Technometrics, ISSN 0040-1706
Volume 61, number 1 (febrero 2019)

Abstract
We investigate the merits of replication, and provide methods for optimal design (including replicates), with the goal of obtaining globally accurate emulation of *noisy* computer simulation experiments. We first show that replication can be beneficial from both design and computational perspectives, in the context of Gaussian process surrogate modeling. We then develop a lookahead-based sequential design scheme that can determine if a new run should be at an existing input location (i.e., replicate) or at a new one (explore). When paired with a newly developed heteroscedastic Gaussian process model, our dynamic design scheme facilitates learning of signal and noise relationships which can vary throughout the input space. We show that it does so efficiently, on both computational and statistical grounds. In addition to illustrative synthetic examples, we demonstrate performance on two challenging real-data simulation experiments, from inventory management and epidemiology. Supplementary materials for the article are available online.

Abstract
To identify the robust settings of the control factors, it is very important to understand how they interact with the noise factors. In this article, we propose space-filling designs for computer experiments that are more capable of accurately estimating the control-by-noise interactions. Moreover, the existing space-filling designs focus on uniformly distributing the points in the design space, which are not suitable for noise factors because they usually follow nonuniform distributions such as normal distribution. This would suggest placing more points in the regions with high probability mass. However, noise factors also tend to have a smooth relationship with the response and therefore, placing more points toward the tails of the distribution is also useful for accurately estimating the relationship. These two opposing effects make the experimental design methodology a challenging problem. We propose optimal and computationally efficient solutions to this problem and demonstrate their advantages using simulated examples and a real industry example involving a manufacturing packing line. Supplementary materials for the article are available online.

Abstract
While literature on constructing efficient experimental designs has been plentiful, how best to incorporate prior information when assigning factors to the columns of a nonregular design has received little attention. Following Li, Zhou, and Zhang (2015) we propose the individual generalized word length pattern (iGWLP) for ranking columns of a nonregular design. Taking examples from the literature of recommended orthogonal arrays, we illustrate how iGWLP

helps to identify important differences in the aliasing that is likely otherwise missed. Given the complexity of characterizing partial aliasing for nonregular designs, iGWLP will help practitioners make more informed assignment of factors to columns when using nonregular fractions. We provide theoretical justification of the proposed iGWLP. A theorem is given to relate the proposed iGWLP criterion to the expected bias caused by model misspecifications. We also show that the proposed criterion may lead to designs having better projection properties in the factors considered most likely to be important. Furthermore, we discuss how iGWLP can be used for design selection. We propose a criterion for choosing best designs when the focus is on a small set of important factors, for which the aliasing of effects involving these factors is minimized.

Response Surface Optimization in the Presence of Internal Noise With Application to Optimal Alignment of Carbon Nanotubes P. 50-65

Sobambo Sosina, E. Marielle Remillard, Qiaoying Zhang, Chad Vecitis, and Tirthankar Dasgupta

Abstract

Internal noise, which means fluctuation of input factors around their set values, is common in many experiments in the physical and engineering sciences. Existing methods for response surface optimization in the presence of internal noise typically adopt a two-step approach: (a) fitting a response model as a function of the set value and (b) using Monte Carlo methods to account for internal noise while optimizing the response. In this article, motivated by a problem in optimizing alignment of carbon nanotubes (CNT), we propose a Bayesian approach for response surface optimization in the presence of internal noise. A unit-free and interpretable measure to quantify the strength of internal noise is proposed. Suitable objective functions or performance measures consistent with the overall goal of optimizing the response function are identified, methods for estimating them from available experimental data are suggested, and simulations are conducted to compare them with respect to their ability to account for internal noise in the optimization problem. The loss accrued by ignoring the internal noise in the optimization problem is quantified and studied via simulation. The proposed method is demonstrated through its application in the CNT alignment problem.

Sliced Rotated Sphere Packing Designs P. 66-76

Xu He

Abstract

Space-filling designs are popular choices for computer experiments. A sliced design is a design that can be partitioned into several subdesigns. We propose a new type of sliced space-filling design called sliced rotated sphere packing designs. Their full designs and subdesigns are rotated sphere packing designs. They are constructed by rescaling, rotating, translating, and extracting the points from a sliced lattice. We provide two fast algorithms to generate such designs. Furthermore, we propose a strategy to use sliced rotated sphere packing designs adaptively. Under this strategy, initial runs are uniformly distributed in the design space, follow-up runs are added by incorporating information gained from initial runs, and the combined design is space-filling for any local region. Examples are given to illustrate its potential application.

***d*-QPSO: A Quantum-Behaved Particle Swarm Technique for Finding *D*-Optimal Designs With Discrete and Continuous Factors and a Binary Response** P. 77-87

Joshua Lukemire,Abhyuday Mandal&Weng Kee Wong

Abstract

Identifying optimal designs for generalized linear models with a binary response can be a challenging task, especially when there are both discrete and continuous independent factors in the model. Theoretical results rarely exist for such models, and for the handful that do, they usually come with restrictive assumptions. In this article, we propose the *d*-

QPSO algorithm, a modified version of quantum-behaved particle swarm optimization, to find a variety of *D*-optimal approximate and exact designs for experiments with discrete and continuous factors and a binary response. We show that the *d*-QPSO algorithm can efficiently find locally *D*-optimal designs even for experiments with a large number of factors and robust pseudo-Bayesian designs when nominal values for the model parameters are not available. Additionally, we investigate robustness properties of the *d*-QPSO algorithm-generated designs to various model assumptions and provide real applications to design a bio-plastics odor removal experiment, an electronic static experiment, and a 10-factor car refueling experiment. Supplementary materials for the article are available online.

Most Recent Changepoint Detection in Panel Data P. 88-98

Lawrecen Bardwell,Paul Fearnhead ,Idris A. Eckley,Simon Smith &Martin Spott

Abstract

Detecting recent changepoints in time-series can be important for short-term prediction, as we can then base predictions just on the data since the changepoint. In many applications, we have panel data, consisting of many related univariate time-series. We present a novel approach to detect sets of most recent changepoints in such panel data that aims to pool information across time-series, so that we preferentially infer a most recent change at the same time-point in multiple series. Our approach is computationally efficient as it involves analysing each time-series independently to obtain a profile-likelihood like quantity that summarizes the evidence for the series having either no change or a specific value for its most recent changepoint. We then post-process this output from each time-series to obtain a potentially small set of times for the most recent changepoints, and, for each time, the set of series that has their most recent changepoint at that time. We demonstrate the usefulness of this method on two datasets: forecasting events in a telecommunications network and inference about changes in the net asset ratio for a panel of US firms.

Linear Hypothesis Testing With Functional Data P. 99-110

Łukasz Smaga and Jin-Ting Zhang

Abstract

In real data analysis, it is often interesting to consider a general linear hypothesis testing (GLHT) problem for functional data, which includes the one-way ANOVA, post hoc, or contrast analysis as special cases. Existing tests for this GLHT problem include an *L2*-norm-based test and an *F*-type test but their theoretical properties have not been investigated. In addition, for functional one-way ANOVA, simulation studies in the literature indicate that they are less powerful than the globalizing pointwise *F* (GPF) test and the *F*max -test. The GPF and *F*max -test enjoy several other good properties. They are scale-invariant in the sense that their test statistics do not change if we multiply each of functional curves with a nonzero function of the observed locations. In this article, the GPF and *F*max -test are adapted to the above GLHT problem. Their theoretical properties, for example, root-*n* consistency as well as those of the *L2*-norm-based and *F*-type tests are established. Intensive simulation studies are carried out to compare the finite-sample behavior of the tests under consideration in scenarios reflecting various practical characteristics of functional data. Simulation results indicate that the GPF test has higher power than other tests when the functional data are less correlated, and the *F*max -test has higher power than other tests when the functional data are moderately or highly correlated. These results are also confirmed by application of the GPF and *F*max tests to the corneal surface data coming from medical industry. This application suggests the new methods may help to make more clear and sure decisions in practice. For a convenient application of the considered testing procedures, their implementation is developed in the R programming language. Supplementary materials for the article are available online.

Robust Estimation Using Modified Huber's Functions With New Tails P. 111-122

Yunlu Jiang,You-Gan Wang,Liya Fu &Xueqin Wang

Abstract

It is traditionally believed that robustness is obtained by sacrificing efficiency. Estimators with high breakdown point

and high efficiency are therefore highly desirable. We investigate a new estimation procedure based on Huber's robust approach, but with tail functions replaced by the exponential squared loss. The tuning parameters are data-dependent to achieve high efficiency even in nonnormal cases. In the regression framework, we show that our hybrid estimator is of high efficiency, reaching the highest asymptotic breakdown point of 50%. We have also established the$n-\!\!-\!\!\sqrt{}n$-consistency and asymptotic normality of our estimator under regularity conditions. Extensive numerical studies are carried out to compare the performances of our method and other existing methods in terms of the standard errors and relative efficiency, and the results reveal that the newly proposed method has smaller standard errors and higher relative efficiency than its competitors when the sample size is sufficiently large. Finally, we present three real examples for demonstration. Supplementary materials for the article are available online.

Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions P. 123-135

Anna Kiriliouk,Holger Rootzén,Johan Segers&Jennifer L. Wadsworth

Abstract

When assessing the impact of extreme events, it is often not just a single component, but the combined behavior of several components which is important. Statistical modeling using multivariate generalized Pareto (GP) distributions constitutes the multivariate analogue of univariate peaks over thresholds modeling, which is widely used in finance and engineering. We develop general methods for construction of multivariate GP distributions and use them to create a variety of new statistical models. A censored likelihood procedure is proposed to make inference on these models, together with a threshold selection procedure, goodness-of-fit diagnostics, and a computationally tractable strategy for model selection. The models are fitted to returns of stock prices of four UK-based banks and to rainfall data in the context of landslide risk estimation. Supplementary materials and codes are available online.