Technometrics, ISSN 0040-1706
Volume 59, number 2 (may 2017)

Sequential Designs Based on Bayesian Uncertainty Quantification in Sparse Representation Surrogate Modeling

P. 139-152

Ray-Bing Chen, Weichung Wang & C. F. Jeff Wu

Abstract
A numerical method, called overcomplete basis surrogate method (OBSM), was recently proposed, which employs overcomplete basis functions to achieve sparse representations. While the method can handle nonstationary response without the need of inverting large covariance matrices, it lacks the capability to quantify uncertainty in predictions. We address this issue by proposing a Bayesian approach that first imposes a normal prior on the large space of linear coefficients, then applies the Markov chain Monte Carlo (MCMC) algorithm to generate posterior samples for predictions. From these samples, Bayesian credible intervals can then be obtained to assess prediction uncertainty. A key application for the proposed method is the efficient construction of sequential designs. Several sequential design procedures with different infill criteria are proposed based on the generated posterior samples. Numerical studies show that the proposed schemes are capable of solving problems of positive point identification, optimization, and surrogate fitting.

Emulation of Numerical Models With Over-Specified Basis Functions

P. 153-164

Avishek Chakraborty, Derek Bingham, Soma S. Dhavala, Carolyn C. Kuranz, R. Paul Drake, Michael J. Grosskopf, Erica M. Rutter, Ben R. Torralva, James P. Holloway, Ryan G. McClarren & Bani K. Mallick

Abstract
Mathematical models are frequently used to explore physical systems, but can be computationally expensive to evaluate. In such settings, an emulator is used as a surrogate. In this work, we propose a basis-function approach for computer model emulation. To combine field observations with a collection of runs from the numerical model, we use the proposed emulator within the Kennedy-O'Hagan framework of model calibration. A novel feature of the approach is the use of an over-specified set of basis functions where number of bases used and their inclusion probabilities are treated as unknown quantities. The new approach is found to have smaller predictive uncertainty and computational efficiency than the standard Gaussian process approach to emulation and calibration. Along with several simulation examples focusing on different model characteristics, we also use the method to analyze a dataset on laboratory experiments related to astrophysics.

Lifted Brownian Kriging Models

P. 165-177

Matthew Plumlee & Daniel W. Apley

Abstract
Gaussian processes have become a standard framework for modeling deterministic computer simulations and producing predictions of the response surface. This article investigates a new covariance function that is shown to offer superior prediction compared to the more common covariances for computer simulations of real physical systems. This is demonstrated via a gamut of realistic examples. A simple, closed-form expression for the covariance

is derived as a limiting form of a Brownian-like covariance model as it is extended to some hypothetical higher-dimensional input domain, and so we term it a lifted Brownian covariance. This covariance has connections with the multiquadric kernel. Through analysis of the kriging model, this article offers some theoretical comparisons between the proposed covariance model and existing covariance models. The major emphasis of the theory is explaining why the proposed covariance is superior to its traditional counterparts for many computer simulations of real physical systems.

Maximum Likelihood Estimation for Stochastic Differential Equations Using Sequential Gaussian-Process-Based Optimization P. 178-188

Grant Schneider, Peter F. Craigmile & Radu Herbei

Abstract

Stochastic differential equations (SDEs) are used as statistical models in many disciplines. However, intractable likelihood functions for SDEs make inference challenging, and we need to resort to simulation-based techniques to estimate and maximize the likelihood function. While importance sampling methods have allowed for the accurate evaluation of likelihoods at fixed parameter values, there is still a question of how to find the maximum likelihood estimate. In this article, we propose an efficient Gaussian-process-based method for exploring the parameter space using estimates of the likelihood from an importance sampler. Our technique accounts for the inherent Monte Carlo variability of the estimated likelihood, and does not require knowledge of gradients. The procedure adds potential parameter values by maximizing the so-called expected improvement, leveraging the fact that the likelihood function is assumed to be smooth. Our simulations demonstrate that our method has significant computational and efficiency gains over existing grid- and gradient-based techniques. Our method is applied to the estimation of ocean circulation from Lagrangian drift data in the South Atlantic ocean.

Box–Cox Transformation in Big Data P. 189-201

Tonglin Zhang & Baijian Yang

Abstract

The Box–Cox transformation is an important technique in linear regression when assumptions of a regression model are seriously violated. The technique has been widely accepted and extensively applied since it was first proposed. Based on the maximum likelihood approach, previous methods and algorithms for the Box–Cox transformation are mostly developed for small or moderate data. These methods and algorithms cannot be applied to big data because of the memory and storage capacity barriers. To overcome these difficulties, the present article proposes new methods and algorithms, where the basic idea is to construct and compute a set of summary statistics, which is termed as the Box–Cox information array in the article. According to the property of the maximum likelihood approach, the computation of the Box–Cox information array is the only issue to be considered in reading of data. Once the Box–Cox information array is obtained, the optimal power transformation as well as the corresponding estimates of model parameters can be quickly computed. Since the whole dataset is scanned only once, the proposed methods and algorithms can be extremely efficient and fast even when multiple models are considered. It is expected that the basic knowledge gained in this article will have a great impact on the development of statistical methods and algorithms for big data.

Estimating a Parametric Component Lifetime Distribution from a Collection of Superimposed Renewal Processes P. 202-214

Wei Zhang, Ye Tian, Luis A. Escobar & William Q. Meeker

Abstract

Maintenance data can be used to make inferences about the lifetime distribution of system components. Typically, a fleet contains multiple systems. Within each system, there is a set of nominally identical replaceable components of particular interest (e.g., 2 automobile headlights, 8 dual in-line memory module (DIMM) modules in a computing server, 16 cylinders in a locomotive engine). For each component replacement event, there is system-level information that a component was replaced, but no information on which particular component was replaced. Thus, the observed data are a collection of superpositions of renewal processes (SRP), one for each system in the fleet. This article proposes a procedure for estimating the component lifetime distribution using the aggregated event data from a fleet of systems. We show how to compute the likelihood function for the collection of SRPs and provide suggestions for efficient computations. We compare performance of this incomplete-data maximum likelihood (ML) estimator with the complete-data ML estimator and study the performance of confidence interval methods for estimating quantiles of the lifetime distribution of the component.

System Reliability and Component Importance Under Dependence: A Copula Approach P. 215-224

Xiang Zhang & Alyson Wilson

Abstract

System reliability and component importance are of great interest in reliability modeling, especially when the components within the system are dependent. We characterize the influence of dependence structures on system reliability and component importance in coherent systems with discrete marginal distributions. The effects of dependence are captured through copula theory. We extend our framework to coherent multi-state system. Applications of the derived results are demonstrated using a Gaussian copula, which yields simple interpretations. Simulations and two examples are presented to demonstrate the importance of modeling dependence when estimating system reliability and ranking of component importance. Proofs, algorithms, code, and data are provided in supplementary materials available online.

A Multi-Level Trend-Renewal Process for Modeling Systems With Recurrence Data P. 225-236

Zhibing Xu, Yili Hong, William Q. Meeker, Brock E. Osborn & Kati Illouz

Abstract

A repairable system is a system that can be restored to an operational state after a repair event. The system may experience multiple events over time that are called recurrent events. To model the recurrent event data, the renewal process (RP), the nonhomogenous Poisson process (NHPP), and the trend-renewal process (TRP) are often used. Compared to the RP and NHPP, the TRP is more flexible for modeling, because it includes both RP and NHPP as special cases. However, for a multi-level system (e.g., system, subsystem, and component levels), the original TRP model may not be adequate if the repair is effected by a subsystem replacement and if subsystem-level replacement events affect the rate of occurrence of the component-level replacement events. In this article, we propose a general class of models to describe replacement events in a multi-level repairable system by extending the TRP model. We also develop procedures for parameter estimation and the prediction of future events based on historical data. The proposed model and method are validated by simulation studies and are illustrated by an industrial application. This article has online supplementary materials.

An Ameliorated Improvement Factor Model for Imperfect Maintenance and Its Goodness of Fit P. 237-246

Mimi Zhang & Min Xie

Abstract

Maintenance actions can be classified, according to their efficiency, into three categories: perfect maintenance, imperfect maintenance, and minimal maintenance. To date, the literature on imperfect maintenance is voluminous, and many models have been developed to treat imperfect maintenance. Yet, there are two important problems in the community of maintenance that still remain wide open: how to give practical grounds for an imperfect-maintenance model, and how to test the fit of a real dataset to an imperfect-maintenance model. Motivated by these two pending problems, this work develops an imperfect-maintenance model by taking a physically meaningful approach. For the practical implementation of the developed model, we advance two methods, called QMI method and spacing-likelihood algorithm, to estimate involved unknown parameters. The two methods complete each other and are widely applicable. To offer a practical guide for testing fit to an imperfect-maintenance model, this work promotes a bootstrapping approach to approximating the distribution of a test statistic. The attractions and dilemmas of QMI method and spacing-likelihood algorithm are revealed via simulated data. The utility of the developed imperfect-maintenance model is evidenced via a real dataset. This article has a supplementary material online.

Statistical Framework for Improved Automatic Flaw Detection in Nondestructive Evaluation Images P. 247-261

Ye Tian, Ranjan Maitra, William Q. Meeker & Stephen D. Holland

Abstract

Nondestructive evaluation (NDE) techniques are widely used to detect flaws in critical components of systems like aircraft engines, nuclear power plants, and oil pipelines to prevent catastrophic events. Many modern NDE systems generate image data. In some applications, an experienced inspector performs the tedious task of visually examining every image to provide accurate conclusions about the existence of flaws. This approach is labor-intensive and can cause misses due to operator ennui. Automated evaluation methods seek to eliminate human-factors variability and improve throughput. Simple methods based on peak amplitude in an image are sometimes employed and a trained-operator-controlled refinement that uses a dynamic threshold based on signal-to-noise ratio (SNR) has also been implemented. We develop an automated and optimized detection procedure that mimics these operations. The primary goal of our methodology is to reduce the number of images requiring expert visual evaluation by filtering out images that are overwhelmingly definitive on the existence or absence of a flaw. We use an appropriate model for the observed values of the SNR-detection criterion to estimate the probability of detection. Our methodology outperforms current methods in terms of its ability to detect flaws. Supplementary materials for this article are available online.

Consistent Testing for Pairwise Dependence in Time Series P. 262-270

K. Fokianos & M. Pitsillou

We consider the problem of testing pairwise dependence for stationary time series. For this, we suggest the use of a Box–Ljung-type test statistic that is formed after calculating the distance covariance function among pairs of observations. The distance covariance function is a suitable measure for detecting dependencies between observations as it is based on the distance between the characteristic function of the joint distribution of the random variables and the product of the marginals. We show that, under the null hypothesis of independence and under mild regularity conditions, the test statistic converges to a normal random variable. The results are complemented by several examples. This article has supplementary material online.