Technometrics, ISSN 0040-1706
Volume 60, number 3 (august 2018)

Abstract
We consider Bayesian design of experiments problems in which we maximize the prior expectation of a utility function over a set of permutations, for example, when sequencing a number of tasks to perform. When the number of tasks is large and the expected utility is expensive to compute, it may be unreasonable or infeasible to evaluate the expected utility of all permutations. We propose an approach to emulate the expected utility using a surrogate function based on a parametric probabilistic model for permutations. The surrogate function is fitted by maximizing the correlation with the expected utility over a set of training points. We propose a suitable transformation of the expected utility to improve the fit. We provide results linking the correlation between the two functions and the number of expected utility evaluations to undertake. The approach is applied to the sequencing of reliability growth tasks in the development of hardware systems, in which there are a large number of potential tasks to perform and engineers are interested in meeting a reliability target subject to minimizing costs and time. An illustrative example shows how the approach can be used and a simulation study demonstrates the performance of the approach more generally. Supplementary materials for this article are available online.

Abstract
The calibration of computer models using physical experimental data has received a compelling interest in the last decade. Recently, multiple works have addressed the functional calibration of computer models, where the calibration parameters are functions of the observable inputs rather than taking a set of fixed values as traditionally treated in the literature. While much of the recent work on functional calibration was focused on estimation, the issue of sequential design for functional calibration still presents itself as an open question. Addressing the sequential design issue is thus the focus of this article. We investigate different sequential design approaches and show that the simple separate design approach has its merit in practical use when designing for functional calibration. Analysis is carried out on multiple simulated and real-world examples.

Abstract
Exact G-optimal designs have rarely, if ever, been employed in practical applications. One reason for this is that, due to the computational difficulties involved, no statistical software system currently provides capabilities for constructing them. Two algorithms for exact G-optimal design construction of small designs involving one to three factors have been discussed in the literature: one employing a genetic algorithm and one employing a coordinate-exchange algorithm.

However, these algorithms are extremely computer intensive in small experiments and do not scale beyond two or three factors. In this article, we develop a new method for constructing exact G-optimal designs using the integrated variance criterion, I_λ-optimality. We show that with careful selection of the weight function, a difficult exact G-optimal design construction problem can be converted to an equivalent exact I_λ-optimal design problem, which is easily and quickly solved. We illustrate the use of the algorithm for full quadratic models in one to five factors. The MATLAB codes used to implement our algorithm and the exact G-optimal designs produced by the algorithm for each test case are available online as supplementary material. imal designs produced by the algorithm for each test case are available online as supplementary material.

Bayesian Spatial Multivariate Receptor Modeling for Multisite Multipollutant Data P. 306-318

Eun Sug Park, Philip K. Hopke, Inyoung Kim, Shuman Tan & Clifford H. Spiegelman

Abstract

For the development of effective air pollution control strategies, it is crucial to identify the sources that are the principal contributors to air pollution and estimate how much each source contributes. Multivariate receptor modeling aims to address these problems by decomposing ambient concentrations of multiple air pollutants into components associated with different source types. With the expanded monitoring efforts that have been established over the past several decades, extensive multivariate air pollution data obtained from multiple monitoring sites (multisite multipollutant data) are now available. Although considerable research has been conducted on modeling multivariate space-time data in other contexts, there has been little research on spatial multivariate receptor models for multisite, multipollutant data. We present a Bayesian spatial multivariate receptor modeling (BSMRM) approach that can incorporate spatial correlations in multisite, multipollutant data into the estimation of source composition profiles and contributions, based on discrete process convolution models for multivariate spatial processes. The new BSMRM approach enables predictions of source contributions at unmonitored sites as well as simultaneously dealing with model uncertainty caused by the unknown number of sources and identifiability conditions. The new approach can also provide uncertainty estimates for the predicted source contributions at any location, which was not possible in previous multivariate receptor modeling approaches. The proposed approach is applied to 24-hour ambient air concentrations of 17 Volatile Organic Compounds (VOCs) measured at nine monitoring sites in Harris County, Texas, between 2003 and 2005. Supplementary materials for this article, including real data and MATLAB codes for implementing BSMRM, are available online on the journal web site.

ADMM for High-Dimensional Sparse Penalized Quantile Regression P. 319-331

Yuwen Gu, Jun Fan, Lingchen Kong, Shiqian Ma & Hui Zou

Abstract

Sparse penalized quantile regression is a useful tool for variable selection, robust estimation, and heteroscedasticity detection in high-dimensional data analysis. The computational issue of the sparse penalized quantile regression has not yet been fully resolved in the literature, due to nonsmoothness of the quantile regression loss function. We introduce fast alternating direction method of multipliers (ADMM) algorithms for computing the sparse penalized quantile regression. The convergence properties of the proposed algorithms are established. Numerical examples demonstrate the competitive performance of our algorithm: it significantly outperforms several other fast solvers for high-dimensional penalized quantile regression. Supplementary materials for this article are available online.

Directional Statistics of Preferential Orientations of Two Shapes in Their Aggregate and Its Application to Nanoparticle Aggregation P. 332-344

Ali Esmaieeli Sikaroudi, David A. Welch, Taylor J. Woehl, Roland Faller, James E. Evans, Nigel D. Browning & Chiwoo Park

Abstract
Nanoscientists have long conjectured that adjacent nanoparticles aggregate with one another in certain preferential directions during a chemical synthesis of nanoparticles, which is referred to the oriented attachment. For the study of the oriented attachment, the microscopy and nanoscience communities have used dynamic electron microscopy for direct observations of nanoparticle aggregation and have been so far relying on manual and qualitative analysis of the observations. We propose a statistical approach for studying the oriented attachment quantitatively with multiple aggregation examples in imagery observations. We abstract an aggregation by an event of two primary geometric objects merging into a secondary geometric object. We use a point set representation to describe the geometric features of the primary objects and the secondary object, and formulated the alignment of two point sets to one point set to estimate the orientation angles of the primary objects in the secondary object. The estimated angles are used as data to estimate the probability distribution of the orientation angles and test important hypotheses statistically. The proposed approach was applied for our motivating example, which demonstrated that nanoparticles of certain geometries have indeed preferential orientations in their aggregates.

Abstract
In manufacturing industries, images are commonly used for quality control purposes. In such applications, if the quality of the products is good, then their images should be all similar to the image of a good-quality product. Therefore, comparison of images is a fundamental task in image-based quality control. This problem, however, is complicated in the sense that (1) observed images often contain noise, and (2) the related images need to be geometrically matched up first because images of different products could be geometrically mismatched because the relative positions between a camera and different products are often not exactly the same. The first issue requires a statistical method that can remove noise, and the second issue is related to the so-called image registration problem in the image processing literature. In this article, we propose effective methods for detecting difference between two images of products, and our proposed methods can accommodate both noise and geometric mismatch mentioned above. Theoretical results and numerical examples show that they can work effectively in applications.

Abstract
Monitoring multichannel profiles has important applications in manufacturing systems improvement, but it is nontrivial to develop efficient statistical methods because profiles are high-dimensional functional data with intrinsic inner- and interchannel correlations, and that the change might only affect a few unknown features of multichannel profiles. To tackle these challenges, we propose a novel thresholded multivariate principal component analysis (PCA) method for multichannel profile monitoring. Our proposed method consists of two steps of dimension reduction: It first applies the functional PCA to extract a reasonably large number of features under the in-control state, and then uses the soft-thresholding techniques to further select significant features capturing profile information under the out-of-control state. The choice of tuning parameter for soft-thresholding is provided based on asymptotic analysis, and extensive numerical studies are conducted to illustrate the efficacy of our proposed thresholded PCA methodology.

Abstract
The ability to detect incipient and critical changes in real world process—esessential for system integrity assurance—is

currently impeded by the mismatch between the key assumption of stationarity underlying most change detection methods and the nonlinear and nonstationary (transient) dynamics of most real-world processes. The current approaches are slow or outright unable to detect qualitative changes in the behaviors that lead to anomalies. We present a Dirichlet process Gaussian state machine (DPGSM) model to represent dynamic intermittency, which is one of the most ubiquitous real-world transient behaviors. The DPGSM model treats a signal as a random walk among a Dirichlet process mixture of Gaussian clusters. Hypothesis tests and a numerical scheme based on this nonparametric representation were developed to detect subtle changes in the transient (intermittent) dynamics. Experimental investigations suggest that the DPGSM approach can consistently detect incipient, critical changes in intermittent signals some 50–2000 ms (20–90%) ahead of competing methods in benchmark test cases as well as a variety of real-world applications, such as in alternation patterns (e.g., ragas) in a music piece, and in the vibration signals capturing the initiation of product defects in an ultraprecision manufacturing process. A supplementary file to this article, available online, includes a Matlab implementation of the presented DPGSM.

Nonparametric Dynamic Curve Monitoring P. 386-397

Peihua Qiu, Xuemin Zi & Changliang Zou

Abstract

Rapid sequential comparison between the longitudinal pattern of a given subject and a target pattern has become increasingly important in modern scientific research for detecting abnormal activities in many data-rich applications. This article focuses on this problem when observations are collected sequentially with uncorrelated or correlated noise involved. A dynamic monitoring procedure is developed after connecting the curve monitoring problem to curve comparison. Under the framework of generalized likelihood ratio testing, we suggest a new exponentially weighted moving average (EWMA) control chart that can accommodate unequally spaced design points. An adaptive parameter selection feature is built in the proposed control chart so that the chart can detect a wide range of longitudinal pattern shifts effectively. To furnish fast computation, recursive formulas are derived for computing the charting statistic. Numerical studies show that the proposed method can deliver a satisfactory performance, and it outperforms existing methods in various cases. An example from the semiconductor manufacturing industry is used for the illustration of its implementation. Supplementary materials for this article are available online.

Multi-Parameter One-Sided Monitoring Tests P. 398-407

Guangyu Zhu & Jiahua Chen

Abstract

Multi-parameter one-sided hypothesis test problems arise naturally in many applications. We are particularly interested in effective tests for monitoring multiple quality indices in forestry products. Our search reveals that there are many effective statistical methods in the literature for normal data, and that they can easily be used to test hypotheses regarding parameter values permitting asymptotically normal estimators. We find that the classical likelihood ratio test is unsatisfactory, because to control the size, it must cope with the least favorable distributions at the cost of power. In this article, we find a novel way to slightly ease the size control, obtaining a much more powerful test. Simulation confirms that the new test retains good control of the Type I error and is markedly more powerful than the likelihood ratio test as well as many competitors based on normal data. The new method performs well in the context of monitoring multiple quality indices.