Technometrics, ISSN 0040-1706
Volume 57, number 2 (may 2015)

Abstract
Many industrial experiments involve some factors whose levels are harder to set than others. The best way to deal with these is to plan the experiment carefully as a split-plot, or more generally a multistratum, design. Several different approaches for constructing split-plot type response surface designs have been proposed in the literature since 2001, which has allowed experimenters to make better use of their resources by using more efficient designs than the classical balanced ones. One of these approaches, the stratum-by-stratum strategy has been shown to produce designs that are less efficient than locally *D*-optimal designs. An improved stratum-by-stratum algorithm is given, which, though more computationally intensive than the old one, makes better use of the advantages of this approach, that is, it can be used for any structure and does not depend on prior estimates of the variance components. This is shown to be almost as good as the locally optimal designs in terms of their own criteria and more robust across a range of criteria. Supplementary materials for this article are available online.

Abstract
For deterministic computer simulations, Gaussian process models are a standard procedure for fitting data. These models can be used only when the study design avoids having replicated points. This characteristic is also desirable for one-dimensional projections of the design, since it may happen that one of the design factors has a strongly nonlinear effect on the response. Latin hypercube designs have uniform one-dimensional projections, but are not efficient for fitting low-order polynomials when there is a small error variance. *D*-optimal designs are very efficient for polynomial fitting but have substantial replication in projections. We propose a new class of designs that bridge the gap between *D*-optimal designs and *D*-optimal Latin hypercube designs. These designs guarantee a minimum distance between points in any one-dimensional projection allowing for the fit of either polynomial or Gaussian process models. Subject to this constraint they are *D*-optimal for a prespecified model.

Abstract
Pipeline integrity is important because leaks can result in serious economic or environmental losses. Inspection information from a sample of locations along the pipeline can be used to estimate corrosion levels. The traditional parametric model method for this problem is to estimate parameters of a specified corrosion distribution and then to use these parameters to estimate the minimum thickness in a pipeline. Inferences using this method are, however, highly sensitive to the distributional assumption. Extreme value modeling provides a more robust method of estimation if a sufficient amount of data is available. For example, the block-minima method produces a more robust method to estimate the minimum thickness in a pipeline. To use the block-minima method, however, one must carefully choose

the size of the blocks to be used in the analysis. In this article, we use simulation to compare the properties of different models for estimating minimum pipeline thickness, investigate the effect of using different size blocks, and illustrate the methods using pipeline inspection data.

Statistical Methods for Degradation Data With Dynamic Covariates Information and an Application to Outdoor Weathering Data P. 180-193

Yili Hong, Yuanyuan Duan, William Q. Meeker, Deborah L. Stanley & Xiaohong Gu

Abstract

Degradation data provide a useful resource for obtaining reliability information for some highly reliable products and systems. In addition to product/system degradation measurements, it is common nowadays to dynamically record product/system usage as well as other life-affecting environmental variables, such as load, amount of use, temperature, and humidity. We refer to these variables as dynamic covariate information. In this article, we introduce a class of models for analyzing degradation data with dynamic covariate information. We use a general path model with individual random effects to describe degradation paths and a vector time series model to describe the covariate process. Shape-restricted splines are used to estimate the effects of dynamic covariates on the degradation process. The unknown parameters in the degradation data model and the covariate process model are estimated by using maximum likelihood. We also describe algorithms for computing an estimate of the lifetime distribution induced by the proposed degradation path model. The proposed methods are illustrated with an application for predicting the life of an organic coating in a complicated dynamic environment (i.e., changing UV spectrum and intensity, temperature, and humidity). This article has supplementary material online.

The Uncertainty of Storm Season Changes: Quantifying the Uncertainty of Autocovariance Changepoints P. 194-206

Christopher F. H. Nam, John A. D. Aston, Idris A. Eckley & Rebecca Killick

Abstract

In oceanography, there is interest in determining storm season changes for logistical reasons such as equipment maintenance scheduling. In particular, there is interest in capturing the uncertainty associated with these changes in terms of the number and location of them. Such changes are associated with autocovariance changes. This article proposes a framework to quantify the uncertainty of autocovariance changepoints in time series motivated by this oceanographic application. More specifically, the framework considers time series under the locally stationary wavelet (LSW) framework, deriving a joint density for scale processes in the raw wavelet periodogram. By embedding this density within a hidden Markov model (HMM) framework, we consider changepoint characteristics under this multiscale setting. Such a methodology allows us to model changepoints and their uncertainty for a wide range of models, including piecewise second-order stationary processes, for example, piecewise moving average processes.

Large-Vector Autoregression for Multilayer Spatially Correlated Time Series P. 207-216

Rodrigue Ngueyep & Nicoleta Serban

Abstract

One of the most commonly used methods for modeling multivariate time series is the vector autoregressive model (VAR). VAR is generally used to identify lead, lag, and contemporaneous relationships describing Granger causality within and between time series. In this article, we investigate the VAR methodology for analyzing data consisting of multilayer time series that are spatially interdependent. When modeling VAR relationships for such data, the dependence between time series is both a curse and a blessing. The former because it requires modeling the between time-series correlation or the contemporaneous relationships which may be challenging when using likelihood-based methods. The latter because the spatial correlation structure can be used to specify the lead–lag relationships within and between time series, within and between layers. To address these challenges, we propose an $L_1 \backslash L_2$ regularized likelihood estimation method. The lead, lag, and contemporaneous relationships are estimated using an efficient

algorithm that exploits sparsity in the VAR structure, accounts for the spatial dependence, and models the error dependence. We consider a case study to illustrate the applicability of our method. In the supplementary materials available online, we assess the performance of the proposed VAR model and compare it with existing methods within a simulation study.

Bayesian Estimation of Density via Multiple Sequential Inversions of Two-Dimensional Images With Application to Electron Microscopy P. 217-233

Dalia Chakrabarty, Nare Gabrielyan, Fabio Rigat, Richard Beanland & Shashi Paul

Abstract

We present a new Bayesian methodology to learn the unknown material density of a given sample by inverting its two-dimensional images that are taken with a scanning electron microscope. An image results from a sequence of projections of the convolution of the density function with the unknown microscopy correction function that we also learn from the data; thus, learning of the unknowns demands multiple inversions. We invoke a novel design of experiment, involving imaging at multiple values of the parameter that controls the subsurface depth from which information about the density structure is carried, to result in the image. Real-life material density functions are characterized by high-density contrasts and are highly discontinuous, implying that they exhibit correlation structures that do not vary smoothly. In the absence of training data, modeling such correlation structures of real material density functions is not possible. So we discretize the material sample and treat values of the density function at chosen locations inside it as independent and distribution-free parameters. Resolution of the available image dictates the discretization length of the model; three models pertaining to distinct resolution classes (at micrometer to nanometer scale lengths) are developed. We develop priors on the material density, such that these priors adapt to the sparsity inherent in the density function. The likelihood is defined in terms of the distance between the convolution of the unknown functions and the image data. The posterior probability density of the unknowns given the data is expressed using the developed priors on the density and priors on the microscopy correction function as elicited from the microscopy literature. We achieve posterior samples using an adaptive Metropolis-within-Gibbs inference scheme. The method is applied to learn the material density of a three-dimensional sample of a nano-structure, using real image data. Illustrations on simulated image data of alloy samples are also included.

Robust Optimization of Biological Protocols P. 234-244

Patrick Flaherty & Ronald W. Davis

Abstract

When conducting high-throughput biological experiments, it is often necessary to develop a protocol that is both inexpensive and robust. Standard approaches are either not cost-effective or arrive at an optimized protocol that is sensitive to experimental variations. Here, we describe a novel approach that directly minimizes the cost of the protocol while ensuring the protocol is robust to experimental variation. Our approach uses a risk-averse conditional value-at-risk criterion in a robust parameter design framework. We demonstrate this approach on a polymerase chain reaction protocol and show that our improved protocol is less expensive than the standard protocol and more robust than a protocol optimized without consideration of experimental variation.

CUSUM Statistical Monitoring of M/M/1 Queues and Extensions P. 245-256

Nan Chen & Shiyu Zhou

Abstract

Many production and service systems can be modeled as queueing systems. Their operational efficiency and performance are often measured using queueing performance metrics (QPMs), such as average cycle time, average waiting length, and throughput rate. These metrics need to be quantitatively evaluated and monitored in real time to continuously improve the system performance. However, QPMs are often highly stochastic, and hence are difficult to monitor using existing methods. In this article, we propose the cumulative sum (CUSUM) schemes to efficiently

monitor the performance of typical queueing systems based on different sampling schemes. We use M/M/1 queues to illustrate how to design the CUSUM chart and compare their performance with several alternative methods. We demonstrate that the performance of CUSUM is superior, responding faster to many shift patterns through extensive numerical studies. We also briefly discuss the extensions of CUSUM charts to more general queues, such as M/G/1, G/G/1, or M/M/c queues. We use case studies to demonstrate the applications of our approach. Supplementary materials for this article are available online.

Engineering-Driven Statistical Adjustment and Calibration P. 257-267

V. Roshan Joseph & Huan Yan

Abstract

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 for the discrepancies between the engineering model and observed 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 article proposes a new framework for model calibration and statistical adjustment. It tries to open up the black box using simple main effects analysis and graphical plots 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 micro-machining process. This article has supplementary material online.

Cook's Distance Measures for Varying Coefficient Models With Functional Responses P. 268-280

Qibing Gao, Mihye Ahn & Hongtu Zhu

Abstract

The aim of this article is to develop Cook's distance measures for assessing the influence of both atypical curves and observations under varying coefficient model with functional responses. Our Cook's distance measures include Cook's distances for deleting multiple curves and for deleting multiple grid points, and their scaled Cook's distances. We systematically investigate some theoretical properties of these diagnostic measures. Simulation studies are conducted to evaluate the finite sample properties of these Cook's distances under different scenarios. A real diffusion tensor tract dataset is analyzed to illustrate the use of our diagnostic measures.

The Mahalanobis Distance for Functional Data With Applications to Classification P. 281-291

Pedro Galeano, Esdras Joseph & Rosa E. Lillo

Abstract

This article presents a new semidistance for functional observations that generalizes the Mahalanobis distance for multivariate datasets. The main characteristics of the functional Mahalanobis semidistance are shown. To illustrate the applicability of this measure of proximity between functional observations, new versions of several well-known functional classification procedures are developed using the functional Mahalanobis semidistance. A Monte Carlo study and the analysis of two real examples indicate that the classification methods used in conjunction with the functional Mahalanobis semidistance give better results than other well-known functional classification procedures. This article has supplementary material online.