Technometrics, ISSN 0040-1706
Volume 62, number 2 (may 2020)

Abstract

Exposure assessment models are deterministic models derived from physical–chemical laws. In real workplace settings, chemical concentration measurements can be noisy and indirectly measured. In addition, inference on important parameters such as generation and ventilation rates are usually of interest since they are difficult to obtain. In this article, we outline a flexible Bayesian framework for parameter inference and exposure prediction. In particular, we devise Bayesian state space models by discretizing the differential equation models and incorporating information from observed measurements and expert prior knowledge. At each time point, a new measurement is available that contains some noise, so using the physical model and the available measurements, we try to obtain a more accurate state estimate, which can be called filtering. We consider Monte Carlo sampling methods for parameter estimation and inference under nonlinear and non-Gaussian assumptions. The performance of the different methods is studied on computer-simulated and controlled laboratory-generated data. We consider some commonly used exposure models representing different physical hypotheses.

Abstract

Water pollution is a major global environmental problem, and it poses a great environmental risk to public health and biological diversity. This work is motivated by assessing the potential environmental threat of coal mining through increased sulfate concentrations in river networks, which do not belong to any simple parametric distribution. However, existing network models mainly focus on binary or discrete networks and weighted networks with known parametric weight distributions. We propose a principled nonparametric weighted network model based on exponential-family random graph models and local likelihood estimation, and study its model-based clustering with application to large-scale water pollution network analysis. We do not require any parametric distribution assumption on network weights. The proposed method greatly extends the methodology and applicability of statistical network models. Furthermore, it is scalable to large and complex networks in large-scale environmental studies. The power of our proposed methods is demonstrated in simulation studies and a real application to sulfate pollution network analysis in Ohio watershed located in Pennsylvania, United States.

Abstract

The development of mobile network technologies has made it possible to collect location data of mobile devices through various positioning technologies. The location data can be used to estimate the spatial density of mobile devices, which in turn can be used by mobile service providers to plan for network capacity improvements. The two most prevalent positioning technologies are the assisted global positioning system (AGPS) and cell tower triangulation (CTT) methods. AGPS data provide more accurate location information than CTT data but can cover only a fraction of mobile devices, while CTT data can cover all mobile devices. Motivated by this problem, we propose a Bayesian nonparametric mixture measurement error model to estimate the spatial density function by integrating both noise-free data (i.e., AGPS data) and data contaminated with measurement errors (i.e., CTT data). The proposed model estimates the true latent locations from contaminated data, and the estimated latent locations, combined with noise-free data, are used to infer the model parameters. We model the true density function using a Dirichlet process (DP) mixture model with a bivariate beta distribution for the mixture kernel and a DP prior for the mixing distribution. The use of bivariate beta distributions for the mixture kernel allows the density function to have various shapes with a bounded support. Moreover, the use of a DP prior for the mixing distribution allows the number of mixture components to be determined automatically without being specified in advance. Therefore, the proposed model is very flexible. We demonstrate the effective performance of the proposed model via simulated and real-data examples.

Modeling and Change Detection for Count-Weighted Multilayer Networks P. 184-195

Hang Dong, Nan Chen & Kaibo Wang

Abstract

In a typical network with a set of individuals, it is common to have multiple types of interactions between two individuals. In practice, these interactions are usually sparse and correlated, which is not sufficiently accounted for in the literature. This article proposes a multilayer weighted stochastic block model (MZIP-SBM) based on a multivariate zero-inflated Poisson (MZIP) distribution to characterize the sparse and correlated multilayer interactions of individuals. A variational-EM algorithm is developed to estimate the parameters in this model. We further propose a monitoring statistic based on the score test of MZIP-SBM model parameters for change detection in multilayer networks. The proposed model and monitoring scheme are validated using extensive simulation studies and the case study from Enron E-mail network.

Matrix Linear Discriminant Analysis P. 196-205

Wei Hu, Weining Shen, Hua Zhou & Dehan Kong

Abstract

We propose a novel linear discriminant analysis (LDA) approach for the classification of high-dimensional matrix-valued data that commonly arises from imaging studies. Motivated by the equivalence of the conventional LDA and the ordinary least squares, we consider an efficient nuclear norm penalized regression that encourages a low-rank structure. Theoretical properties including a nonasymptotic risk bound and a rank consistency result are established. Simulation studies and an application to electroencephalography data show the superior performance of the proposed method over the existing approaches.

Analysis of Large Heterogeneous Repairable System Reliability Data With Static System Attributes and Dynamic Sensor Measurement in Big Data Environment P. 206-222

Xiao Liu & Rong Pan

Abstract

In the age of Big Data, one pressing challenge facing engineers is to perform reliability analysis for a large fleet of heterogeneous repairable systems with covariates. In addition to static covariates, which include time-invariant system attributes such as nominal operating conditions, geo-locations, etc., the recent advances of sensing technologies have also made it possible to obtain dynamic sensor measurement of system operating and environmental conditions. As a common practice in the Big Data environment, the massive reliability data are typically stored in some distributed

storage systems. Leveraging the power of modern statistical learning, this article investigates a statistical approach which integrates the random forests algorithm and the classical data analysis methodologies for repairable system reliability, such as the nonparametric estimator for the mean cumulative function and the parametric models based on the nonhomogeneous Poisson process. We show that the proposed approach effectively addresses some common challenges arising from practice, including system heterogeneity, covariate selection, model specification and data locality due to the distributed data storage. The large sample properties as well as the uniform consistency of the proposed estimator are established. Two numerical examples and a case study are presented to illustrate the application of the proposed approach. The strengths of the proposed approach are demonstratted by comparison studies. Datasets and computer code have been made available on GitHub.

Student-*t* Processes for Degradation Analysis P. 223-235

Chien-Yu Peng & Ya-Shan Cheng

Abstract

Stochastic processes are widely used to analyze degradation data, and the Gaussian process is a particularly common one. In this article, we propose a robust statistical model using a Student-*t* process to assess the lifetime information of highly reliable products. This model is statistically plausible and demonstrates a substantially improved fit when applied to real data. A computationally accurate approach is proposed to calculate the first-passage-time density function of the Student-*t* degradation-based process; related properties are investigated as well. In addition, this article provides parameter estimation using the EM-type algorithm and a simple model-checking procedure to evaluate the appropriateness of the model assumptions. Several case studies are performed to demonstrate the flexibility and applicability of the proposed model with random effects and explanatory variables.

Process Monitoring ROC Curve for Evaluating Dynamic Screening Methods P. 236-248

Peihua Qiu, Zhiming Xia & Lu You

Abstract

In practice, we often need to sequentially monitor the performance of individual subjects or processes, so that interventions can be made in a timely manner to avoid unpleasant consequences (e.g., strokes or airplane crashes) once the longitudinal patterns of their performance variables deviate significantly from the regular patterns of well-functioning subjects or processes. Some statistical methods are available to handle this dynamic screening (DS) problem. Because the performance of the DS methods is related to their signal times, the conventional false positive rate (FPR) and false negative rate (FNR) cannot be effective in measuring their performance. So far, there is no existing metrics in the literature for properly measuring the performance of DS methods. In this article, we aim to fill this gap by proposing a new performance evaluation approach, called process monitoring receiver operating characteristic curve, which properly combines the signal times with (FPR,FNR). Numerical examples and theoretical justifications show that this approach provides an effective tool for measuring the performance of DS methods.

An Effective Method for Online Disease Risk Monitoring P. 249-264

Lu You & Peihua Qiu

Abstract

Many diseases can be prevented or treated if they can be detected early or signaled before their occurrence. Disease early detection and prevention (DEDAP) is thus important for health improvement of our society. Traditionally, people are encouraged to check their health conditions regularly so that readings of relevant medical indices can be compared with certain threshold values and any irregular readings can trigger further medical tests to find root causes or diseases. One limitation of such traditional DEDAP methods is that they focus mainly on the data collected at the current time point and historical data are not fully used. Consequently, irregular longitudinal pattern of the medical indices could be neglected and certain diseases could be left undetected. In this article, we suggest a novel and effective new method for DEDAP. To detect a disease by this method, a patient's risk to the disease is first quantified at

each time point, and then the longitudinal pattern of the risk is monitored sequentially over time. A signal will be triggered by a large cumulative difference between the longitudinal risk pattern of the patient under monitoring and the longitudinal risk pattern of a typical person without the disease in concern. Both theoretical arguments and numerical studies show that it works well in practice.

Additive Heredity Model for the Analysis of Mixture-of-Mixtures Experiments P. 265-276

Sumin Shen, Lulu Kang & Xinwei Deng

Abstract

The mixture-of-mixtures (MoM) experiment is different from the classical mixture experiment in that the mixture component in MoM experiments, known as the major component, is made up of subcomponents, known as the minor components. In this article, we propose an additive heredity model (AHM) for analyzing MoM experiments. The proposed model considers an additive structure to inherently connect the major components with the minor components. To enable a meaningful interpretation for the estimated model, the hierarchical and heredity principles are applied by using the nonnegative garrote technique for model selection. The performance of the AHM was compared to several conventional methods in both unconstrained and constrained MoM experiments. The AHM was then successfully applied in two real-world problems studied previously in the literature.