Technometrics, ISSN 0040-1706
Volume 60, number 1 (february 2018)

Abstract

We develop a supervised-learning-based approach for monitoring and diagnosing texture-related defects in manufactured products characterized by stochastic textured surfaces that satisfy the locality and stationarity properties of Markov random fields. Examples of stochastic textured surface data include images of woven textiles; image or surface metrology data for machined, cast, or formed metal parts; microscopy images of material microstructure samples; etc. To characterize the complex spatial statistical dependencies of in-control samples of the stochastic textured surface, we use rather generic supervised learning methods, which provide an implicit characterization of the joint distribution of the surface texture. We propose two spatial moving statistics, which are computed from residual errors of the fitted supervised learning model, for monitoring and diagnosing local aberrations in the general spatial statistical behavior of newly manufactured stochastic textured surface samples in a statistical process control context. We illustrate the approach using images of textile fabric samples and simulated two-dimensional stochastic processes, for which the algorithm successfully detects local defects of various natures. Supplemental discussions, results, data and computer codes are available online.

Abstract

With the rapid advancement of sensor technology, a huge amount of data is generated in various applications, which poses new and unique challenges for statistical process control (SPC). In this article, we propose a nonparametric adaptive sampling (NAS) strategy to online monitor nonnormal big data streams in the context of limited resources, where only a subset of observations are available at each acquisition time. In particular, this proposed method integrates a rank-based CUSUM scheme and an innovative idea that corrects the anti-rank statistics with partial observations, which can effectively detect a wide range of possible mean shifts when data streams are exchangeable and follow arbitrary distributions. Two theoretical properties on the sampling layout of the proposed NAS algorithm are investigated when the process is in control and out of control. Both simulations and case studies are conducted under different scenarios to illustrate and evaluate the performance of the proposed method. Supplementary materials for this article are available online.

Abstract

Expectile, first introduced by Newey and Powell in 1987 Newey, W. K., and Powell, J. L. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847.[Crossref], [Web of Science ®], [Google Scholar] in the

econometrics literature, has recently become increasingly popular in risk management and capital allocation for financial institutions due to its desirable properties such as coherence and elicitability. The current standard tool for expectile regression analysis is the multiple linear expectile regression proposed by Newey and Powell in 1987 Newey, W. K., and Powell, J. L. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847. [Crossref], [Web of Science ®], [Google Scholar]. The growing applications of expectile regression motivate us to develop a much more flexible nonparametric multiple expectile regression in a reproducing kernel Hilbert space. The resulting estimator is called KERE, which has multiple advantages over the classical multiple linear expectile regression by incorporating nonlinearity, nonadditivity, and complex interactions in the final estimator. The kernel learning theory of KERE is established. We develop an efficient algorithm inspired by majorization-minimization principle for solving the entire solution path of KERE. It is shown that the algorithm converges at least at a linear rate. Extensive simulations are conducted to show the very competitive finite sample performance of KERE. We further demonstrate the application of KERE by using personal computer price data. Supplementary materials for this article are available online.

Robust Lasso Regression Using Tukey's Biweight Criterion P. 36-47

Le Chang, Steven Roberts & Alan Welsh

Abstract

The adaptive lasso is a method for performing simultaneous parameter estimation and variable selection. The adaptive weights used in its penalty term mean that the adaptive lasso achieves the oracle property. In this work, we propose an extension of the adaptive lasso named the Tukey-lasso. By using Tukey's biweight criterion, instead of squared loss, the Tukey-lasso is resistant to outliers in both the response and covariates. Importantly, we demonstrate that the Tukey-lasso also enjoys the oracle property. A fast accelerated proximal gradient (APG) algorithm is proposed and implemented for computing the Tukey-lasso. Our extensive simulations show that the Tukey-lasso, implemented with the APG algorithm, achieves very reliable results, including for high-dimensional data where $p > n$. In the presence of outliers, the Tukey-lasso is shown to offer substantial improvements in performance compared to the adaptive lasso and other robust implementations of the lasso. Real-data examples further demonstrate the utility of the Tukey-lasso. Supplementary materials for this article are available online.

Complex-Valued Wavelet Lifting and Applications P. 48-60

Jean Hamilton, Matthew A. Nunes, Marina I. Knight, and Piotr Fryzlewicz

Abstract

Signals with irregular sampling structures arise naturally in many fields. In applications such as spectral decomposition and nonparametric regression, classical methods often assume a regular sampling pattern, thus cannot be applied without prior data processing. This work proposes new complex-valued analysis techniques based on the wavelet lifting scheme that removes "one coefficient at a time." Our proposed lifting transform can be applied directly to irregularly sampled data and is able to adapt to the signal(s)' characteristics. As our new lifting scheme produces complex-valued wavelet coefficients, it provides an alternative to the Fourier transform for irregular designs, allowing phase or directional information to be represented. We discuss applications in bivariate time series analysis, where the complex-valued lifting construction allows for coherence and phase quantification. We also demonstrate the potential of this flexible methodology over real-valued analysis in the nonparametric regression context. Supplementary materials for this article are available online.

Design for Sequential Follow-Up Experiments in Computer Emulations P. 61-69

Xiangshun Kong, Mingyao Ai & Kwok Leung Tsui

Abstract

Sequential experiments composed of initial experiments and follow-up experiments are widely adopted for economical

computer emulations. Many kinds of Latin hypercube designs with good space-filling properties have been proposed for designing the initial computer experiments. However, little work based on Latin hypercubes has focused on the design of the follow-up experiments. Although some constructions of nested Latin hypercube designs can be adapted to sequential designs, the size of the follow-up experiments needs to be a multiple of that of the initial experiments. In this article, a general method for constructing sequential designs of flexible size is proposed, which allows the combined designs to have good one-dimensional space-filling properties. Moreover, the sampling properties and a type of central limit theorem are derived for these designs. Several improvements of these designs are made to achieve better space-filling properties. Simulations are carried out to verify the theoretical results. Supplementary materials for this article are available online.

Pairwise Estimation of Multivariate Gaussian Process Models With Replicated Observations: Application to Multivariate Profile Monitoring P. 70-78

Yongxiang Li, Qiang Zhou, Xiaohu Huang & Li Zeng

Abstract

Profile monitoring is often conducted when the product quality is characterized by profiles. Although existing methods almost exclusively deal with univariate profiles, observations of multivariate profile data are increasingly encountered in practice. These data are seldom analyzed in the area of statistical process control due to lack of effective modeling tools. In this article, we propose to analyze them using the multivariate Gaussian process model, which offers a natural way to accommodate both within-profile and between-profile correlations. To mitigate the prohibitively high computation in building such models, a pairwise estimation strategy is adopted. Asymptotic normality of the parameter estimates from this approach has been established. Comprehensive simulation studies are conducted. In the case study, the method has been demonstrated using transmittance profiles from low-emittance glass. Supplementary materials for this article are available online.

A Conjugate Model for Dimensional Analysis P. 79-89

Weijie Shen & Dennis K. J. Lin

Abstract

Dimensional analysis (DA) is a methodology widely used in physics and engineering. The main idea is to extract key variables based on physical dimensions. Its overlooked importance in statistics has been recognized recently. However, most literature treats DA as merely a preprocessing tool, leading to multiple statistical issues. In particular, there are three critical aspects: (a) the nonunique choice of basis quantities and dimensionless variables; (b) the statistical representation and testing of DA constraints; (c) the spurious correlations between post-DA variables. There is an immediate need for an appropriate statistical methodology that integrates DA and the quantitative modeling. In this article, we propose a power-law type of "DA conjugate" model that is useful for incorporating dimensional information and analyzing post-DA variables. Adapting the similar idea of "conjugacy" in Bayesian analysis, we show that the proposed modeling technique not only produces flexible and effective results, but also provides good solutions to the above three issues. A modified projection pursuit regression analysis is implemented to fit the additive power-law model. A numerical study on ocean wave speed is discussed in detail to illustrate and evaluate the advantages of the proposed procedure. Supplementary materials for this article are available online.

Analysis of Field Return Data With Failed-But-Not-Reported Events P. 90-100

Xin Wang, Zhi-Sheng Ye, Yi-Li Hong & Loon-Ching Tang

Abstract

Warranty data contain valuable information on product field reliability and customer behaviors. Most previous studies on analysis of warranty data implicitly assume that all failures within the warranty period are reported and recorded. However, the failed-but-not-reported (FBNR) phenomenon is quite common for a product whose price is not very high.

Ignorance of the FBNR phenomenon leads to an overestimate of product reliability based on field return data or an overestimate of warranty cost based on lab data or tracking data. Being an indicator of customer satisfaction, the FBNR proportion provides valuable managerial insights. In this study, statistical inference for the FBNR phenomenon as well as field lifetime distribution is described. We first propose a flexible FBNR function to model the time-dependent FBNR behavior. Then, a framework for data analysis is developed. In the framework, both semiparametric and parametric approaches are used to jointly analyze warranty claim data and supplementary tracking data from a follow-up of selected customers. The FBNR problem in the tracking data is minimal and thus the data can be used to effectively decouple the FBNR information from the warranty claim data. The proposed methods are illustrated with an example. Supplementary materials for this article are available online.

Some Multivariate Tests of Independence Based on Ranks of Nearest Neighbors P. 101-111

Soham Sarkar & Anil K. Ghosh

Abstract

Several parametric and nonparametric tests of independence between two random vectors are available in the literature. But, many of them perform poorly for high-dimensional data and are not applicable when the dimension exceeds the sample size. In this article, we propose some tests based on ranks of nearest neighbors, which can be conveniently used in high dimension, low sample size situations. Several simulated and real datasets are analyzed to show the utility of the proposed tests. Codes for implementation of the proposed tests are available as supplementary materials.

A Unified Analysis of Structured Sonar-Terrain Data Using Bayesian Functional Mixed Models P. 112-123

Hongxiao Zhu, Philip Caspers, Jeffrey S. Morris, Xiaowei Wu & Rolf Müller

Abstract

Sonar emits pulses of sound and uses the reflected echoes to gain information about target objects. It offers a low cost, complementary sensing modality for small robotic platforms. Although existing analytical approaches often assume independence across echoes, real sonar data can have more complicated structures due to device setup or experimental design. In this article, we consider sonar echo data collected from multiple terrain substrates with a dual-channel sonar head. Our goals are to identify the differential sonar responses to terrains and study the effectiveness of this dual-channel design in discriminating targets. We describe a unified analytical framework that achieves these goals rigorously, simultaneously, and automatically. The analysis was done by treating the echo envelope signals as functional responses and the terrain/channel information as covariates in a functional regression setting. We adopt functional mixed models that facilitate the estimation of terrain and channel effects while capturing the complex hierarchical structure in data. This unified analytical framework incorporates both Gaussian models and robust models. We fit the models using a full Bayesian approach, which enables us to perform multiple inferential tasks under the same modeling framework, including selecting models, estimating the effects of interest, identifying significant local regions, discriminating terrain types, and describing the discriminatory power of local regions. Our analysis of the sonar-terrain data identifies time regions that reflect differential sonar responses to terrains. The discriminant analysis suggests that a multi- or dual-channel design achieves target identification performance comparable with or better than a single-channel design. Supplementary materials for this article are available online.