ABSTRACTS FOR PLENARY, INVITED AND CONTRIBUTED TALKS[1]

(SRC 2013 — UCLA)

——— DAY 1: THURSDAY, JUNE 20 ———

—— Thursday, June 20, 9:00-10:00 ——

Plenary Session 1. Chair: Peter Z. Qian (University of Wisconsin)

From Big Data to Big Statistics
John Sall (SAS Institute Inc.)

When you scale up the analysis, you have a lot of issues to address. When you have a lot of data, even a small difference is significant. When you screen a lot of hypotheses, adjusting for selection or multiple test bias is an issue. When you have a lot of bad data, making the analysis automatically robust becomes importance. When you have big data, you need to make the computer work fast to get the job done. When you have thousands of results, you need to create compact summaries to show you all the results in one page, or at least produce the results sorted by significance. All these issues need to be resolved and the solutions encapsulated into a workflow for engineers and scientists that deal with more data each year.

—— Thursday, June 20, 10:30-12:00 ——

Invited Session 1. Journal of Quality of Technology Invited Session
Organizer and Chair: Brad Jones, JMP Division, SAS Institute Inc.

Definitive Screening Designs with Added Two-Level Categorical Factors
Christopher J. Nachtsheim, University of Minnesota

Recently, Jones and Nachtsheim (2011) proposed a new class of designs called definitive screening designs (DSDs). These designs have three levels, provide estimates of main effects that are unbiased by any second-order effect, require only one more than twice as many runs as there are factors, and avoid confounding of any pair of second order effects. For designs having six factors or more, these designs project to efficient response surface designs with three or fewer factors. A limitation of these designs is that all factors must be quantitative. In this paper we develop column-augmented DSDs that can accommodate any number of two-level qualitative factors using two methods. The DSD-augment method provides highly efficient designs that are still definitive in the sense that the estimates of all main effects continue to be unbiased by any active second-order effects. An alternative procedure, the ORTH-augment approach, leads to designs that are orthogonal linear main effects plans; however some partial aliasing between main effects and interactions involving the categorical factors is present.

A GLR Control Chart for Monitoring the Mean Vector of a Multivariate Normal Process

[1] In order of presentation time

Marion R. Reynolds, Virginia Tech

This talk develops a statistical process control (SPC) chart based on a generalized likelihood ratio (GLR) statistic to monitor the mean vector of a multivariate normal process. The performance of the GLR chart is compared with the performance of several alternative charts. Unlike most of the other options, the GLR chart does not require specification of tuning-parameter values by the user. The GLR chart also has the advantage in process diagnostics: at the time of a signal, estimates of change-point and out-of-control mean vector are immediately available to the user. All these advantages of the GLR chart make it a favorable option for practitioners. For the design of the GLR chart, a series of easy-to-use equations are provided to users for calculating the control limit to achieve a desired in-control performance.

Invited Session 2. Statistics in Cognition and Vision
Organizer and Chair: Yingnian Wu, UCLA Department of Statistics

Bayesian Analogy with Relational Transformations
Hongjing Lu, Departments of Psychology and Statistics, UCLA

How can humans acquire relational representations that enable analogical inference and other forms of high-level reasoning? Using comparative relations as a model domain, we explore the possibility that bottom-up learning mechanisms applied to objects coded as feature vectors can yield representations of relations sufficient to solve analogy problems. We introduce Bayesian analogy with relational transformations (BART) and apply the model to the task of learning first-order comparative relations (e.g., larger, smaller, fiercer, meeker) from a set of animal pairs. Inputs are coded by vectors of continuous valued features, based either on human magnitude ratings, or outputs of the topics model. Bootstrapping from empirical priors, the model is able to induce first-order relations represented as probabilistic weight distributions, even when given positive examples only. These learned representations allow classification of novel instantiations of the relations and yield a symbolic distance effect of the sort obtained with both humans and other primates. BART then transforms its learned weight distributions by importance-guided mapping, thereby placing distinct dimensions into correspondence. These transformed representations allow BART to reliably solve 4-term analogies (e.g., larger:smaller::fiercer:meeker), a type of reasoning that is arguably specific to humans. Our results provide a proof-of-concept that structured analogies can be solved with representations induced from unstructured feature vectors by mechanisms that operate in a largely bottom-up fashion. I will discuss potential implications for algorithmic and neural models of relational thinking, as well as for the evolution of abstract thought.

Perturb-and-MAP Random Fields: The Interplay between Random Sampling and Optimization, with Applications in Computer Vision
George Papandreou, Department of Statistics, UCLA

Machine learning plays an increasingly important role in computer vision, allowing us to build complex vision systems that better capture the properties of images. Probabilistic Bayesian methods such as Markov random fields are well suited for describing ambiguous images and videos, providing us with the natural conceptual framework for representing the uncertainty in interpreting them and automatically learning model parameters from training data. However, Bayesian techniques pose significant computational challenges in computer vision applications and alternative deterministic energy minimization techniques are often preferred in practice.

I will present a new computationally efficient probabilistic random field model, which can be best described as a "Perturb-and-MAP" generative process: We obtain a random sample from the whole field at once by first injecting noise into the system's energy function, then solving an optimization problem to find the least energy configuration of the perturbed system. With Perturb-and-MAP random fields we thus turn powerful deterministic energy minimization methods into efficient probabilistic random sampling algorithms that bypass costly Markov-chain Monte-Carlo (MCMC) and can generate in a fraction of a second independent random samples from mega-pixel sized images. I will discuss how the Perturb-and-MAP model relates to the standard Gibbs MRF and how it can be used in conjunction with other approximate Bayesian computation techniques. I will illustrate these ideas with applications in image inpainting and deblurring, image segmentation, and scene labeling, showing how the Perturb-and-MAP model makes large-scale Bayesian inference computationally tractable for challenging computer vision problems.

How Preschoolers' Reasoning Can Inform Scientific Causal Discovery
Patricia Cheng, Department of Psychology, UCLA

Scientists' concern with objectivity has led to the dominance of associative statistics, which define the basic concept of independence on observations. In this talk we show that to infer causation, the relevant concept of independence is causal invariance: a causal mechanism remaining unchanged across contexts. Our analysis reveals that to infer causes of a binary outcome, (1) the associative definition of independence results in a logical inconsistency – even for data from an ideal experiment – for both frequentist and Bayesian statistics, and (2) removing the logical incoherence requires defining independence on strictly imaginary causal events. We encapsulated the distinction between the two rather abstract definitions of independence in a simple scenario involving the effects of two treatments. The associative and causal definitions yield opposite recommendations regarding which treatment best achieves a desired outcome. The distinction between the definitions is so basic that the recommendations remain divergent no matter how large the sample size, and hold for either frequentist or Bayesian measures. We posed the scenario as a problem to preschool children in story form. The rationality of the correct answer is so compelling that even preschoolers, in agreement with adults, choose the treatment according to the logically coherent causal definition over that according to the incoherent associative definition. The preschoolers' decision cannot be explained by heuristic shortcuts. Our theoretical and empirical findings together indicate that the causal definition is adaptive, and suggest that introducing a new causal statistics would result in more consistent and generalizable causal discoveries in medicine and other sciences.

Contributed Session 1. Capacity Optimization and Bayesian Methods

Nonlinear Mixed Effects Models as a Basis for Capacity Optimization in the Processing Industry: a Case Study
Bart De Ketelaere, Faculty of Bioscience Engineering, Katholieke Universiteit Leuven, Belgium

Capacity optimization is a crucial aspect in modern manufacturing and requires to find a balance between (1) operating at high speeds lowering production time but increasing the risk of nonconforming products or machine downtime, and (2) using lower production speeds that might improve the uptime of the machine but requires more processing time and, thus, money. In any case, it is important that the process under study is well understood so that data driven decisions can be made. In highly variable environments where data are sparse and potentially unequally sampled building such a model is not evident.

In this contribution, we will show the benefit of (nonlinear) mixed effects models for model building, and will apply the concept to a real life case from the wire processing industry where production is based on runs (cycles). Each cycle is characterized by a given speed of operation and this speed has an influence on the machine temperature increase during processing the run. If the operating temperature exceeds a fixed threshold, the machine has to be stopped which is highly undesirable. After stopping the machine, the operator has to wait long enough so that the temperature of the machine is low enough to start up a new cycle. The essence now is to determine the maximal operating speed of the machine before the cycle starts, given the starting temperature recorded.

Simulation Comparison of EVOP and Simplex–Type Optimization Methods
Koen Rutten, Faculty of Bioscience Engineering, Katholieke Universiteit Leuven, Belgium

In industry, running a process at optimal settings is a crucial aspect for delivering optimal quality to the customer. However, for a broad range of processes it is not only important to find this optimum at a given time, but also to adjust inputs to keep track of this optimum when external (often unknown) disturbances are present. In the 1950's G.E.P. Box proposed a technique called Evolutionary Operation (EVOP) that uses small input perturbations to indicate the direction of a process optimum during production. In later years, Simplex optimization was proposed as an alternative to EVOP. The EVOP and Simplex schemes have been developed in the classical situation at that time – sparse sampling of process inputs and outputs and limited computational power to derive information. This situation has changes drastically during the last decades, and although EVOP and related methods did not attract large interest, we feel that it can bring substantial added value in the data rich world we live. From that perspective, this contribution aims at comparing EVOP and Simplex for a wide variety of cases, including the influence of noise level, dimensionality and step size. To our best knowledge, this is the first time those methods are compared to each other in a systematic way. Joint work with J. De Baerdemaeker and B. De Ketelaere

A Simple Bayesian Decision-Theoretic Design for Dose Finding Trials
Shenghua (Kelly) Fan, Dept. of Statistics and Biostatistics, California State University, East Bay

A flexible and simple Bayesian decision-theoretic design for dose finding trials is proposed in this paper. In order to reduce the computational complexity, we adopt a working model with conjugate priors, which is flexible to fit all monotonic dose-toxicity curves and produces analytic posterior distributions. We also discuss how to use a proper utility function to reflect the interest of the trial. Patients are allocated based on not only the utility function but also the chosen dose selection rule. The most popular dose selection rule is the one-step-look-ahead (OSLA), which selects the best so far dose. A more complicated rule such as two-step-look-ahead (TSLA) is theoretically more efficient than OSLA only when the required distributional assumptions are met, which is, however often not the case in practice.We carry out extensive simulation studies to evaluate these two dose selection rules and have found that OSLA is often more efficient than TSLA under the proposed Bayesian structure. Moreover, our simulation results show that the proposed Bayesian method performs superior to several popular Bayesian methods and the negative impact of prior mis-specification can be predicted and managed in the design stage.

Bayesian Family Factor Models for Multiple Outcomes
Qiaolin Chen, Department of Biostatistics, UCLA

The UCLA Neurocognitive Family Study is a family study collected neurocognitive performance measures and MRI measures on relatives of schizophrenia patients and relatives of matched control subjects in order to study the transmission of vulnerability factors for schizophrenia. There are two

types of correlations: the correlations among measurements on individuals from the same family and the correlations among outcome measurements within subjects. Standard analysis techniques for multiple outcomes do not take into account associations among members in a family, while standard analyses of familial data usually model outcomes separately and do not provide information about the relationship among outcomes. I construct new Bayesian Family Factor Models (BFFMs), which apply Bayesian inferences to confirmatory factor analysis (CFA) models. The basic assumption is that both family member type (that is, proband, sibling, father or mother) and outcome measurement type account for variations in the observed data. I consider analyzing the data using confirmatory factor analysis (CFA) with two independent groups of factors: (i) four inter-correlated family-member-specific factors and (ii) inter-correlated outcome-specific factors, one for each outcome. However, classical non-Bayesian CFA approach to this kind of data usually has the problem of empirical unidentifiability. Therefore, I incorporate Bayesian inference into confirmatory factor analysis models with inter-correlated family-member factors and inter-correlated outcome factors to construct the new Bayesian Family Factor Models. Results of model fitting on synthetic data show that the Bayesian factor analysis model can reasonably estimate parameters and that it works as well as the classic confirmatory factor analysis model using SPSS AMOS, while being able to handle missing data. The developed methods will be generally useful in analyzing family data with multiple outcomes.

—— Thursday, June 20, 13:30-15:00 ——

Invited Session 3. Designing Efficient Discrete Choice Experiments for Estimating Interactions

Organizer and Chair: Christopher J. Nachtsheim, University of Minnesota

Conjoint Analysis and Discrete Choice Experiments for Quality Improvement

William Li, Carlson School of Management, University of Minnesota

Conjoint analysis and discrete choice experiments, which were developed in fields such as marketing and economics, are useful for understanding the voice of the customer to guide quality-improvement efforts. Unfortunately, these methods have received relatively little attention in the quality area. In this article, we provide some guidelines for the use of conjoint analysis and discrete choice experiments. We discuss what they are, why they are useful methodologies for quality improvement, and how a discrete choice experiment can be carried out. We demonstrate the methodology by discussing a real case study in quality improvement in detail. We then introduce a new class of designs for discrete choice experiments that are robust for a class of possible models. We provide several examples in which an optimal design based on the main-effects only models is shown to have limited capability for estimation of two-factor interactions, whereas the proposed robust designs perform well in the presence of two-factor interactions.

Application of Blocked Fractional Factorial Designs for Discrete Choice Experiments

Jessica Jaynes, Department of Statistics, University of California, Los Angeles

Discrete choice experiment is an attribute based method that gives further insight into how individuals develop preferences for particular attributes. They are used in the traditional areas in health economics, transportation, marketing, and increasingly beyond these areas. However, there is limited work on designs for such studies to date. Motivated by the need for smaller optimal dis-

crete choice experiments, we propose a novel application of blocked factorial designs for designing choice sets for estimating main effects and main effects plus some two-factor interactions, with 100% efficiency. Our method produces smaller numbers of choice sets thereby making discrete choice experiments easier and cheaper to administer. In addition, our designs can have different levels of resolution capable of estimating all main effects clearly and unbiasedly, and in many cases, can also estimate some two-factor interactions clearly as well. We illustrate our approach using a consumer preference study for trans-Atlantic flights and demonstrate how these techniques can be extended for three-level attributes as well as asymmetric attributes.

Maximin Model Robust Discrete Choice Experiments
Anna Errore, University of Palermo, Italy, and University of Minnesota

Discrete choice experiments have their genesis in conjoint analysis, which was based on first-order models. For this and other reasons, discrete choice experiments in practice are usually designed for estimation of main effects. In this paper, we explore the construction of maximin model robust designs when the experimenter is concerned about the presence of interactions. We consider three classes of models—main effects models, main effects models plus first-order interactions, and second-order models, and we construct designs that maximize the minimum efficiency of the design for the three competing models. We do so first with standard linear models and then extend our analysis to nonlinear models and, in particular, to models for discrete choice experiments. We compare our results with existing approaches in the literature, such as Li et al., 2013.

Invited Session 4. Critical Role and Impact of Statistics in Preclinical and Nonclinical Pharmaceutical Research
Organizer and Chair: Lei Zhou, Amgen

Statistical problems in Carcinogenicity studies
Lanju Zhang and Lei Shu, AbbVie*

Carcinogenicity studies are required by regulatory agencies in approval submissions to predict carcinogenic effect on human of pharmaceutical products administered six months or longer. Two studies in two different species (usually mouse and rat) are needed for a 2-year period. The long duration, high cost (millions of dollars per study), and extreme complexities of such studies have commanded significant regulatory and industry attention and resources to improve study design, data analysis and result interpretation. In this talk, we will present some challenging problems from statistical perspective (high dose selection, sample size, group comparison and trend test, multiplicity issue, and use of historical controls) and share some of our experiences with regulatory agencies.

Discovery and Characterization of Novel Synergistic Combinations Through In-vitro Cell Line Profiling
Cheng Su and Sean Caenepeel (Amgen, Inc.)*

In vitro combination screens in cancer cell lines provide an opportunity to discover novel synergistic interactions and gain insights into the mechanism behind these synergies. Challenges in this discovery process include variability in the experimental system, variation in the level of single drug response, different reference models and methods for defining additivity, interpretation of results and follow-up strategies to translate findings to the in vivo and clinical settings. We compared different analysis methodologies, including Bliss Independence and Loewe Additivity, and two different activity definitions, illustrating how they led to different emphasis on anti-proliferative or cell

killing effects. Other important practical considerations will be discussed, including quality control, data representation and visualization, and strategies for following-up on in vitro findings.

Multiple-response process optimization using process capability: the pharmaceutical industry design space story
John Peterson, GlaxoSmith Kline

Since the beginning of the 21st century, pharmaceutical regulators have been stressing the importance of designing quality into the development of drug manufacturing processes. A key concept that has arisen, and is documented in the ICH Q8 Guidance for Pharmaceutical Development, is "design space". ICH Q8 design space is defined as the "The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality". In other words, design space is the collection of possible manufacturing recipes that are associated with a satisfactory likelihood of producing product that meets all of its critical quality specifications. In industrial engineering terms, this can be thought of as a set of conditions for good process capability.

It is interesting that the seemingly simple and desirable goal of creating a good ICH Q8 design space, is actually fraught with statistical challenges that are conceptual, theoretical, and practical in nature. This talk will discuss these challenges and outline strategies for coping with them.

Contributed Session 2. Environmental Statistics

Statistical Analysis and Modeling of Atmospheric Gases
Rebecca Wooten, Department of Mathematics and Statistics, University of South Florida

Trend Analysis of Atmospheric Gases: Greenhouse Gases and Ozone including Fourier series and Statistical Analysis of the Relationship between these Gases using Standard Regression Analysis, Non-response Analysis (Non-functional Analysis) and Rotational Analysis, which is an extension of regression theory which allows us to analyze the constant nature of a variable, the co-dependent relationship that exist amount the various gases including standard regression analysis in modeling to estimate the various gases as a function of the other gases. Non-Response Analysis and Rotational Analysis are new statistical procedures for analyzing functional and nonfunctional relationships among variables found in real world issues involving large data sets.

Spatio–Temporal Statistical Analysis Combining Physical Models and Sensor Measurements
Youngdeok Hwang, IBM, Thomas J. Watson Research Center

Under the Smarter Planet Campaign, IBM is involved in many large-scale physical system. In such problems, the available information often includes various types of physics-based computer models as well as measurements from sensors and instruments. In this talk, we present a general spatio-temporal statistical analysis framework that can readily incorporate computer models and station measurements taken over space and time to mimic a complex physical system such as air quality. The utility of this approach is illustrated in the context of real-world problem solving.

Daily Collision Prediction by SARIMAX and GLM Models Based on Temporal and Weather Variables
Yongsheng Chen, Office of Traffic Safety, City of Edmonton, Canada

Short-term collision prediction remains an uncharted territory in the traffic safety field due to

the high randomness of data and methodological complexity. Sparked by real requirements from frontline traffic operations and enforcement services, the City of Edmonton maintains an advanced collision data system, which is used in this study to develop models to predict citywide daily total and fatal & injury (FI) collision. The study began with time series data decomposition analysis to determine trends, seasonality and randomness of the citys daily collisions, before proceeding to an investigation of potential collision contributors. Temporal features (months, weekdays and holidays) and weather forecasts (daily mean temperature, amount of rainfall and amount of snowfall) were selected as predictive factors. Accordingly, the Seasonal Autoregressive Integrated Moving Average model with external factors (SARIMAX) was identified as the preferable predictive tool for daily collisions. A series of SARIMAX models with different orders was estimated and diagnosed. Generalized linear models (GLM) were also developed and compared with SARIMAX models using goodness-of-fit measures. Finally, a validating and calibrating mechanism was constructed to optimize predictions. While predictive accuracy for FI collisions still needs improvement, the models developed in this paper have proven to be functional tools for predicting both total and FI collisions. These tools support traffic operation, enforcement scheduling, and resource and personnel scheduling and also provide situational awareness for all road users and stakeholders. Joint work with Karim El-Basyounyb - Department of Civil & Environmental Engineering, University of Alberta.

Refinement of a Method for Identifying Probable Archaeological Sites from Remotely Sensed Data
Li Chen, John Hopkins University

To discover and locate archaeological sites, we aim to develop scientific and efficient approaches to identify these sites with high accuracy. In this article, we present a statistical learning model consisting of imaging processing, feature extraction and classification. Our analysis uses the remotely sensed data composed of 8 WorldView-2 imageries and one slope band. In the imaging processing step, we use a particular annuli technique; in the feature extraction step, principal component analysis is applied; in the classification step, linear discriminant analysis is carried out. We assess this procedure on 33 lithic sites, 16 habitation sites and 100 non-sites from the western portion of Ft. Irwin, CA, USA. The receiver operating characteristic curve, used for accessing the performance of the algorithm, shows that our new approach generates higher classification power than the Archaeological Predictive Model (APM). When APM is convexly combined with our new model, the classification accuracy is even higher.

—— Thursday, June 20, 15:30-17:00 ——

Invited Session 5. Special Session on Experimental Design in Honor of Ching-Shui Cheng
Organizer and Chair: Boxin Tang, Simon Fraser University

Using graphs to find optimal block designs
Rosemary A. Bailey, University of St Andrews

Ching-Shui Cheng was one of the pioneers of using graph theory to prove results about optimal incomplete-block designs. There are actually two graphs associated with an incomplete-block design, and either can be used. A block design is D-optimal if it maximizes the number of spanning trees; it is A-optimal if it minimizes the total of the pairwise resistances when the graph is thought of as an

electrical network. I shall report on some surprising results about optimal designs when replication is very low.

Finding GMA Orthogonal Arrays by Enumeration
Dursun Bulutoglu, Department of Mathematics and Statistics, Air Force Institute of Technology

Factorial designs are used extensively in a wide range of scientific and industrial investigations for extracting as much information as possible at a fixed cost. Orthogonal arrays (OAs) that sequentially minimize the generalized wordlength pattern (GWP) are called generalized minimum aberration (GMA) OAs. GMA OAs are the most efficient factorial designs for certain statistical models. The GMA property of OAs is preserved under design isomorphism, hence classifying OAs up to isomorphism allows a GMA OA to be found. The most efficient algorithms in the literature extend all non-isomorphic OAs by columns via integer linear programming with isomorphism rejection or constraint programming with isomorphism rejection. However, both algorithms suffer from the exponential growth in the number of OAs that need to be enumerated with increasing number of runs (rows). New transformations based on the equivalence of orthogonal designs (ODs) are proposed to overcome this growth. When applied to enumerating all non-isomorphic OAs in our test bed of problems, these new transformations decreased the number of intermediate designs that needed to be enumerated by a factor of 10 to 15. This, in turn, reduced the overall solution times by the same factor.

C.-S. Cheng: Selected Highlights
John Stufken, University of Georgia

With approximately 85 publications to his credit, of which many have appeared in top journals, professor Ching-Shui Cheng's contributions to design of experiments are widespread and influential. He has made seminal contributions to the theory of optimal design of experiments for block designs, row-column designs, crossover designs, fractional factorial designs, and supersaturated designs, among others. He has also been a leader in discovering and exploring connections between research in design of experiments and research in areas of mathematics, most notably graph theory. Part of the importance of his work stems from the fact that he has often been able to state and solve problems in a more general framework, thereby generalizing or correcting results in the literature that were based on considerations of special cases. Without any illusion of being able to do justice to his many outstanding contributions, I will attempt to highlight a few selected results that stand out to me in Professor Cheng's rich publication record.

Contributed Session 3. Biostatistics and Biomedical Applications

Predicting Sparse Protein-DNA Binding Landscapes
Matthew Levinson, UCLA Department of Statistics

Study Objectives: All cellular processes are controlled, directly or indirectly, by the binding of hundreds of different DNA binding factors (DBFs) to the genome. One key to furthering our understanding of the workings of the cell is to discover where, when, and how strongly these DBFs bind to the DNA sequence. High-throughput direct measurement of DBF binding sites (e.g. through ChIP-Chip or ChIP-Seq experiments) is noisy and is not available for every DBF in every cell type. Computational approaches to predict DBF binding landscapes are thus needed to aid the study of such processes as gene regulation and chromatin remodeling.

Methods: Computational limits have led to predicting binding of DBFs separately, considering only

a small number jointly, or considering many DBFs only on smaller genomes with DBF concentration known a priori or estimated separately. We have developed a penalized Bayesian method for predicting binding landscapes that does not require the investigator to know which DBFs bind in the genomic regions of interest nor to know or measure their concentrations a priori. Our method selects the DBFs active in the regions of interest and jointly predicts their binding landscape in these regions by combining their sequence affinities, ChIP-Seq data (if available), and nucleosome occupancy models. This method employs penalty counts in iterative sampling to achieve a sparse selection of DBFs from a large candidate set and a more accurate prediction of the binding landscape for the selected DBFs. The local concentration of each selected DBF is estimated simultaneously.

Results and Conclusions: Results from analysis of selected regions in mouse embryonic stem cells (ESCs) and simulated data show that our method selects DBFs with true binding sites in the regions of interest and predicts the location of these sites more accurately than existing methods. Our method dramatically outperforms the naTve motif scanning method and the comparable computational approaches, increasing the true positives and reducing the false positive in the DBFs selected. Our penalized Bayesian approach then also significantly improves prediction of true binding sites for the selected DBFs. One goal of future applications is to detect regulatory failures that contribute to complex disorders with a genetic component.

Joint work with Qing Zhou.

Swarm Techniques for Finding Bayesian Optimal Designs for Compartmental Models
Yuanyuan Fan, UCLA Department of Biostatistics

Bayesian optimal designs have been gaining attention in recent years in industry and in biomedical studies. Bayesian optimal design methodology utilizes prior information of the unknown parameters to construct a more efficient design at minimal cost. However, there are computational difficulties in finding Bayesian optimal designs and currently there are virtually no effective algorithm that can find such optimal designs for a general statistical model and a user-selected design criterion.

I present a nature-inspired meta-heuristic algorithm based on Particle Swarm Optimization (PSO) that potentially can find Bayesian optimal designs effectively. Using ideas from PSO, I implemented a modified PSO algorithm to find Bayesian optimal designs. As illustrative examples, I will discuss applications to find Bayesian D-optimal designs for a couple of compartmental models.

Estimating treatment effect size for single-case research: special considerations for participants with autism spectrum disorder
Monica E. Carr, Monash Universtiy, Melbourne, Australia

This study has examined the application of treatment effect size calculations to supplement visual analysis in single-case research designs (SCD) for participants with Autism Spectrum Disorder (ASD). Guidelines developed by the U.S. Department of Education, What Works Clearinghouse, to inform research evidence standards have acknowledged that there is currently no agreed-upon method for treatment effect size estimation in single-case research (Kratochwill et al., 2010). A data set was developed using SCD data from self-management interventions that were conducted with participants with ASD, obtained through a systematic search of peer-reviewed literature. The feasibility of regression based effect size calculations has been explored using the baseline and treatment data. Results indicate a declining trend in the length of baseline measurements in interventions over the last 20 years. The adequacy of volume of data points from baseline and treatment phases (Parker et al., 2005) was explored, and results indicated that the majority of studies were not suit-

able for regression calculations. Ethical considerations of the unique behaviours associated with ASD are reviewed in relation to the collection of baseline data. Subsequently, three non-parametric effect calculations have been reviewed. The traditional PND method (Scruggs & Mastropieri, 1985), which has been widely used in published literature, has been compared to more recently developed calculation methods that better account for outlying variables and trend. PAND (Parker et al., 2007) and NAP (Parker et al., 2011) were selected on merit given their ease of hand calculation, making this metric widely accessible to clinicians and teachers, in addition to researchers. Treatment effect sizes derived from each method were compared, and implications for selection of a treatment effect metric in SCDs for participants with ASD are discussed.

How Technology Involved in Real Time Intensive Longitudinal Data Collection
Yan Wang, UCLA Department of Biostatistics

Traditional longitudinal data are always limited to fewer waves of measurements and are usually analyzed after all the measurements having been collected. The delay between the event occurring time and data collecting time is a potential problem for the validity of intervention. The complexity and variety of individual behavior often needs to move beyond simple time-graded effects and often requires immediate and even simultaneously action. For example, the adherence to medication is crucial for Human Immunodeficiency Virus (HIV) viral load suppression. The frequently delayed or skipped medication may lead to treatment failure and viral rebound. Modern technologies not only assist us to collect intensive longitudinal data (ILD) on adherence, such as Medication Events Monitoring System (MEMS), but allow us to record the intensive data points as soon as the events occur, such as Wisepill device, and eventually allow us to make intelligent intervention and effective treatment. Wisepill device is based on the mobile phone and Internet technologies to provide real-time medication management. We develop the web system based on the real time feedback from Wisepill device as soon as the medication taking event occurring. After analyze the event, we are able to send real time and customized reward and encouraging text messages to patients' cell phone regarding to the particular medication event using packages from Python. The combined technologies will eventually provide effective and enhanced support to the real time intervention for those who are substance users and are potentially at risk for poor medication adherence that most adherence focused interventions have only modest effects. Joint work with Jie Shen and Honghu Lui

DAY 2: FRIDAY, JUNE 21

Friday, June 21, 9:00-10:00

Plenary Session 2. Chair: Weng Kee Wong (UCLA Department of Biostatistics)

Envelopes and Partial Least Squares Regression
R. Dennis Cook (University of Minnesota)

We will begin by discussing a recently proposed context for estimation in the classical multivariate linear model, which leads to estimators of the coefficient matrix with the potential to be substantially less variable than the standard estimator. The context arises by recognizing that the responses and predictors may contain information that is immaterial to the purpose of estimating the coefficients, while still introducing substantial extraneous variation into estimation. This leads to a general parametric construct – an envelope – for removing the immaterial information and thereby reducing estimative variation.

Partial least squares (PLS) is well-known iterative methodology that is used for regression throughout the applied sciences. We will describe how PLS regression depends fundamentally on an envelope at the population level and how this envelope can be used as a well-defined parameter that characterize PLS. The establishment of an envelope as the nucleus of PLS will then open the door to pursuing the same goals as PLS but using envelope estimators that can significantly improve PLS predictions.

The discussion will include several examples for illustration. Emphasis will be placed on the concepts and their potential impact on data analysis.

Friday, June 21, 10:30-12:00

Invited Session 6. Technometrics Randy Sitter Invited Session
Organizer: Hugh Chipman, Acadia University
Chair: William Q. Meeker (Iowa State University)

Experimental Design for Engineering Dimensional Analysis
Mark Albrecht (National Marrow Donor Program, Minneapolis, MN)
Co-authors: Christopher Nachtsheim, Thomas Albrecht and Dennis Cook

Dimensional Analysis (DA) is a fundamental method in the engineering and physical sciences for analytically reducing the number of experimental variables affecting a given phenomenon prior to experimentation. Two powerful advantages associated with the method, relative to standard design of experiment (DOE) approaches are: (1) a priori dimension reduction, (2) scalability of results. The latter advantage permits the experimenter to effectively extrapolate results to similar experimental systems of differing scale. Unfortunately, DA experiments are underutilized because very few statisticians are familiar with them. In this paper, we first provide an overview of DA and give basic recommendations for designing DA experiments. Next we consider various risks associated with the DA approach, the foremost among them is the possibility that the analyst might omit a key explanatory variable, leading to an incorrect DA model. When this happens, the DA model will

fail and experimentation will be largely wasted. To protect against this possibility, we develop a robust-DA design approach, that combines the best of the standard empirical DOE approach with our suggested design strategy. Results are illustrated with some straightforward applications of DA.

Discussants: Tim Davis (We Predict Ltd), Bradley Jones (SAS Institute Inc.), Matthew Plumlee (Georgia Tech)

Invited Session 7. Bioinformatics and Data Mining in the Medical Sciences
Organizer: Christina Kitchen, UCLA Department of Biostatistics
Chair: Donatello Telesca, UCLA Department of Biostatistics

A Bayesian regression tree approach to identify the effect of nanoparticles properties on toxicity profiles
Cecile Low-Kam, Montreal Heart Institute

The increasing use of engineered nanomaterials (ENM) has recently raisen concern about their potential environmental hazard. Currently, there is a growing interest in developing models to relate physiochemical characteristics of ENM, such as size or dissolution rate, to their toxicity profile, measured over dose and time exposure escalation designs.

We introduce a Bayesian regression tree model that integrates Bayesian trees for modeling threshold effects and interactions of physiochemical properties, and penalized B-splines for dose and time-response surfaces smoothing. Unlike conventional models that rely on data summaries, our model solves the low sample size issue and avoids arbitrary loss of information by combining all measurements across doses, times of exposure, and replicates. The resulting posterior distribution is sampled via a Markov Chain Monte Carlo algorithm. This method allows for inference on a number of quantities of potential interest to substantive nanotoxicology, such as the importance of physiochemical properties and their marginal effect on toxicity. We illustrate the application of our method to the analysis of a library of 24 metal oxides.

multiGeMS: Multi-Sample SNP Detection Using Genotype Model Selection on High-Throughput Sequencing Data
Xinping Cui, Department of Statistics, UC Riverside

Recent advances in high-throughput sequencing (HTS) promise revolutionary impacts in science and technology, including the areas of disease diagnosis, pharmacogenomics, and mitigating antibiotic resistance. In particular, it has greatly expanded our ability to perform genome-wide survey of single nucleotide polymorphisms (SNPs) to better understand the biological mechanisms through which genetic polymorphisms affect phenotypes. Considering a selection of popular HTS SNP calling procedures, it becomes clear that many rely mainly on base-calling and read mapping quality values. Thus there is a need to consider other sources of error, such as those occurring during genomic sample preparation, which might increases false positive SNP calls. On the other hand, many studies prepare and sequence multiple samples with the same protocol. Pooling information across samples can contribute better learning on the error properties of the data and therefore help reduce the number of false positive SNP calls. We propose a multi-sample based Genotype Model Selection (multiGeMS) method, a method of consensus and SNP calling which accounts for sequencing, mapping and other possible sources of errors with the aid of cross-sample information. Through simulation studies, we demonstrate that multiGeMS outperforms a selection of popular SNP callers.

Using the random generalized linear model to predict clinical outcomes based on genomic data

Steve Horvath, Peter Langfelder and Lin Song, UCLA Department of Genetics*

Ensemble predictors such as the random forest are known to have superior accuracy but their black-box predictions are difficult to interpret. In contrast, a generalized linear model (GLM) is very interpretable especially when forward feature selection is used to construct the model. However, forward feature selection tends to overfit the data and leads to low predictive accuracy. Therefore, it remains an important research goal to combine the advantages of ensemble predictors (high accuracy) with the advantages of forward regression modeling (interpretability). To address this goal several articles have explored GLM based ensemble predictors. Since limited evaluations suggested that these ensemble predictors were less accurate than alternative predictors, they have found little attention in the literature. Here we use comprehensive evaluations involving hundreds of genomic data sets to re-evaluate GLM based ensemble predictors. The random generalized linear model (RGLM) predictor is a bootstrap aggregated GLM predictor that incorporates several elements of randomness and instability (random subspace method, optional interaction terms, forward variable selection). It often outperforms a host of alternative prediction methods including random forests and penalized regression models. Further, the RGLM predictor provides variable importance measures that can be used to define a "thinned" ensemble predictor (involving few features) that retains excellent predictive accuracy.

RGLM is a state of the art predictor that shares the advantages of a random forest (excellent predictive accuracy, feature importance measures, out-of-bag estimates of accuracy) with those of a forward selected generalized linear model (interpretability). These methods are implemented in the freely available R software package random GLM.

Relevant reference: BMC Bioinformatics 2013, 14:5

Contributed Session 4. Screening Design, Latin Hypercube Design and Kriging

A Systematic Approach for the Construction of Definitive Screening Designs

Frederik K. H. Phoa, Institute of Statistical Science, Academia Sinica, Taiwan

Definitive screening (DS) designs, first proposed by Jones and Nachtsheim (2011), draw numerous attentions from the researches of designs of experiments due to its good design properties and run-size economy. This paper investigates in the structure of DS designs and suggests a theoretically-driven approach to construct DS designs for any number of run size. This approach is applicable for both even and odd number of factors. The D-efficiencies of some DS designs constructed via our approach are equivalently optimal to existing theoretical results and higher than those reported in Jones and Nachtsheim (2011) and Xiao et al. (2012).

Optimal Sliced Latin Hypercube Designs

Shan Ba, The Fariborz Maseeh Dept. of Mathematics and Statistics, Portland State University

Sliced Latin hypercube designs (SLHDs) have important applications in designing computer experiments with quantitative and qualitative factors and in running ensembles of multiple computer codes. However, a randomly generated SLHD can be poor in terms of space-filling, and it is hard to find the optimal SLHDs based on the existing construction method. In this work, we develop a new construction approach which first generates the small Latin hypercube design in each slice and then arranges them together to form the SLHD. The new approach is very intuitive and it enables us to

develop efficient algorithms to generate the optimal SLHDs. In addition, it can also be adapted to produce the orthogonal SLHDs and the orthogonal array-based SLHDs

Sliced Full Factorial-Based Latin Hypercube Designs as a Framework for a Batch Sequential Design Algorithm

Weitao Duan, Dept. of Industrial Engineering and Management Science, Northwestern University

When fitting complex models, such as finite element or discrete event simulations, the experiment design should exhibit good properties of both projectivity and orthogonality. To reduce experimental effort, sequential de- sign strategies allow experimenters to collect data only until some measure of prediction precision is reached. In this article, we present a batch sequential experiment design method that is based on sliced Full Factorial-Based Latin Hypercube Designs (FFLHD), which are an extension to the concept of sliced Orthogonal Array-Based Latin Hypercube Designs (OALHD). At all stages of the sequential design, good univariate stratification is achieved. The structure of the FF-based LHD also tends to produce uniformity in higher dimensions especially at certain stages of the design. We show that the designs have good sampling and fitting qualities through both empirical studies and theoretical arguments. Joint work with Bruce Ankenmana, Paul Sanchez, and Susan Sanchez.

An Information Bound for Kriging

Selden Crary, Newall Street Inc, Palo Alto, CA

Haaland and Qian's (H&Q) Y2011 Annals paper (Annals of Statistics **39**, pp. 2974-3002) showed that, as the number N of equispaced, collected data points in an interval of length Δ increases, there is a point at which ill-conditioning is so severe that fixed-θ Kriging fits to the data deteriorate badly. This indicates that Kriging may have inherent, information upper bounds. We report that such bounds exist and how they scale with N, θ, patch volume, and the precision of the computer in use. Our information-bound analysis makes use of our new, exact results for the L_1- and $L-\infty$-norm condition numbers of the covariance matrices of gridded data under the fixed-θ, polynomial-exponential class of covariance functions $\sigma_Z^2(1+q_1\theta|x_i-x_j|+q_2\theta^2|x_i-x_j|^2+\cdots)e^{-\theta|x_i-x_j|^p}$. In 1D and $p=1$, the leading term of the condition number, in an expansion in nearest-neighbor spacing δ, is proportional to $(\theta\delta^2)^{-1}$, independent of N, whereas for $p=2$, the leading term is proportional to $(\theta\delta^2)^{-(N-1)}$. We present a new theorem showing how the condition numbers scale with the number of spatial dimensions. We also report that $p=1$ is, in general, the largest exponent that does not introduce, almost always, a long-memory effect, i.e., a far-field different from the average of the responses.

—— Friday, June 21, 13:30-15:00 ——

Invited Session 8. Uncertainty Quantification: Using Complex Models to Aid Statistical Inference

Organizer and Chair: Dave Higdon, Los Alamos National Lab

Engineering-Driven Statistical Adjustment

V. Roshan Joseph and Huan Yan, Georgia Institute of Technology*

Engineering model development involves several simplifying assumptions for the purpose of mathematical tractability which are often not realistic in practice. This leads to discrepancies in the model

predictions. A commonly used statistical approach to overcome this problem is to build a statistical model over the engineering model using real data. In contrast, an engineering approach would be to find the causes of discrepancy and fix the engineering model using first principles. However, the engineering approach is time consuming, whereas the statistical approach is fast. The drawback of the statistical approach is that it treats the engineering model as a black box and therefore, the statistically adjusted models lack physical interpretability. This paper proposes a new framework for model calibration and statistical adjustment. It tries to open up the black box and introduces statistical models inside the engineering model. This approach leads to simpler adjustment models that are physically more interpretable. The approach is illustrated using a model for predicting the cutting forces in a laser-assisted mechanical micromachining process.

Bayesian Experimental Design Using Gaussian Processes
Brian Weaver, Statistical Sciences Group, Los Alamos National Laboratory

Bayesian experimental design provides an approach for incorporating prior information into design selection before performing an experiment. As intuitive and appealing as these methods are, Bayesian experimental design quickly becomes analytically and numerically intractable. For example, how does one minimize a posterior variance prior to observing any data? Simulation based methods have been developed to provide a computationally viable solution to these difficulties, for example density approximation based methods as well as chaotic polynomial approximations to the criterion we wish to optimize. In this talk we present methods using Gaussian processes to approximate the surface of the criteria function. These methods are illustrated with a simple example. This is joint work with Brian Williams, Dave Higdon, and Christine Anderson-Cook of Los Alamos National Laboratory.

Projective Updating of Probabilistic Models
Evangelia Kalligiannaki and Roger Ghanem, University of Southern California*

We describe the updating of polynomial chaos (PC) representations of probabilistic models. The updates are carried out to augment prior representations with constraints that can take various forms. The prior PC models are typically the output of computer models that implement stochastic Galerkin projections predicated on specific physics models and specific probabilistic models of associated parameters. These PC describe a mean-square convergent approximation of the map from input to output parameters. The additional constraints reflect new knowledge concerning either the underlying physics model or the probabilistic models of parameters. Given the Hilbert construction of associated with the PC Galerkin projection, we implement the constraints as projections onto suitable sets. We also explore the connection between this updating procedure and situations under which it coincides with statistical conditioning. In this latter case, we explore the connection with Bayes theorem. We note that projective updating, in the manner described here, is a more logical mechanism for satisfying constraints in the L_2 setting already in place to construct the PC decompositions.

Invited Session 9. Modern Meta-Heuristic Algorithms For Statistical Applications
Organizer and Chair: Ray-Bing Chen (National Cheng Kung University) and Weng Kee Wong (UCLA)

Constrained Multi-objective Designs for Functional MRI Experiments via A Modified Non-dominated Sorting Genetic Algorithm
Ming-Hung Kao, School of Mathematical & Statistical Sciences, Arizona State University

Functional magnetic resonance imaging (fMRI) is an advanced technology for studying brain functions. Due to the complexity and high cost of fMRI experiments, high quality multi-objective fMRI designs are in great demand; they help to render precise statistical inference, and are keys to the success of fMRI experiments. Here, we propose an efficient approach for obtaining multi-objective fMRI designs. In contrast to existing methods, the proposed approach does not require users to specify weights for the different objectives, and can easily handle constraints to fulfill customized requirements. Moreover, the underlying statistical models that we consider are more general. We can thus obtain designs for cases where brief, long or varying stimulus durations are utilized. The usefulness of our approach is illustrated using various experimental settings.

Optimizing Latin hypercube Designs by Particle Swarm

Ray-Bing Chen, Department of Statistics, National Cheng-Kung University, Tainan, Taiwan

Latin hypercube designs (LHDs) are widely used in many applications. As the number of design points or factors becomes large, the total number of LHDs grows exponentially. The large number of feasible designs makes the search for optimal LHDs a difficult discrete optimization problem. To tackle this problem, we propose a new population-based algorithm named LaPSO that is adapted from the standard particle swarm optimization (PSO) and customized for LHD. Moreover, we accelerate LaPSO via a graphic processing unit (GPU). According to extensive comparisons, the proposed LaPSO is more stable than existing approaches and is capable of improving known results.

Global Optimization for Nonlinear Mixed-Effects Pharmacokinetic - Pharmacodynamic Models Using Particle Swarm Optimization

Seongho Kim, Department of Bioinformatics and Biostatistics, University of Louisville

The nonlinear mixed-effects model is a powerful statistical method in analyzing repeated-measurements data that arise in the field of population pharmacokinetics/pharmacodynamics (PK/PD). The most popular approaches to the nonlinear mixed-effects PK/PD analysis is NONMEM. However, NONMEM usually requires an initial value close enough to the global optimum since it is a local optimization algorithm. In this talk, we will present a novel global search algorithm called P-NONMEM. It combines the global search strategy by particle swarm optimization (PSO) and the local estimation strategy of NONMEM. In the developed algorithm, initial values (particles) are generated randomly by PSO, and NONMEM is implemented for each particle to find a local optimum for fixed effects and variance parameters. P-NONMEM guarantees the global optimization for fixed effects and variance parameters. Under certain regularity conditions, it also leads to global optimization for random effects. Because P-NONMEM does not run PSO search for random effect estimation, it avoids tremendous computational burden. In the simulation studies, we will show that P-NONMEM has much improved convergence performance than NONMEM. Even when the initial values were far away from the global optimal, P-NONMEM converged nicely for all fixed effects, random effects, and variance components.

Let Darwin and the Bees Help Improve your Designs: Nature Inspired Optimization Techniques in Engineering

Yahya Rahmat-Samii, UCLA Department of Electrical Engineering

Engineers are constantly challenged with the temptation to search for optimum solutions for complex engineering system designs. The ever increasing advances in computational power have fueled this temptation. The well-known brute force design methodologies are systematically being replaced by the state-of-the-art Evolutionary Optimization (EO) techniques. In recent years, EO techniques are finding growing applications to the design of all kind of systems with increasing complexity. Among

various EO's, nature inspired techniques such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have attracted considerable attention. GA utilizes an optimization methodology which allows a global search of the cost surface via the mechanism of the statistical random processes dictated by the Darwinian evolutionary concept (adaptation, selection, survivability and mutation). PSO is a robust stochastic evolutionary computation technique based on the movement and intelligence of swarms of bees looking for the most fertile feeding location applying their cognitive and social knowledge. This presentation will focus on: (a) an engineering introduction to GA and PSO by describing in a novel fashion the underlying concepts and recent advances for those who have used these techniques and for those who have not had any experiences in these areas, (b) demonstration of the potential applications of GA's and PSO's to a variety of electromagnetic and antenna designs and (c) assessment of the advantages and the limitations of these techniques.

Contributed Session 5. Risk Analysis and Reliability

On a Two–Unit Cold Standby System with Three Modes and Imperfect Switch Considering Preventive Maintenance

Mahmoud Mahmoud, Mathematics Dept, Faculty of Science, Al-Azhar University, Cairo, Egypt

In this paper the stochastic analysis of a two-unit cold standby system with three modes and imperfect switch considering preventive maintenance (PM) is studied. All time distributions are arbitrary. Various reliability measures of this system, such as mean time to system failure, steady state availability, busy periods and the net gain, are deduced based on regenerative point technique. The effect of PM on the reliability measures is also discussed. Finally graphical representation is given to illustrate the results. Joint work with A.A. Sedeka, and A.E.B. Abdelghany

Degradation Data Analysis Using Nonlinear Mixed–effects Modelsvia Shape–restricted Splines

Zhibing Xu, Department of Statistics, Virginia Tech

Compared with traditional lifetime data, degradation data not only can provide reliability information in a short time period, but also can give better estimation and prediction for the lifetime of products. With the improvement of technology, many life–affecting environmental variables and usage information can be recorded as well as the degradation measure over time, which is referred to as dynamic covariates. For example, the highly variable ultraviolet exposure can be recorded, which will cause the degradation of polymer coating. Thus, it is important to incorporate those environmental variables into the degradation path model. In literature, a linear mixed effect model with shape-restricted splines is used to incorporate the dynamic covariates. Although the linear mixed effects model fits the data well, it does not have the physical or chemical meaning for the degradation path. Compared to the linear relationship, nonlinear relationship is more common in reality, especially in the pharmacokinetic research. In this paper, we propose a nonlinear mixed effects model combing with dynamic covariates to model the degradation data. Shape–restricted splines are used in the proposed model and a modified alternating algorithm is proposed. The convergence of the algorithm is validated by simulations. An outdoor weathering dataset is used to illustrate the proposed method. This is a joint work with Yili Hong.

Reliability Assessments with Restricted Degradation Measurements

Shuen-Lin Jeng, Department of Statistics, National Cheng Kung University

Complete degradation measurements may not be available for certain devices during a life test.

The devices under testing may be inoperable when the cumulative degradation crosses a threshold level (TL) or may have a mode change after a particular degradation level. The testing budget will also limit the sample size and the inspection frequency in order to reduce costs. In this research we investigate the effects of the restricted degradation measurements (RDMs) on the life time inferences based on a nonhomogeneous compound Poisson model. Two data sets, ultra-thin gate oxides (UTGOs) and metallized film pulse capacitors (MFPCs), are analyzed by our approach. The UTGO data is recorded from continued inspections and the compound part of the process is modeled by a Weibull distribution. The MFPC data is recorded from scheduled inspections and the compound part of the process is modeled by a gamma distribution. Simulations based on the failure mechanisms of UTGO and MFPC are carried out to explore the effects of sample size, restricted degradation level, and inspection frequency on the lifetime estimates.

Optimal Design for Accelerated Destructive Degradation Tests
Chien-Tai Lin, Dept. of Mathematics, Tamkang University, New Taipei City, Taiwan

Degradation tests are powerful and useful tools for lifetime assessment of highly reliable products. In some applications, the degradation measurement process would destroy the physical characteristic of units when tested at higher than usual stress levels of an accelerating variable such as temperature, so that only one measurement can be made on each tested unit during the degradation testing. An accelerated degradation test giving rise to such a degradation data is called an accelerated destructive degradation test (ADDT). The specification of the size of the total sample, the frequency of destructive measurements, the number of measurements at each stress level, and other decision variables are very important to plan and conduct an ADDT efficiently. A wrong choice of these decision variables may not only result in increasing the experimental cost, but may also yield an imprecise estimate of the reliability of the product at the use condition. Motivated by a polymer data, this article deals with the problem of designing an ADDT with a nonlinear model. Under the constraint that the total experimental cost does not exceed a pre-fixed budget, the optimal test plan is obtained by minimizing the asymptotic variance of the estimated 100pth percentile of the productÕs lifetime distribution at the use condition. A sensitivity analysis is also carried out to examine the effects of changes in the decision variables on the precision of the estimator of the 100pth percentile. Joint work with Chih-Chun Tsai, Sheng-Tsaing Tseng, and N. Balakrishnan

—— Friday, June 21, 15:30-17:00 ——

Invited Session 10. Experimental Design
Organizer and Chair: Hongquan Xu, UCLA

Nearly Orthogonal Space-Filling Designs Uniform in Two-Dimensional Projections
Brad Jones, JMP Division, SAS Institute Inc.

Latin Hypercube designs (LHDs) are the default design for exploring deterministic computer simulation models. One reason for their popularity is that they project uniformly on every factor. A random LHD for k factors is comprised of k suitably scaled permutations of the integers from 1 to n where n is the number of experimental runs. Since there are a very large number of these, it is desirable to choose a secondary criterion to make the design choice unique. For example, there are orthogonal (or nearly orthogonal) LHDs that try to make the columns of the design uncorrelated. Orthogonality and uniform one-dimensional projections seem like desirable properties, yet in

combination they can result in designs with questionable coverage of any pair of factors.

This talk introduces a new kind of space filling design that provides zero or negligible correlations between pairs of factors along with uniform coverage in all projections to two-dimensions. These design are produced algorithmically using a multiple objective optimization approach. The approach allows for substantial flexibility in the number of simulation runs and the number of factors.

A new framework on calibration for computer models: estimation and convergence properties

C. F. Jeff Wu, Georgia Institute of Technology

Calibration parameters in deterministic computer experiments are those attributes that cannot be measured or available in physical experiments or observations. Kennedy-O'Hagan (2001) suggested an approach to estimate them by using data from physical experiments and computer simulations. We study the estimation problem and show that a simplified version of the original KO method leads to asymptotically inconsistent calibration. This calibration inconsistency can be remedied by modifying the original estimation procedure. A novel calibration method, called the L2 calibration, is proposed and proven to be consistent and enjoys optimal convergence rate. A numerical example and some mathematical analysis are used to illustrate the source of the inconsistency problem.

(joint work with Rui Tuo, Chinese Academy of Sciences)

Design of experiments for Bayesian model discrimination

Dave Woods, Southampton Statistical Sciences Research Institute, University of Southampton, UK

The design of any experiment is implicitly Bayesian, with prior knowledge being used informally to aid decisions such as which factors to vary and the choice of plausible causal relationships between the factors and measured responses. Bayesian methods allow uncertainty in these decisions to be incorporated into design selection through prior distributions that encapsulate information available from scientific knowledge or previous experimentation. Further, a design may be explicitly tailored to the aim of the experiment through a decision-theoretic approach with an appropriate loss function.

This talk will describe a new decision-theoretic criterion for design selection when the aim of the experiment is discrimination between rival statistical models. Motivated by an experiment from materials science, we consider the problem of early stage screening experimentation to choose an appropriate linear model, potentially including interactions, to describe the dependence of a response on a set of factors. We adopt an expected loss for model selection which is a weighted sum of posterior model probabilities and introduce the Penalised Model Discrepancy (PMD) criterion for design selection.

The use of this criterion is explored through a variety of issues pertinent to screening experiments, including the choice of initial and follow-up designs and the robustness of design performance to prior information. Designs from the PMD criterion are compared with those from existing approaches through examples. We also investigate reducing the computational burden of the method for experiments with a large number of contending models, through both the use of informative prior distributions and the approximation of the PMD objective function. Some directions of current and future research will also be discussed.

Contributed Session 6. Machine Learning: Methods and Applications

Feature Selection by Scheduled Elimination
Adrian Barbu and Yiyuan She, Department of Statistics, Florida State University*

Many imaging problems are faced with learning classifiers from large datasets, with millions of observations and features. In this work we propose a novel efficient algorithm for variable selection and learning on such datasets, optimizing a constrained penalized likelihood without any sparsity inducing priors. The iterative suboptimal algorithm alternates parameter updates with tightening the constraints by gradually removing variables based on a criterion and a schedule. We present a generic approach applicable to any differentiable loss function and present an application to logistic regression. Experiments on real and synthetic data show that the proposed method outperforms Logitboost and penalized methods for both variable selection and prediction while being computationally faster.

Casual Search, Bootstrap Aggregation and Bayesian Networks: An Application to the Consumer Packaged Goods Industry
Irina Kukuyeva, Director, Ipsos Science Centre

Variable importance has multiple valid definitions in statistics – but for businesses looking to increase their share of consumers' wallets, driver importance is typically understood as identifying which variables have the largest potential to boost business metrics. Accordingly, clients expect key driver analysis to capture the magnitude of this potential causal impact. Previous research by the authors' organization has identified Bayesian networks as one of the most promising methods of estimating this impact (Egner, Porter and Hart 2011), particularly when the method is modified to (1) run off of the causal search algorithms developed by Pearl (2009) and others; (2) use bootstrap aggregation (or "bagging") for greater robustness (3) employ smoothing algorithms to adjust conditional probabilities when sample sizes are small; and (4) generate true simulations rather than mere conditional subsets (Egner and Hart, 2012). However, these previous analyses were run on simulated data alone. In this paper, we demonstrate for the first time how to identify the practical value of this approach using real data from a multibillion-dollar Consumer Packaged Goods company. Even when the "true" value of the drivers is not known, it is still possible to see how Bayesian networks run in the absence of these modifications, particularly the absence of bootstrap aggregation, can produce potentially misleading results for clients. Joint work with Michael Egner, Ph.D. and Richard Timpone, Ph.D.

Extension of Sliced Inverse Regression for High Dimensional but Low Sample Size Data
Jingyi Zhu, Department of Statistics, Purdue University

Sliced inverse regression (SIR) is known to be an effective dimension reduction method that extracts features of high dimensional data from low dimensional projections. SIR is based on eigenvector decomposition of sample covariance matrix and only works for classical data structures, when the dimension p is smaller than the sample size n. In our research, we extend the SIR method to high dimensional but low sample size data. When $p \gg n$, the standard sample covariance matrix is not of full rank nor a consistent estimate. To overcome the difficulty of estimating large covariance matrices and their inverses, we employ the least squares formulation of SIR and apply an alternating least squares method to obtain SIR estimators. The alternating least squares method is not limited by high dimensional but low sample size data hence more general than the eigenvector decomposition algorithm. We improve the alternating least squares by using a consistent estimate of the large

covariance matrix and regularizing the least squares by appropriate penalty functions. Our method is demonstrated in simulation studies and a pharmacogenomics study of bortezomib for multiple myeloma.

Simultaneous Envelopes for Multivariate Linear Regression
Xin Zhang, School of Statistics, University of Minnesota

We introduce envelopes for simultaneously reducing the predictors and the responses in multivariate linear regression, so the regression then depends only on estimated linear combinations of X and Y. We use a likelihood-based objective function for estimating envelopes and then propose algorithms for estimation of a simultaneous envelope as well as for basic Grassmann manifold optimization. The asymptotic properties of the resulting estimator are studied under normality and extended to general distributions. We also investigate likelihood ratio tests and information criteria for determining the simultaneous envelope dimensions. Simulation studies and real data examples show substantial gain over the classical methods like, partial least squares and canonical correlation analysis.

DAY 3: SATURDAY, JUNE 22

Saturday, June 22, 8:30-10:00

Invited Session 11. New Statistical Procedures for Analyzing Big Data
Organizer: Runze Li, Pennsylvania State University
Chair: Peter Z. Qian (University of Wisconsin)

Estimating Single Index Models in High Dimensions
Peter Radchenko, University of Southern California

The talk will address the problem of fitting nonlinear regression models in high-dimensional situations, where the number predictors may be much larger than the number of observations. Most of the research in this area has been conducted under the assumption that the regression function has a simple additive structure. We will focus instead on single index models, which are becoming increasingly popular in many scientific fields including financial econometrics, economics and biostatistics. Novel methodology will be presented for fitting high-dimensional single index models by defining the estimator as the minimizer of a constrained least-squares criterion. This approach will be demonstrated to be more appealing than the L1 penalization approach, which results in solutions that are discontinuous with respect to the tuning parameter. A computationally efficient algorithm will be provided for constructing the solution path, and high-dimensional theoretical properties will be discussed. A comprehensive simulation study and an empirical analysis of the new estimator and some of the competitors will be presented.

The Analysis of Group fMRI Data Using L1, L2, and Mixed-Type Penalties
Ronghui "Lily" Xu, University of California at San Diego

Motivated by clinical studies aimed at incorporating fMRI technology into the treatment of conditions such as pain, we consider the analysis of group fMRI data. 'Group' here refers to different groups of subjects under different treatments. As different studies may aim at different characteristics of neurological response to a treatment, we seek to characterize the underlying time course response curves (and their derivatives). Using mixed effects to represent individual subject effects, as well as serial correlation for the errors, we consider L1, L2, and mixed L1 and L2 type penalties in estimating the group response curve. Different model selection techniques will be discussed, and emphasis will be on interval estimation as anticipated in clinical applications.

High-Dimensional Sparse Additive Hazards Regression
Jinchi Lv, University of Southern California

High-dimensional sparse modeling with censored survival data is of great practical importance, as exemplified by modern applications in high-throughput genomic data analysis and credit risk analysis. In this article, we propose a class of regularization methods for simultaneous variable selection and estimation in the additive hazards model, by combining the nonconcave penalized likelihood approach and the pseudoscore method. In a high-dimensional setting where the dimensionality can grow fast, polynomially or nonpolynomially, with the sample size, we establish the weak oracle property and oracle property under mild, interpretable conditions, thus providing strong performance guarantees for the proposed methodology. Moreover, we show that the regularity conditions required by the L_1 method are substantially relaxed by a certain class of sparsity-inducing concave penalties.

As a result, concave penalties such as the smoothly clipped absolute deviation (SCAD), minimax concave penalty (MCP), and smooth integration of counting and absolute deviation (SICA) can significantly improve on the L_1 method and yield sparser models with better prediction performance. We present a coordinate descent algorithm for efficient implementation and rigorously investigate its convergence properties. The practical utility and effectiveness of the proposed methods are demonstrated by simulation studies and a real data example. This is a joint work with Wei Lin.

Invited Session 12. Applied Bayesian Design of Experiments
Organizer and Chair: Dave Woods, University of Southampton

Optimal Bayesian design using Gaussian process emulators
Antony Overstall, University of St Andrews, UK

Bayesian optimal design criteria usually aim to maximise some objective function that is related to the expected utility of an outcome of the experiment. This expectation is usually with respect to both the unobserved data and the unknown parameters. In all but the simplest cases, evaluating the objective function is analytically intractable as there is no closed-form solution to the integrals involved.

One approach is to approximate the objective function using Monte Carlo integration methods. Unfortunately, this introduces a level of computational expense that makes the optimisation both tedious and extremely time-consuming. To overcome this problem, we use a Gaussian process emulator of the 'noisy', computationally expensive, Monte Carlo approximation. We can then use the computationally cheap emulator to approximately maximise the objective function.

We demonstrate this approach by producing designs for both generalised linear models and non-linear models.

Joint work with Dave Woods (University of Southampton, UK)

A D-optimal design for estimation of parameters of an exponential-linear growth curve of nanostructures
Tirthankar Dasgupta, Harvard University

We consider the problem of determining an optimal experimental design for estimation of parameters of a complex curve characterizing nanowire growth that is partially exponential and partially linear. A locally D-optimal design for the non-linear change-point growth model is obtained by using a geometric approach. Further, a Bayesian sequential algorithm is proposed for obtaining the D-optimal design. The advantages of the proposed algorithm over traditional approaches adopted in recently reported nano-experiments are demonstrated using Monte-Carlo simulations.

Joint work with Li Zhu (Harvard) and Qiang Huang (University of Southern California).

A particle filter for Bayesian sequential design
James McGree, Queensland University of Technology, Australia

A particle filter algorithm will be presented that can be used for Bayesian sequential design in the presence of model and parameter uncertainty. Examples will be presented that consider different design objectives such as parameter estimation and model discrimination, but the algorithm can be applied more generally. Particle filters are run in parallel for each model and the algorithm relies on a convenient estimator of the marginal likelihood of each model. Approximating posterior model probabilities in this way allows us to use, for example, model discrimination utilities derived from

information theory that were previously difficult to compute for all but the simplest of models. We motivate our research by the application of sequential design to a number of real world examples where various rival nonlinear models are contemplated.

Joint work with Chris Drovandi and Tony Pettitt (Queensland University of Technology).

Contributed Session 7. Fractional Factorial Designs

A Potential Outcomes Perspective of the Analysis of Unreplicated Factorial Designs
Valeria Espinosa Mateos, Department of Statistics, Harvard University

Factorial designs have been widely used in many scientific and industrial settings. The traditional ways of analyzing such experiments assume an underlying normal population and restrict the range of effects that can be tested to the means. We explore two methods based on the Rubin Causal Model framework which allow the relaxation of these assumptions : a potential outcome - single imputation version of the Loughin and Noble (1997) proposal and a Sequential Posterior Predictive Check approach. Both methods, as the Loughin and Noble, have the same starting point: a Fisher randomization test for a sharp null hypothesis. Different discrepancy measures and stopping rules are compared to other standard analysis methods such as the original and Step–Down Lenth methods. Simulation results across a variety of alternatives will be presented.

Some Results on Conditional Main Effect Analysis
Heng Su, Georgia Tech

Ever since the founding work by Finney, it has been widely known and accepted that aliased effects in two-level regular designs cannot be de-aliased without adding more runs. A surprising result by Wu in his 2011 Fisher Lecture showed that aliased effects can sometimes be "de-aliased" using a new framework based on the concept of conditional main effects (CMEs). This talk will report some further work. Wu's approach expands the design space of the aliased effects and their corresponding main effects into the CMEs, and use variable selection to identify the significant ones. The complex aliasing patterns between the CMEs will be analyzed, and the orthogonality property of the expanded design space will be studied. Some practical rules as to when and whether to use the CME analysis will be shown with examples. Under certain circumstances, the CME analysis can offer substantial increase in the R-squared value with fewer effects in the chosen models. Moreover, the selected significant effects are often interpretable. (Joint work with C. F. Jeff Wu)

Semifoldover Two-level Regular Blocked Factorial Designs
Po Yang, Department of Statistics, University of Manitoba Winnipeg

Foldover designs have been in the literatures for many years, the new added runs are used to de-alias confounded effects. Semifoldover designs are partial foldover designs obtained by adding half of the new runs. It is known that, in many cases, a semifoldover design can de-alias as many main factors or two-factor interactions as the corresponding foldover design. In this talk, we consider the impact of blocking effect on semifoldover designs. We show that a semifoldover blocked design can de-alias as many two-factor interactions as the corresponding foldover blocked design when the initial design is a resolution IV design. The optimal blocking and semifoldover plans in terms of the clear effect criterion are tabulated for 16 and 32 runs.

Interesting Insights in Indicators: Indicator Functions and the Algebra of the Linear-Quadratic Parameterization
Arman Sabbaghi, Department of Statistics, Harvard University

Indicator functions provide a general perspective on the design and analysis of fractional factorials. An important characteristic of indicator functions is that, once constructed, they yield immediate information on partial aliasing of contrasts for a design. Previous construction methods for the linear-quadratic parameterization of contrasts involve tedious and routine calculations, necessarily performed on a case-by-case basis for various designs. We demonstrate a simpler approach that yields deep insights on the aliasing structure of contrasts for large classes of fractional factorials. Novel algebraic operators are introduced that eliminate tedious and routine calculations, and provide immediate structural results on partial aliasing of contrasts by simple inspection of the design matrix. Ultimately, these operators connect design construction and analysis methods under the linear-quadratic system, and help establish simple and explicit conditions for the estimability of interactions. Joint work with Tirthankar Dasgupta and Chien-Fu Jeff Wu.

—— Saturday, June 22, 10:30-12:00 ——

Invited Session 13. Reliability and SPC Applications
Organizer and Chair: Yili Hong, Virginia Tech

Estimating Failure-Time Distributions and Predicting Field Failures with Censored Data and an Unknown Retirement Times

William Q. Meeker, Department of Statistics, Iowa State University

Nowadays, many consumer products are designed and manufactured so that the probability of failure during the technological life of the product is small. Most product units in the field retire before they fail. Even though the number of failures from such products is small, there is still the need to model and predict field failures in applications that involve safety and risk assessment. Challenges in modeling and predictions of failures arise because the retirement times are often unknown, and there are delays in field failure reporting. Motivated by an application to assess the risk of failure, we develop a statistical procedure to predict the field failures of products, which considers the impact of product retirements and reporting delays. Based on the developed method, we provide the point prediction for cumulative number of reported failures in a future time and the corresponding prediction interval to quantify uncertainty. We also conduct sensitivity analysis to show the effects of different assumptions on failure-time and retirement distributions and the values of their parameters.

This is joint work with Zhibing Xu and Yili Hong

Estimation of Warranty Repair Demand Considering New Sales and Failed-but-not-reported Phenomena

Haitao Liao, Department of Systems and Industrial Engineering, University of Arizona

When customers buy a product, they are often eligible for free repairs for a certain warranty period. In this paper, we study some important aspects in estimating warranty repair demands that are of interest to the manufacturer and/or the designated service provider. The study is limited to a new product under a non-renewable free minimal-repair warranty policy. We consider that the installed base of the product varies with time due to both new sales and units being taken out of service. We explicitly address the fact that customers may not always request repairs for failed units and formulate the corresponding warranty repair demands. For the case where the product failure time is exponential, we derive the closed-form expression for the repair demands for both an individual

unit and the installed base. The relevant statistics and parameter estimation method, and the insights into some risk-related quantities are presented in detail. Numerical examples are provided to illustrate that joint consideration for both new sales and failed-but-not-reported phenomena is important for managing the obligatory repair services.

Nonparametric CUSUM Control Charts for Periodic SPC Applications
Daniel R. Jeske, Department of Statistics, University of California, Riverside

We propose how to use the CUSUM for sequential monitoring of independent and identically distributed observations when the underlying in-control density is arbitrary and unknown, and the out-of-control density is either an additive or a multiplicative transformation of the in-control density. We develop our methodology under a framework where periodic cycles of finite length are considered. Our approach utilizes a smooth bootstrap algorithm along with an adaptive nonparametric kernel density estimator to make the CUSUM work for reasonably sized sets of in-control data. We demonstrate the effectiveness of our approach with a simulation study.

This is a joint work with Tatevik Ambartsoumian.

Invited Session 14. Computer Experiments
Organizer and Chair: Roshan Joseph Vengazhiyil (Georgia Tech)

Assessing the chance of a satellite collision using a computational propagator model
Earl Lawrence, Dave Higdon (Los Alamos National Laboratory), Mike Shoemaker*

The orbit path of satellites is predicted using high fidelity computer models that propagate an object according to the effects of gravity and atmospheric drag. At any given time, the exact position and velocity of any satellite is uncertain; so is the state of the atmosphere. Hence the future path of the satellite is also uncertain. While the probability of a collision between two satellites can be computed, in principle, via Monte Carlo, the computational demands of the propagator make direct Monte Carlo infeasible. This talk surveys current approaches for estimating collision probabilities, and suggests an approach based on importance sampling.

Statistical Emulation of Materials Science Computer Experiments
Shane Reese, Bringham Young University

Our society's economic progress depends critically on the high-performance materials such as lightweight alloys, high-energy-density battery materials, recyclable motor vehicle and building components, and energy-efficient lighting. Industrial growth areas depend, in part, on fundamental understanding of materials science and the atomic particle behavior. We discuss the role of statistical emulation of complex computational models of crystal structure in material properties. As opposed to Density Functional Theory, the Gaussian Process formulation allows rapid assessment of a wide variety of alloy combinations and prediction of important materials science properties. The methods are illustrated by applying the GP approach to two common systems and extensions to 700 other systems are discussed. Extensions to physically interpretable bases and sparseness prior implementations are presented and demonstrated to be faster and more feasible for large systems with small training datasets.

OEM for Big Data
Peter Z. Qian, University of Wisconsin

Big data with large sample size arise in internet, marketing, engineering and many other fields.

We propose an algorithm, called OEM (a.k.a., orthogonalizing EM), for analyzing big data. This algorithm employs a procedure, named active orthogonalization, to expand an arbitrary matrix to an orthogonal matrix. The orthogonality of the completed matrix yields closed-form solutions to ordinary and various penalized least squares problems. The maximum number of points required in active orthogonalization is bounded by the number of columns of the original matrix, which makes OEM particularly appealing for large n problems. Attractive theoretical properties of OEM include (1) providing a local solution to the SCAD or MCP with guaranteed oracle property, (2) convergence to the Moore-Penrose generalized inverse estimator for a singular regression matrix and (3) convergence to a point having grouping coherence for a fully aliased regression matrix. OEM can be modified to fit massive Gaussian process models in computer experiments and solve other singularity problems. The effectiveness of OEM will be illustrated through examples.

Accelerating Inference by Changing Designs: Sparse Grid Designs
Matthew Plumlee, Georgia Institute of Technology

Traditional analysis of a computer experiment experiences difficulty in high dimensions due to the computational burden of inference. This work shows by using a class of experimental designs, termed sparse grids, the computational cost of inference scales significantly better in high dimensions (exact prediction with over 250,000 design points is possible in seconds using only a laptop computer). Compared to the more common space filling Latin hypercube designs, sparse grid designs seem to be competitive in terms of mean square prediction error in both simulation and analytic results. Additionally, analysis of the Fisher information matrix demonstrates the advantages of using sparse grid designs for variable screening.

Contributed Session 8. High Throughput Screening and Optimal Designs

Plate Designs in High Throughput Screening Experiments
Xianggui Qu, Department of Mathematics and Statistics, Oakland University, MI 48309

High-throughput screening (HTS) is a large-scale process that screens hundreds of thousands to millions of compounds in order to identify potentially leading candidates rapidly and accurately. There are many statistically challenging issues in HTS. In this talk, I will focus the spatial effect in primary HTS. I will discuss why the current experimental design fails to eliminate these spatial effects. A new class of designs will be proposed for elimination of spatial effects. The new designs have the advantages such as all compounds are comparable within each microplate in spite of the existence of spatial effects; the maximum number of compounds in each microplate is attained, etc. Several designs are recommended for HTS experiments.

Robust Analysis of High Throughput Screening (HTS) Assay Data
Changwon Lim, Department of Mathematics and Statistics, Loyola University Chicago

Quantitative high throughput screening (qHTS) assays use cells or tissues to screen thousands of compounds in a short period of time. Data generated from qHTS assays are evaluated using nonlinear regression models such as the Hill model and decisions regarding toxicity are made using the estimates of the parameters of the model. For any given compound, the variability in the observed response may either be constant across dose groups (homoscedasticity) or vary with dose (heteroscedasticity). Secondly, since thousands of chemicals are evaluated simultaneously, it is not uncommon to find outliers and influential observations in the data. It is well-known that the variance structure and outliers play an important role in the analysis of nonlinear regression models. Since

thousands of chemicals are processed in a given run of the assay, it is necessary to develop and use methodology that is robust to variance structure and outliers so that a researcher does not have to perform model diagnostics for each individual model. In this talk we describe such a robust methodology and illustrate it using a data set obtained from the U.S. National Toxicology Program. Joint work with Pranad Sen, Shymal Peddada

Optimal Design for the M/M/1 Queue

B.M. Parker, University of Southampton, UK

Queues occur frequently in many areas of nature and technology; they are often used in, for example, transport modelling, but much current use of the mathematical properties of queues is made in modelling communications networks.

Whilst a lot of probabilistic results can be derived from queueing models, there has been limited research about the optimal times to measure queues in order to make inference about the parameters which determine the behaviour of these queues. In particular, in some applications (e.g. communications networks) measuring queues can require adding customers to the queue to act as survey customers. This has the effect of altering the future behaviour of the queue, and potentially changing the optimal measurement pattern of the queues: observations interfere with the experiment. We look in some detail at this interesting interfering case.

We examine the optimal design of measurements on queues with particular reference to the M/M/1 queue. Using the statistical theory of design of experiments, we calculate numerically the Fisher information matrix for an estimator of the arrival rate and the service rate to find optimal times to measure the queue when the number of measurements are limited for both interfering and non-interfering measurements. Our results show that, if measurements are intrusive, that measuring queues at a fixed interval is generally not optimal. We compute optimal designs for a variety of queuing situations and give results obtained under the D- and Ds- optimality criteria.

Finding D-optimal design for Multi-toxicant Poisson Model via Particle Swarm Optimization

Jiaheng Qiu, Department of Biostatistics, UCLA

Poisson regression models are widely used in modeling count data in social science, epidemiology and bioscience. In toxicity studies, Poisson models can be written as $y_{ij} \sim \text{Poisson}(\exp(x_i^T \beta))$, where y_{ij} is the number of organisms/cells that survive the experiment for the experiment for the jth replicate at ith design point and x_i is the regressor vector at the ith design point. D-optimal designs for estimating β or up to only two toxicants with a two way interaction term are available in the literature. We developed a version of the Particle Swarm Optimization methodology to find D-optimal designs for estimating the parameters in a 3, 4 or 5-toxicant Poisson model with and without two-way interactions when the induced design space is unrestricted and restricted and so generalize results from Wang et al. (2004) and Russell et al. (2009). We also performed simulation studies and provide guidelines for the choice of different tuning parameters in the PSO algorithm for searching such optimal designs. The sample codes can be found on the website http://www.math.ntu.edu.tw/~optdesign.