Technometrics, ISSN 0040-1706
Volume 63, number 4 (2021)

Strategies for Supersaturated Screening: Group Orthogonal and Constrained Var(*s*) Designs P. 443-455

Maria L. Weese Jonathan W. Stallrich, Byran J. Smucker & David J. Edwards

Abstract
Despite the vast amount of literature on supersaturated designs (SSDs), there is a scant record of their use in practice. We contend this imbalance is due to conflicting recommendations regarding SSD use in the literature as well as the designs' inabilities to meet practitioners' analysis expectations. To address these issues, we first summarize practitioner concerns and expectations of SSDs as determined via an informal questionnaire. Next, we discuss and compare two recent SSDs that pair a design construction method with a particular analysis method. The choice of a design/analysis pairing is shown to depend on the screening objective. Group orthogonal SSDs, when paired with our new, modified analysis, are demonstrated to have high power even with many active factors. Constrained positive Var(*s*)-optimal designs, when paired with the Dantzig selector, are recommended when effect directions can be credibly specified in advance; this strategy reasonably controls Type I error rates while still identifying a high proportion of active factors.

Aggregate Inverse Mean Estimation for Sufficient Dimension Reduction P. 456-465

Qin Wang & Xiangrong Yin

Abstract

Many well-known sufficient dimension reduction methods investigate the inverse conditional moments of the predictors given the response. The required linearity condition, the number and arrangement of slices, and the inability to detect symmetric dependence are among several long-standing issues that have negatively impacted on the use of these approaches. Motivated by two recent works dealing with the choice of number of slices, we propose a novel and effective method based on the aggregation of inverse mean estimation. The new approach can substantially improve the estimation accuracy, break down the symmetry to achieve exhaustive estimation, and is much less sensitive to the violation of the linearity condition. Both simulation studies and a real data application show the efficacy of the newly proposed approach.

Elastic Depths for Detecting Shape Anomalies in Functional Data P. 466-476

Trevor Harris, J. Derek Tucker, Bo Li & Lyndsay Shand

Abstract
We propose a new family of depth measures called the elastic depths that can be used to greatly improve shape anomaly detection in functional data. Shape anomalies are functions that have considerably different geometric forms

or features from the rest of the data. Identifying them is generally more difficult than identifying magnitude anomalies because shape anomalies are often not distinguishable from the bulk of the data with visualization methods. The proposed elastic depths use the recently developed elastic distances to directly measure the centrality of functions in the amplitude and phase spaces. Measuring shape outlyingness in these spaces provides a rigorous quantification of shape, which gives the elastic depths a strong theoretical and practical advantage over other methods in detecting shape anomalies. A simple boxplot and thresholding method is introduced to identify shape anomalies using the elastic depths. We assess the elastic depth's detection skill on simulated shape outlier scenarios and compare them against popular shape anomaly detectors. Finally, we use hurricane trajectories to demonstrate the elastic depth methodology on manifold valued functional data.

Joint Models for Event Prediction From Time Series and Survival Data P. 477-486

Xubo Yue & Raed Al Kontar

Abstract
We present a nonparametric prognostic framework for individualized event prediction based on joint modeling of both time series and time-to-event data. Our approach exploits a multivariate Gaussian convolution process (MGCP) to model the evolution of time series signals and a Cox model to map time-to-event data with time series data modeled through the MGCP. Taking advantage of the unique structure imposed by convolved processes, we provide a variational inference framework to simultaneously estimate parameters in the joint MGCP-Cox model. This significantly reduces computational complexity and safeguards against model overfitting. Experiments on synthetic and real world data show that the proposed framework outperforms state-of-the art approaches built on two-stage inference and strong parametric assumptions. Technical details are available in the supplementary materials.

Template Priors in Bayesian Curve Registration P. 487-499

W. Zachary Horton, Garritt L. Page, C. Shane Reese, Lindsey K. Lepley & McKenzie White

Abstract

In experiments where observations on each experimental unit are functional in nature, it is often the case that, in addition to variability along the horizontal axis (height or amplitude variability), there are also lateral displacements/deformations in curves (referred to as phase variability). Unlike the former, the latter form of variability is often treated as a nuisance parameter when making inferences. Therefore, it is common in functional data analysis to reduce this variability by aligning curves through a process called curve registration. Often, expert knowledge regarding the location and time that certain curve features occur is available to guide the curve realignment. We propose a Bayesian model that permits incorporating this knowledge when registering curves using a Gaussian process prior formulation. This novel approach capitalizes on the interpolation property of predictive distributions from Gaussian processes while still preserving the flexibility found in modern registration techniques. We detail computational strategies and illustrate the utility of the method through a simulation study and an analysis of knee-power biomechanics. Supplementary materials the article are available online.

A Random Fourier Feature Method for Emulating Computer Models With Gradient Information P. 500-509

Tzu-Hsiang Hung & Peter Chien

Abstract
Computer models with gradient information are increasingly used in engineering and science. The gradient-enhanced

Gaussian process emulator can be used for emulating such models. Because the size of the covariance matrix increases proportionally with the dimension of inputs and the sample size, it is computationally challenging to fit such an emulator for large datasets. We propose a random Fourier feature method to mitigate this difficulty. The key idea of the proposed method is to employ random Fourier features to obtain an easily computable, low-dimensional feature representation for shift-invariant kernels involving gradients. The effectiveness of the proposed method is illustrated by several examples.

Bayesian Analysis of Multifidelity Computer Models With Local Features and Nonnested Experimental Designs: Application to the WRF Model P. 510-522

Bledar A. Konomi & Georgios Karagiannis

Abstract

Motivated by a multi-fidelity Weather Research and Forecasting (WRF) climate model application where the available simulations are not generated based on hierarchically nested experimental design, we develop a new co-kriging procedure called augmented Bayesian treed co-kriging. The proposed procedure extends the scope of co-kriging in two major ways. We introduce a binary treed partition latent process in the multifidelity setting to account for nonstationary and potential discontinuities in the model outputs at different fidelity levels. Moreover, we introduce an efficient imputation mechanism which allows the practical implementation of co-kriging when the experimental design is nonhierarchically nested by enabling the specification of semiconjugate priors. Our imputation strategy allows the design of an efficient reversible jump Markov chain Monte Carlo implementation that involves collapsed blocks and direct simulation from conditional distributions. We develop the Monte Carlo recursive emulator which provides a Monte Carlo proxy for the full predictive distribution of the model output at each fidelity level, in a computationally feasible manner. The performance of our method is demonstrated on benchmark examples and used for the analysis of a large-scale climate modeling application which involves the WRF model.

Assurance for Sample Size Determination in Reliability Demonstration Testing P. 523-535

Kevin J. Wilson & Malcolm Farrow

Abstract

Manufacturers are required to demonstrate that products meet reliability targets. A way to achieve this is with reliability demonstration tests (RDTs), where a number of products are put on test and the test is passed or failed according to a decision rule based on the observed outcomes. There are various methods for determining the sample size for RDTs, typically based on the power of a hypothesis test following the RDT or risk criteria. Bayesian risk criteria approaches combine the choice of sample size with the analysis of the test data while relying on the specification of acceptable and rejectable reliability levels. In this article, we offer an alternative approach to sample size determination based on the idea of assurance. This approach chooses the sample size to provide a specified probability that the RDT will result in a successful outcome. It separates the design and analysis of the RDT, allowing different priors for the producer and consumer. We develop the assurance approach for sample size calculations in RDTs for binomial and Weibull likelihoods and propose appropriate prior distributions for the design and analysis of the test. In each case, we illustrate the approach with an example based on real data.

Modality-Constrained Density Estimation via Deformable Templates P. 536-547

Sutanoy Dasgupta, Debdeep Pati, Ian H. Jermyn & Anuj Srivastava

Abstract

Estimation of a probability density function (pdf) from its samples, while satisfying certain shape constraints, is an important problem that lacks coverage in the literature. This article introduces a novel geometric, deformable template constrained density estimator (dtcode) for estimating pdfs constrained to have a given number of modes. Our approach explores the space of thus-constrained pdfs using the set of shape-preserving transformations: an arbitrary template from the given shape class is transformed via a shape-preserving transformation to obtain the final optimal estimate. The search for this optimal transformation, under the maximum-likelihood criterion, is performed by mapping transformations to the tangent space of a Hilbert sphere, where they are effectively linearized, and can be expressed using an orthogonal basis. This framework is first applied to (univariate) unconditional densities and then extended to conditional densities. We provide asymptotic convergence rates for dtcode, and an application of the framework to the speed distributions for different traffic flows on Californian highways.

Estimation of Spatial Deformation for Nonstationary Processes via Variogram Alignment

P. 548-561

Ghulam A. Qadir, Ying Sun & Sebastian Kurtek

Abstract

In modeling spatial processes, a second-order stationarity assumption is often made. However, for spatial data observed on a vast domain, the covariance function often varies over space, leading to a heterogeneous spatial dependence structure, therefore requiring nonstationary modeling. Spatial deformation is one of the main methods for modeling nonstationary processes, assuming the nonstationary process has a stationary counterpart in the deformed space. The estimation of the deformation function poses severe challenges. Here, we introduce a novel approach for nonstationary geostatistical modeling, using space deformation, when a single realization of the spatial process is observed. Our method is based on aligning regional variograms, where warping variability of the distance from each subregion explains the spatial nonstationarity. We propose to use multi-dimensional scaling to map the warped distances to spatial locations. We assess the performance of our new method using multiple simulation studies. Additionally, we illustrate our methodology on precipitation data to estimate the heterogeneous spatial dependence and to perform spatial predictions.