Technometrics, ISSN 0040-1706
Volume 60, number 4 (november 2018)

Permutation and Grouping Methods for Sharpening Gaussian Process Approximations P. 415-429

Joseph Guinness

Abstract
Vecchia's approximate likelihood for Gaussian process parameters depends on how the observations are ordered, which has been cited as a deficiency. This article takes the alternative standpoint that the ordering can be tuned to sharpen the approximations. Indeed, the first part of the article includes a systematic study of how ordering affects the accuracy of Vecchia's approximation. We demonstrate the surprising result that random orderings can give dramatically sharper approximations than default coordinate-based orderings. Additional ordering schemes are described and analyzed numerically, including orderings capable of improving on random orderings. The second contribution of this article is a new automatic method for grouping calculations of components of the approximation. The grouping methods simultaneously improve approximation accuracy and reduce computational burden. In common settings, reordering combined with grouping reduces Kullback–Leibler divergence from the target model by more than a factor of 60 compared to ungrouped approximations with default ordering. The claims are supported by theory and numerical results with comparisons to other approximations, including tapered covariances and stochastic partial differential equations. Computational details are provided, including the use of the approximations for prediction and conditional simulation. An application to space-time satellite data is presented.

Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets P. 430-444

Rajarshi Guhaniyogi & Sudipto Banerjee

Abstract
Spatial process models for analyzing geostatistical data entail computations that become prohibitive as the number of spatial locations becomes large. There is a burgeoning literature on approaches for analyzing large spatial datasets. In this article, we propose a divide-and-conquer strategy within the Bayesian paradigm. We partition the data into subsets, analyze each subset using a Bayesian spatial process model, and then obtain approximate posterior inference for the entire dataset by combining the individual posterior distributions from each subset. Importantly, as often desired in spatial analysis, we offer full posterior predictive inference at arbitrary locations for the outcome as well as the residual spatial surface after accounting for spatially oriented predictors. We call this approach "spatial meta-kriging" (SMK). We do not need to store the entire data in one processor, and this leads to superior scalability. We demonstrate SMK with various spatial regression models including Gaussian processes with Matern and compactly supported correlation functions. The approach is intuitive, easy to implement, and is supported by theoretical results presented in the supplementary material available online. Empirical illustrations are provided using different simulation experiments and a geostatistical analysis of Pacific Ocean sea surface temperature data. Supplementary materials for this article are available online.

Using History Matching for Prior Choice

P. 445-460

Xueou Wang, David J. Nott, Christopher C. Drovandi, Kerrie Mengersen & Michael Evans

Abstract

It can be important in Bayesian analyses of complex models to construct informative prior distributions which reflect knowledge external to the data at hand. Nevertheless, how much prior information an analyst can elicit from an expert will be limited due to constraints of time, cost and other factors. This article develops effective numerical methods for exploring reasonable choices of a prior distribution from a parametric class, when prior information is specified in the form of some limited constraints on prior predictive distributions, and where these prior predictive distributions are analytically intractable. The methods developed may be thought of as a novel application of the ideas of history matching, a technique developed in the literature on assessment of computer models. We illustrate the approach in the context of logistic regression and sparse signal shrinkage

Degradation in Common Dynamic Environments

P. 461-471

Qingqing Zhai & Zhi-Sheng Ye

Abstract

Degradation studies are often used to assess reliability of products subject to degradation-induced soft failures. Because of limited test resources, several test subjects may have to share a test rig and have their degradation measured by the same operator. The common environments experienced by subjects in the same group introduce significant interindividual correlations in their degradation, which is known as the block effect. In the present article, the Wiener process is used to model product degradation, and the group-specific random environments are captured using a stochastic time scale. Both semiparametric and parametric estimation procedures are developed for the model. Maximum likelihood estimations of the model parameters for both the semiparametric and parametric models are obtained with the help of the EM algorithm. Performance of the maximum likelihood estimators is validated through large sample asymptotics and small sample simulations. The proposed models are illustrated by an application to lumen maintenance data of blue light-emitting diodes. Supplementary materials for this article are available online.

Sequential Bayesian Design for Accelerated Life Tests

P. 472-483

I-Chen Lee, Yili Hong, Sheng-Tsaing Tseng & Tirthankar Dasgupta

Abstract

Most of the recently developed methods on optimum planning for accelerated life tests (ALT) involve “guessing” values of parameters to be estimated, and substituting such guesses in the proposed solution to obtain the final testing plan. In reality, such guesses may be very different from true values of the parameters, leading to inefficient test plans. To address this problem, we propose a sequential Bayesian strategy for planning of ALTs and a Bayesian estimation procedure for updating the parameter estimates sequentially. The proposed approach is motivated by ALT for polymer composite materials, but are generally applicable to a wide range of testing scenarios. Through the proposed sequential Bayesian design, one can efficiently collect data and then make predictions for the field performance. We use extensive simulations to evaluate the properties of the proposed sequential test planning strategy. We compare the proposed method to various traditional non-sequential optimum designs. Our results show that the proposed strategy is more robust and efficient, as compared to existing non-sequential optimum designs. Supplementary materials for this article are available online.

Nonparametric Modeling and Prognosis of Condition Monitoring Signals Using Multivariate Gaussian Convolution Processes

Raed Kontar, Shiyu Zhou, Chaitanya Sankavaram, Xinyu Du & Yilu Zhang

Abstract

Graphical models have been widely used to investigate the complex dependence structure of high-dimensional data, and it is common to assume that observed data follow a homogeneous graphical model. However, observations usually come from different resources and have heterogeneous hidden commonality in real-world applications. Thus, it is of great importance to estimate heterogeneous dependencies and discover a subpopulation with certain commonality across the whole population. In this work, we introduce a novel regularized estimation scheme for learning nonparametric finite mixture of Gaussian graphical models, which extends the methodology and applicability of Gaussian graphical models and mixture models. We propose a unified penalized likelihood approach to effectively estimate nonparametric functional parameters and heterogeneous graphical parameters. We further design an efficient generalized effective expectation-maximization (EM) algorithm to address three significant challenges: high-dimensionality, nonconvexity, and label switching. Theoretically, we study both the algorithmic convergence of our proposed algorithm and the asymptotic properties of our proposed estimators. Numerically, we demonstrate the performance of our method in simulation studies and a real application to estimate human brain functional connectivity from attention deficit hyperactivity disorder (ADHD) imaging data, where two heterogeneous conditional dependencies are explained through profiling demographic variables and supported by existing scientific findings.

Profile Monitoring of Probability Density Functions via Simplicial Functional PCA With Application to Image Data

Alessandra Menafoglio, Marco Grasso, Piercesare Secchi & Bianca Maria Colosimo

Abstract

The advance of sensor and information technologies is leading to data-rich industrial environments, where large amounts of data are potentially available. This study focuses on industrial applications where image data are used more and more for quality inspection and statistical process monitoring. In many cases of interest, acquired images consist of several and similar features that are randomly distributed within a given region. Examples are pores in parts obtained via casting or additive manufacturing, voids in metal foams and light-weight components, grains in metallographic analysis, etc. The proposed approach summarizes the random occurrences of the observed features via their (empirical) probability density functions (PDFs). In particular, a novel approach for PDF monitoring is proposed. It is based on simplicial functional principal component analysis (SFPCA), which is performed within the space of density functions, that is, the Bayes space B^2. A simulation study shows the enhanced monitoring performances provided by SFPCA-based profile monitoring against other competitors proposed in the literature. Finally, a real case study dealing with the quality control of foamed material production is discussed, to highlight a practical use of the proposed methodology. Supplementary materials for the article are available online.

Nonparametric Finite Mixture of Gaussian Graphical Models

Kevin H. Lee & Lingzhou Xue

Abstract

Graphical models have been widely used to investigate the complex dependence structure of high-dimensional data, and it is common to assume that observed data follow a homogeneous graphical model. However, observations usually come from different resources and have heterogeneous hidden commonality in real-world applications. Thus, it is of great importance to estimate heterogeneous dependencies and discover a subpopulation with certain commonality across the whole population. In this work, we introduce a novel regularized estimation scheme for learning nonparametric finite mixture of Gaussian graphical models, which extends the methodology and applicability of Gaussian graphical models and mixture models. We propose a unified penalized likelihood approach to effectively estimate nonparametric functional parameters and heterogeneous graphical parameters. We further design an efficient generalized effective expectation-maximization (EM) algorithm to address three significant challenges: high-dimensionality, nonconvexity, and label switching. Theoretically, we study both the algorithmic convergence of our proposed algorithm and the asymptotic properties of our proposed estimators. Numerically, we demonstrate the performance of our method in simulation studies and a real application to estimate human brain functional connectivity from attention deficit hyperactivity disorder (ADHD) imaging data, where two heterogeneous conditional dependencies are explained through profiling demographic variables and supported by existing scientific findings.

Efficient Blind Image Deblurring Using Nonparametric Regression and Local Pixel Clustering P. 522-531

Yicheng Kang, Partha Sarathi Mukherjee & Peihua Qiu

Abstract
Blind image deblurring is a challenging ill-posed problem. It would have an infinite number of solutions even in cases when an observed image contains no noise. In reality, however, observed images almost always contain noise. The presence of noise would make the image deblurring problem even more challenging because the noise can cause numerical instability in many existing image deblurring procedures. In this article, a novel blind image deblurring approach is proposed, which can remove both pointwise noise and spatial blur efficiently without imposing restrictive assumptions on either the point spread function (psf) or the true image. It even allows the psf to be location dependent. In the proposed approach, a local pixel clustering procedure is used to handle the challenging task of restoring complicated edge structures that are tapered by blur, and a nonparametric regression procedure is used for removing noise at the same time. Numerical examples show that our proposed method can effectively handle a wide variety of blur and it works well in applications. Supplementary materials for this article are available online.

Bayesian Model Building From Small Samples of Disparate Data for Capturing In-Plane Deviation in Additive Manufacturing P. 532-544

Arman Sabbaghi, Qiang Huang & Tirthankar Dasgupta

Abstract
Quality control of geometric shape deviation in additive manufacturing relies on statistical deviation models. However, resource constraints limit the manufacture of test shapes, and consequently impede the specification of deviation models for new shape varieties. We present an adaptive Bayesian methodology that effectively combines in-plane deviation data and models for a small sample of previously manufactured, disparate shapes to aid in the model specification of in-plane deviation for a broad class of new shapes. The power and simplicity of this general methodology is demonstrated with illustrative case studies on in-plane deviation modeling for polygons and straight edges in free-form shapes using only data and models for cylinders and a single regular pentagon. Our Bayesian approach facilitates deviation modeling in general, and thereby can help advance additive manufacturing as a high-quality technology. Supplementary materials for this article are available online.

Analysis of Repeatability and Reproducibility Studies With Ordinal Measurements P. 545-556

Stacey L. Culp, Kenneth J. Ryan, Juan Chen & Michael S. Hamada

Abstract
A Bayesian inferential approach with a noninformative prior is introduced to analyze ordinal repeatability and reproducibility (R&R) data using the De Mast–Van Wieringen model. This approach is extended with a weakly informative prior and random effects to allow for the consideration of a population of raters and prediction of a new rater. This random-effects approach is also shown to result in partial pooling of estimates across raters. In addition, match-probability-based measures to decompose ordinal R&R study data into contributions due to repeatability and due to reproducibility are defined. All extensions involving Bayesian inference (for fixed or random effects) and

measures are illustrated on real and simulated ordinal R&R study data and are applicable in business and industry settings. This methodology can be implemented using the supplemental R package ordinalRR available from CRAN. Additional supplementary material for this article is available online.

Technometrics, ISSN 0040-1706
Volume 60, number 4 (november 2018)