Linguists' Training Outcomes with Linear Regression Models Constructed Using a Single and Double Cross-validation Design
Raymond T. Mirikitani, Defense Language Institute

Administrative databases often contain extant data that are organized for specific report generation or personnel-processing purposes as opposed to research purposes. Efficiency and speed-of-processing tend to be over-riding concerns for database administrators and accurate data retrieval may suffer. In addition, the data are often not structured in a manner that allows for answering the specific research questions being asked. Nevertheless, when a commitment is made to reorganize an administrative database for the purpose of facilitating research, a win-win situation can obtain.

In this paper, I describe empirical research conducted at the request of the Defense Language Institute, Washington D.C. (DLIW) to statistically model the language training of career military linguists. The research was initiated to develop metrics for assessing foreign language training outcomes of students, as measured by their end-of-course Defense Language Proficiency (DLPT) test scores; (1) to facilitate the process of establishing course outcome objectives for language trainees and (2) to improve the methods for measuring progress in meeting course objectives. Students were enrolled in basic, abbreviated basic and refresher courses offered through contract language schools. The 64 target languages ranged from low difficulty, Category I (Spanish, French) to high difficulty, Category IV (Arabic, Chinese). Using correlation and linear stepwise regression procedures, length of training, language aptitude, prior language-training experience and target language difficulty were found to be significant predictors of test outcomes in listening, reading and speaking. Prediction models were developed and cross-validated using administrative data for DLIW students graduating between 2003 and 2008.

Oh Say Can You GEE: Analysis of Visual Acuity Data
Robyn Lee, Brian Hatch, David Hilber, James Elledge, and James Stout, [U.S. Army Center for Health Promotion and Preventive Medicine]

A randomized, single-blind, repeated measures study was conducted to evaluate the effect of visual acuity on target discrimination and marksmanship. Subjects fired at randomized target presentations from 50 to 300 meters while wearing customized spectacles to yield the five visual acuity levels being evaluated. Randomized targets marked as friend or enemy were presented in a similar manner to determine target discrimination for each of the five visual acuity levels. A generalized estimating equations (GEE) approach was used for analyze of this longitudinal data to explore the effects of visual acuity on marksmanship and target discrimination while accounting for the covariates of range distance and other range configurations present in the study.

General Session IV

Spinning Heads and Spinning News: the Use and Abuse of Statistics in the Media
Rebecca Goldin, [George Mason University]

News increasingly depends on a careful dissection of numbers. Statistics are everywhere, from how many people are not covered by health insurance to whether Vitamin E is good for you or not. Yet for being so prevalent, statistics are badly understood by the general public.

In this talk, I'll illustrate how the press often misuses statistics with examples from recent coverage. Since news sources are the main avenue by which the public understands many public health issues, these misguided representations of science can actually shape public policy, legislation, and individual choices. We will see why it is so important that media writers understand basic concepts from statistics, epidemiology and the scientific method.

I will also show how powerful the work can be when the press goes beyond politics and morality to get the science right.

These examples come from my experience as the research director for Statistical Assessment Service (STATS), a nonprofit media education and watchdog group affiliated with George Mason University, where I am a professor of mathematics as well. STATS takes critical aim at the poor use of statistics to justify false claims or to back-up ideological agendas, while serving as a resource for journalists and producers who want to engage in high-level responsible reporting that takes into account what the science says, what it doesn't, and what it can't.