Technometrics, ISSN 0040-1706
Volume 63, number 1 (february 2021)

Abstract
Online experiments and specifically A/B testing are commonly used to identify whether a proposed change to a web page is in fact an effective one. This study focuses on basic settings in which a binary outcome is obtained from each user who visits the website and the probability of a response may be affected by numerous factors. We use Bayesian probit regression to model the factor effects and combine elements from traditional two-level factorial experiments and multiarmed bandits to construct sequential designs that embed attractive features of estimation and exploitation.

Abstract
We consider the problem of estimating the set of all inputs that leads a system to some particular behavior. The system is modeled by an expensive-to-evaluate function, such as a computer experiment, and we are interested in its excursion set, that is, the set of points where the function takes values above or below some prescribed threshold. The objective function is emulated with a Gaussian process (GP) model based on an initial design of experiments enriched with evaluation results at (batch-) sequentially determined input points. The GP model provides conservative estimates for the excursion set, which control false positives while minimizing false negatives. We introduce adaptive strategies that sequentially select new evaluations of the function by reducing the uncertainty on conservative estimates. Following the stepwise uncertainty reduction approach we obtain new evaluations by minimizing adapted criteria. Tractable formulas for the conservative criteria are derived, which allow more convenient optimization. The method is benchmarked on random functions generated under the model assumptions in different scenarios of noise and batch size. We then apply it to a reliability engineering test case. Overall, the proposed strategy of minimizing false negatives in conservative estimation achieves competitive performance both in terms of model-based and model-free indicators. Supplementary materials for this article are available online.

Abstract
Model calibration uses outputs from a simulator and field data to build a predictive model for the physical system and to estimate unknown inputs. The conventional approach to model calibration assumes that the observations are continuous outcomes. In many applications this is not the case. The methodology proposed was motivated by an application in modeling photon counts at the Center for Exascale Radiation Transport. There, high performance computing is used for simulating the flow of neutrons through various materials. In this article, new Bayesian methodology for computer model calibration to handle the count structure of our observed data allows closer fidelity

to the experimental system and provides flexibility for identifying different forms of model discrepancy between the simulator and experiment. Supplementary materials for this article are available online.

Distance-Distributed Design for Gaussian Process Surrogates P. 40-52

Boya Zhang, D. Austin Cole & Robert B. Gramacy

Abstract

A common challenge in computer experiments and related fields is to efficiently explore the input space using a small number of samples, that is, the experimental design problem. Much of the recent focus in the computer experiment literature, where modeling is often via Gaussian process (GP) surrogates, has been on space-filling designs, via maximin distance, Latin hypercube, etc. However, it is easy to demonstrate empirically that such designs disappoint when the model hyperparameterization is unknown, and must be estimated from data observed at the chosen design sites. This is true even when the performance metric is prediction-based, or when the target of interest is inherently or eventually sequential in nature, such as in blackbox (Bayesian) optimization. Here we expose such inefficiencies, showing that in many cases a purely random design is superior to higher-powered alternatives. We then propose a family of new schemes by reverse engineering the qualities of the random designs which give the best estimates of GP length scales. Specifically, we study the distribution of pairwise distances between design elements, and develop a numerical scheme to optimize those distances for a given sample size and dimension. We illustrate how our distance-based designs, and their hybrids with more conventional space-filling schemes, outperform in both static (one-shot design) and sequential settings.

Gaussian Process Modeling of Heterogeneity and Discontinuities Using Voronoi Tessellations P. 53-63

Christopher A. Pope, John Paul Gosling, Stuart Barber, Jill S. Johnson, Takanobu Yamaguchi, Graham Feingold & Paul G. Blackwell

Abstract

Many methods for modeling functions over high-dimensional spaces assume global smoothness properties; such assumptions are often violated in practice. We introduce a method for modeling functions that display heterogeneity or contain discontinuities. The heterogeneity is dealt with by using a combination of Voronoi tessellation, to partition the input space, and separate Gaussian processes to model the function over different regions of the partitioned space. The proposed method is highly flexible since it allows the Voronoi cells to combine to form regions, which enables nonconvex and disconnected regions to be considered. In such problems, identifying the borders between regions is often of great importance and we propose an adaptive sampling method to gain extra information along such borders. The method is illustrated by simulated examples and an application to real data, in which we see improvements in prediction error over the commonly used stationary Gaussian process and other nonstationary variations. In our application, a computationally expensive computer model that simulates the formation of clouds is investigated, the proposed method more accurately predicts the underlying process at unobserved locations than existing emulation methods. Supplementary materials for this article are available online.

Boundary Detection Using a Bayesian Hierarchical Model for Multiscale Spatial Data P. 64-76

Kai Qu, Jonathan R. Bradley & Xufeng Niu

Abstract

Spatial boundary analysis has attained considerable attention in several disciplines including engineering, shape analysis, spatial statistics, and computer science. The inferential question of interest is often to identify rapid surface change of an unobserved latent process. Curvilinear wombling and crisp wombling (or fuzzy) are two major approaches that have emerged in Bayesian spatial statistics literature. These methods are limited to a single spatial scale even though data with multiple spatial scales are often accessible. Thus, we propose a multiscale

representation of the directional derivative Karhunen–Loéve expansion to perform directionally based boundary detection. Taking a multiscale spatial perspective allows us, for the first time, to consider the concept of curvilinear boundary fallacy (CBF) error, which is a boundary detection analog to the ecological fallacy that is often studied in spatial change of support literature. Furthermore, we propose a directionally based multiscale curvilinear boundary error criterion to quantify CBF. We refer to this metric as the criterion for boundary aggregation error (BAGE), and use it to perform boundary detection. Several theoretical results are derived to motivate BAGE. In particular, we show that no BAGE exists when the directional derivatives of eigenfunctions of a KL expansion are constant across spatial scales. We illustrate the use of our model through a simulated example and an analysis of Mediterranean wind measurements data. Supplementary materials for this article are available online.

Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand P. 77-89

Xiaochen Xian, Honghan Ye, Xin Wang & Kaibo Liu

Abstract

Traffic demand prediction has been a crucial problem for the planning, scheduling, and optimization in transportation management. The prediction of traffic demand counts for origin-destination (OD) pairs has been considered challenging due to the high variability and complicated spatiotemporal correlations in the data. Though several articles have considered estimating traffic flows from counts observed at specific locations, existing traffic prediction models seldom dealt with spatiotemporal demand count data of certain OD pairs, or they failed to effectively consider domain knowledge of the traffic network to enhance the prediction accuracy of traffic demand. To tackle the aforementioned challenges, we formulate and propose a multivariate Poisson log-normal model with specific parameterization tailored to the traffic demand problem, which captures the spatiotemporal correlations of the traffic demand across different routes and epochs, and automatically clusters the routes based on the demand correlations. The model is further estimated using an expectation-maximization algorithm and applied for predicting future demand counts at the subsequent epochs. The estimation and prediction procedures incorporate Markov chain Monte Carlo sampling to overcome the computational challenges. Simulations as well as a real application on a New York yellow taxi data are performed to demonstrate the applicability and effectiveness of the proposed method. Supplementary materials for this article are available online.

Bayesian Methods for Planning Accelerated Repeated Measures Degradation Tests P. 90-99

Brian P. Weaver & William Q. Meeker

Abstract

Accelerated repeated measures degradation tests are often used to assess product or component reliability when there would be few or even no failures during a traditional life test. Such tests are used to estimate the failure-time distributions of highly reliable items in applications where it is possible to take repeated measures of some appropriate degradation measure. When engineers have valid prior information about failure mechanisms, it is important that such information be used in inference and test planning. Bayesian methods provide a vehicle for doing this. This article describes methods for selecting a Bayesian repeated measures accelerated degradation test plan when the degradation and acceleration model is linear in the parameters. A Bayesian criterion based on estimation precision of the failure-time quantile at use conditions is selected for finding optimum test plans. We use a large-sample approximation for the posterior distribution to simplify the planning criterion. The general equivalence theorem is used to check for global optimality of the optimum test plan. We also discuss how to find a compromise test plan that satisfies practical constraints while still providing good statistical properties.

Nonparametric Bayesian Modeling and Estimation for Renewal Processes P. 100-115

Sai Xiao, Athanasios Kottas, Bruno Sansó & Hyotae Kim

Abstract

We propose a flexible approach to modeling for renewal processes. The model is built from a structured mixture of Erlang densities for the renewal process inter-arrival density. The Erlang mixture components have a common scale parameter, and the mixture weights are defined through an underlying distribution function modeled nonparametrically with a Dirichlet process (DP) prior. This model specification enables nonstandard shapes for the inter-arrival time density, including heavy tailed and multimodal densities. Moreover, the choice of the DP centering distribution controls clustering or declustering patterns for the point process, which can therefore be encouraged in the prior specification. Using the analytically available Laplace transforms of the relevant functions, we study the renewal function and the directly related K function, which can be used to infer about clustering or declustering patterns. From a computational point of view, the model structure is attractive as it enables efficient posterior simulation while properly accounting for the likelihood normalizing constant implied by the renewal process. A hierarchical extension of the model allows for the quantification of the impact of different levels of a factor. The modeling approach is illustrated with several synthetic datasets, earthquake occurrences data, and coal-mining disaster data.

Detection of latent heteroscedasticity and group-based regression effects in linear models via Bayesian model selection P. 116-126

Thomas A. Metzger & Christopher T. Franck

Abstract

Standard linear modeling approaches make potentially simplistic assumptions regarding the structure of categorical effects that may obfuscate more complex relationships governing data. For example, recent work focused on the two-way unreplicated layout has shown that hidden groupings among the levels of one categorical predictor frequently interact with the ungrouped factor. We extend the notion of a "latent grouping factor" to linear models in general. The proposed work allows researchers to determine whether an apparent grouping of the levels of a categorical predictor reveals a plausible hidden structure given the observed data. Specifically, we offer a Bayesian model selection-based approach to reveal latent group-based heteroscedasticity, regression effects, and/or interactions. Failure to account for such structures can produce misleading conclusions. Since the presence of latent group structures is frequently unknown *a priori* to the researcher, we use fractional Bayes factor methods and mixture g-priors to overcome lack of prior information.

A Simplified Formulation of Likelihood Ratio Confidence Intervals Using a Novel Property P. 127-135

Necip Doganaksoy

Abstract

This article describes a novel property of likelihood ratio (LR) confidence intervals which is subsequently used to formulate an alternative approach for their calculation. It is shown that LR confidence limits can be defined as the minimum and maximum values of a parameter (or a function of parameters) that satisfy a set value of the log-likelihood. The proposed formulation allows straightforward implementation in end-user computing settings and it is particularly useful for the computation of intervals on noninvertible functions of model parameters. The main goal of the article is to expose this little-known property of LR confidence limits to the practitioner and research communities. Two case studies based on applications in product quality and reliability improvement are used for illustration. The first case study deals with interval estimation of the difference between the means of two lognormal populations. The second application concerns interval estimation for misclassification probabilities attributable to measurement error. Supplementary materials for this article are available online.