Technometrics, ISSN 0040-1706
Volume 57, number 1 (february 2015)

Nonparametric Variable Transformation in Sufficient Dimension Reduction P. 1-10
Qing Mai - Hui Zou

Abstract
Sufficient dimension reduction (SDR) techniques have proven to be very useful data analysis tools in various applications. Underlying many SDR techniques is a critical assumption that the predictors are elliptically contoured. When this assumption appears to be wrong, practitioners usually try variable transformation such that the transformed predictors become (nearly) normal. The transformation function is often chosen from the log and power transformation family, as suggested in the celebrated Box–Cox model. However, any parametric transformation can be too restrictive, causing the danger of model misspecification. We suggest a nonparametric variable transformation method after which the predictors become normal. To demonstrate the main idea, we combine this flexible transformation method with two well-established SDR techniques, sliced inverse regression (SIR) and inverse regression estimator (IRE). The resulting SDR techniques are referred to as TSIR and TIRE, respectively. Both simulation and real data results show that TSIR and TIRE have very competitive performance. Asymptotic theory is established to support the proposed method. The technical proofs are available as supplementary materials.

Simultaneous Envelopes for Multivariate Linear Regression P. 11-25
R. Dennis Cook - Xin Zhang

Abstract
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, canonical correlation analysis, and reduced-rank regression. This article has supplementary material available online.

Dynamic Retrospective Regression for Functional Data P. 26-34
Daniel Gervini

Abstract
Samples of curves, or functional data, usually present phase variability in addition to amplitude variability. Existing functional regression methods do not handle phase variability in an efficient way. In this article we propose a functional regression method that incorporates phase synchronization as an intrinsic part of the model, and then attains better predictive power than ordinary linear regression in a simple and parsimonious way. The finite-sample properties of the estimators are studied by simulation. As an example of application, we analyze neuromotor data arising from a study of human lip movement. This article has supplementary materials online.

Analysis of Computer Experiments With Functional Response P. 35-44
Ying Hung - V. Roshan Joseph - Shreyes N. Melkote

Abstract
This article is motivated by a computer experiment conducted for optimizing residual stresses in the machining of metals. Although kriging is widely used in the analysis of computer experiments, it cannot be easily applied to model the residual stresses because they are obtained as a profile. The high dimensionality caused by this functional response introduces severe computational challenges in kriging. It is well known that if the functional data are observed on a regular grid, the computations can be simplified using an application of Kronecker products. However, the case of irregular grid is quite complex. In this article, we develop a Gibbs sampling-based expectation maximization algorithm, which converts the irregularly spaced data into a regular grid so that the Kronecker product-based approach can be employed for efficiently fitting a kriging model to the functional data. Supplementary materials are available online.

Physical Experimental Design in Support of Computer Model Development P. 45-53
Max D. Morris

Abstract
Computer models of physical systems are often written based on known theory or "first principles" of a system, reflecting substantial knowledge of each component or subsystem, but also the need to use a numerical approach to mimic the more complex behavior of the entire system of interest. However, in some cases, there is insufficient known theory to encode all necessary aspects of the system, and empirical studies are required to generate approximate functional forms. We consider the question of how a physical experiment might be designed to approximate one module or subroutine of a computer model that can otherwise be written from first principles. The concept of preposterior analysis is used to suggest an approach to generating a kind of *I*-optimal design for this purpose, when the remainder of the computer model is a composition of nonlinear functions that can be directly evaluated as part of the design process. Extensions are then described for situations in which one or more known components must themselves be approximated by metamodels due to the large number of evaluations needed, and for computer models that have iterative structure. A simple "toy" model is used to demonstrate the ideas. Online supplementary material accompanies this article.

Approximate Model Spaces for Model-Robust Experiment Design P. 54-63
Byran J. Smucker & Nathan M. Drew

Abstract
Optimal designs depend upon a prespecified model form. A popular and effective model-robust alternative is to design with respect to a set of models instead of just one. However, model spaces associated with experiments of interest are often prohibitively large and so algorithmically generated designs are infeasible. Here, we present a simple method that largely eliminates this problem by choosing a small set of models that approximates the full set and finding designs that are explicitly robust for this small set. We build our procedure on a restricted columnwise-pairwise algorithm, and explore its effectiveness for two model spaces in the literature. For smaller full model spaces, we find that the designs constructed with the new method compare favorably with robust designs that use the full model space, with construction times reduced by orders of magnitude. We also construct designs that heretofore have been unobtainable due to the size of their model spaces. Supplementary material (available online) includes code, designs, and additional results.

Sequential Exploration of Complex Surfaces Using Minimum Energy Designs P. 64-74
V. Roshan Joseph - Tirthankar Dasgupta - Rui Tuo - C. F. Jeff Wu

Abstract
A new space-filling design, called minimum energy design (MED), is proposed to explore unknown regions of the

design space of particular interest to an experimenter. The key ideas involved in constructing the MED are the visualization of each design point as a charged particle inside a box, and minimization of the total potential energy of these particles. It is shown through theoretical arguments and simulations that with a proper choice of the charge function, the MED can asymptotically generate any arbitrary probability density function. A version of the MED, which adaptively updates the design by "learning" about the unknown response surface sequentially, is proposed and implemented. Two potential applications of MED in simulation of complex probability densities and optimization of complex response surfaces are discussed and demonstrated with examples. This article has supplementary material online.

Model-Based Sampling Design for Multivariate Geostatistics P. 75-86

Jie Li - Dale L. Zimmerman

Abstract

The quality of inferences made from geostatistical data is affected significantly by the spatial locations, or *design*, of the sites that are sampled. A large body of published work exists on sampling design for univariate geostatistics, but not for multivariate geostatistics. This article considers multivariate spatial sampling design based on criteria targeted at classical co-kriging (prediction with known covariance parameters), estimation of covariance (including cross-covariance) parameters, and empirical co-kriging (prediction with estimated covariance parameters). Through a combination of analytical results and examples, we investigate the characteristics of optimal designs with respect to each criterion, addressing in particular the design's degree of collocation. We also consider the robustness of the optimal design to the strength of spatial correlation and cross-correlation; the effects of smoothness and/or separability of the sampled process on the optimal design; the relationship between optimal designs for the multivariate problems considered here and univariate problems considered previously; and the efficiency of optimal collocated designs. One key finding is that optimal collocated designs are highly efficient in many cases. Supplementary materials are available online.

Geodesic Gaussian Processes for the Parametric Reconstruction of a Free-Form Surface P. 87-99

Enrique del Castillo - Bianca M. Colosimo - Sam Davanloo Tajbakhsh

Abstract

Reconstructing a free-form surface from 3-dimensional (3D) noisy measurements is a central problem in inspection, statistical quality control, and reverse engineering. We present a new method for the statistical reconstruction of a free-form surface patch based on 3D point cloud data. The surface is represented parametrically, with each of the three Cartesian coordinates (x, y, z) a function of surface coordinates (u, v), a model form compatible with computer-aided-design (CAD) models. This model form also avoids having to choose one Euclidean coordinate (say, z) as a "response" function of the other two coordinate "locations" (say, x and y), as commonly used in previous Euclidean kriging models of manufacturing data. The (u, v) surface coordinates are computed using parameterization algorithms from the manifold learning and computer graphics literature. These are then used as locations in a spatial Gaussian process model that considers correlations between two points on the surface a function of their *geodesic* distance on the surface, rather than a function of their Euclidean distances over the xy plane. We show how the proposed geodesic Gaussian process (GGP) approach better reconstructs the true surface, filtering the measurement noise, than when using a standard Euclidean kriging model of the "heights", that is, $z(x, y)$. The methodology is applied to simulated surface data and to a real dataset obtained with a noncontact laser scanner. Supplementary materials are available online.

Inverse Gaussian Processes With Random Effects and Explanatory Variables for Degradation Data P. 100-111

Chien-Yu Peng

Abstract

Degradation models are widely used to assess the lifetime information of highly reliable products. This study proposes a degradation model based on an inverse normal-gamma mixture of an inverse Gaussian process. This article presents the properties of the lifetime distribution and parameter estimation using the EM-type algorithm, in addition to providing a simple model-checking procedure to assess the validity of different stochastic processes. Several case applications are performed to demonstrate the advantages of the proposed model with random effects and explanatory variables. Technical details, data, and R code are available online as supplementary materials.

Statistical Inference for Power-Law Process With Competing Risks P. 112-122

Anupap Somboonsavatdee - Ananda Sen

Abstract

The focus of this article is on failure history of a repairable system for which the relevant data comprise successive event times for a recurrent phenomenon along with an event-count indicator. We undertake an investigation for analyzing failures from repairable systems that are subject to multiple failure modes. Failure data representing a cluster of recurrent events from a single system are studied under the parametric framework of a *power-law process*, a model that has found considerable attention in industrial applications. Some interesting and nonstandard asymptotic results ensue in this context that are discussed in detail. Extensive simulation has been carried out that supplements the theoretical findings. An extension to the case where the specific cause of failure may be missing is investigated in detail. The methodology has been implemented on recurrent failure data obtained from a warranty claim database for a fleet of automobiles. Supplementary material for this article is available online.

Solving the MEG Inverse Problem: A Robust Two-Way Regularization Method P. 123-137

Siva Tian - Jianhua Z. Huang - Haipeng Shen

Abstract

Magnetoencephalography (MEG) is a common noninvasive imaging modality for instantly measuring whole brain activities. One challenge in MEG data analysis is how to minimize the impact of the outliers that commonly exist in the images. In this article, we propose a robust two-way regularization approach to solve the important MEG inverse problem, that is, reconstructing neuronal activities using the measured MEG signals. The proposed method is based on the distributed source model and produces a spatio-temporal solution for all the dipoles simultaneously. Unlike the traditional methods that use the squared error loss function, our proposal uses a robust loss function, which improves the robustness of the results against outliers. To impose desirable spatial focality and temporal smoothness, we then penalize the robust loss through appropriate spatial–temporal two-way regularization. Furthermore, an alternating reweighted least-squares algorithm is developed to optimize the penalized model fitting criterion. Extensive simulation studies and a real-world MEG study clearly demonstrate the advantages of the proposed method over three nonrobust methods.