Technometrics, ISSN 0040-1706
Volume 58, number 2 (may 2016)

Abstract

Partial least squares (PLS) is a widely used method for prediction in applied statistics, especially in chemometrics applications. However, PLS is not invariant or equivariant under scale transformations of the predictors, which tends to limit its scope to regressions in which the predictors are measured in the same or similar units. Cook, Helland, and Su (2013 Cook, R.D., Helland, I.S., Su, Z. (2013), Envelopes and Partial Least Squares Regression, Journal of the Royal Statistical Society, Series B,75, 851–877.[CrossRef]) built a connection between nascent envelope methodology and PLS, allowing PLS to be addressed in a traditional likelihood-based framework. In this article, we use the connection between PLS and envelopes to develop a new method—scaled predictor envelopes (SPE)—that incorporates predictor scaling into PLS-type applications. By estimating the appropriate scales, the SPE estimators can offer efficiency gains beyond those given by PLS, and further reduce prediction errors. Simulations and an example are given to support the theoretical claims.

Abstract

Complex natural phenomena are increasingly investigated by the use of a complex computer simulator. To leverage the advantages of simulators, observational data need to be incorporated in a probabilistic framework so that uncertainties can be quantified. A popular framework for such experiments is the statistical computer model calibration experiment. A limitation often encountered in current statistical approaches for such experiments is the difficulty in modeling high-dimensional observational datasets and simulator outputs as well as high-dimensional inputs. As the complexity of simulators seems to only grow, this challenge will continue unabated. In this article, we develop a Bayesian statistical calibration approach that is ideally suited for such challenging calibration problems. Our approach leverages recent ideas from Bayesian additive regression Tree models to construct a random basis representation of the simulator outputs and observational data. The approach can flexibly handle high-dimensional datasets, high-dimensional simulator inputs, and calibration parameters while quantifying important sources of uncertainty in the resulting inference. We demonstrate our methodology on a CO_2 emissions rate calibration problem, and on a complex simulator of subterranean radionuclide dispersion, which simulates the spatial–temporal diffusion of radionuclides released during nuclear bomb tests at the Nevada Test Site. Supplementary computer code and datasets are available online.

Abstract

Quantile regression is an important tool to determine the quality level of service, product, and operation systems via stochastic simulation. It is frequently known that the quantiles of the output distribution are monotonic functions of certain inputs to the simulation model. Because there is typically high variability in estimation of tail quantiles, it can be valuable to incorporate this information in quantile modeling. However, the existing literature on monotone quantile regression with multiple inputs is sparse. In this article, we propose a class of monotonic regression models, which consists of functional analysis of variance (FANOVA) decomposition components modeled with Bernstein polynomial bases for estimating quantiles as a function of multiple inputs. The polynomial degrees of the bases for the model and the FANOVA components included in the model are selected by a greedy algorithm. Real examples demonstrate the advantages of incorporating the monotonicity assumption in quantile regression and the good performance of the proposed methodology for estimating quantiles. Supplementary materials for this article are available online.

A Change-Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis P. 191-204

Kamran Paynabar, Changliang Zou & Peihua Qiu

Abstract

Process monitoring and fault diagnosis using profile data remains an important and challenging problem in statistical process control (SPC). Although the analysis of profile data has been extensively studied in the SPC literature, the challenges associated with monitoring and diagnosis of multichannel (multiple) nonlinear profiles are yet to be addressed. Motivated by an application in multioperation forging processes, we propose a new modeling, monitoring, and diagnosis framework for phase-I analysis of multichannel profiles. The proposed framework is developed under the assumption that different profile channels have similar structure so that we can gain strength by borrowing information from all channels. The multidimensional functional principal component analysis is incorporated into change-point models to construct monitoring statistics. Simulation results show that the proposed approach has good performance in identifying change-points in various situations compared with some existing methods. The codes for implementing the proposed procedure are available in the supplementary material.

Bayesian Detection of Changepoints in Finite-State Markov Chains for Multiple Sequences P. 205-213

Petter Arnesen, Tracy Holsclaw & Padhraic Smyth

Abstract

We consider the analysis of sets of categorical sequences consisting of piecewise homogenous Markov segments. The sequences are assumed to be governed by a common underlying process with segments occurring in the same order for each sequence. Segments are defined by a set of unobserved changepoints where the positions and number of changepoints can vary from sequence to sequence. We propose a Bayesian framework for analyzing such data, placing priors on the locations of the changepoints and on the transition matrices and using Markov chain Monte Carlo (MCMC) techniques to obtain posterior samples given the data. Experimental results using simulated data illustrate how the methodology can be used for inference of posterior distributions for parameters and changepoints, as well as the ability to handle considerable variability in the locations of the changepoints across different sequences. We also investigate the application of the approach to sequential data from an application involving monsoonal rainfall patterns. Supplementary materials for this article are available online.

Bayesian Inference for a New Class of Distributions on Equivalence Classes of Three-Dimensional Orientations With Applications to Materials Science P. 214-224

Chuanlong Du, Daniel J. Nordman & Stephen B. Vardeman

Abstract

Experiments in materials science investigating cubic crystalline structures often collect data which are in truth *equivalence classes* of crystallographically symmetric orientations. These intend to represent how lattice structures of

particles are orientated relative to a reference coordinate system. Motivated by a materials science application, we formulate parametric probability models for "unlabeled orientation data." This amounts to developing models on equivalence classes of three-dimensional rotations. We use a flexible existing model class for random rotations (called uniform-axis-random-spin models) to induce probability distributions on the equivalence classes of rotations. We develop one-sample Bayesian inference for the parameters in these models, and compare this methodology to some likelihood-based approaches. We also contrast the new parametric analysis of unlabeled orientation data with other analyses that proceed as if the data have been preprocessed into honest orientation data. Supplementary materials for this article are available online.

Prior-Free Probabilistic Prediction of Future Observations P. 225-235

Ryan Martin & Rama T. Lingham

Abstract

Prediction of future observations is a fundamental problem in statistics. Here we present a general approach based on the recently developed inferential model (IM) framework. We employ an IM-based technique to marginalize out the unknown parameters, yielding prior-free probabilistic prediction of future observables. Verifiable sufficient conditions are given for validity of our IM for prediction, and a variety of examples demonstrate the proposed method's performance. Thanks to its generality and ease of implementation, we expect that our IM-based method for prediction will be a useful tool for practitioners. Supplementary materials for this article are available online.

Quantifying Uncertainty in Lumber Grading and Strength Prediction: A Bayesian Approach P. 236-243

Samuel W.K. Wong, Conroy Lum, Lang Wu & James V. Zidek

Abstract

This article presents a joint distribution for the strength of a randomly selected piece of structural lumber and its observable characteristics. In the process of lumber strength testing, these characteristics are ascertained under strict grading protocols, as they have the potential to be strength reducing. However, for practical reasons, only a few such selected characteristics among the many present, are recorded. We present a data-generating mechanism that reflects the uncertainties resulting from the grading protocol. A Bayesian approach is then adopted for model fitting and construction of a predictive distribution for strength that accounts for the unrecorded characteristics. The method is validated on simulated examples, and then applied on a sample of specimens tested for bending and tensile strength. Use of the predictive distribution is demonstrated, and insights gained into the grading process are described. Details of the lumber testing experiments can be found in the online supplementary materials.

Optimum Allocation Rule for Accelerated Degradation Tests With a Class of Exponential-Dispersion Degradation Models P. 244-254

Sheng-Tsaing Tseng & I-Chen Lee

Abstract

Optimum allocation problem in accelerated degradation tests (ADTs) is an important task for reliability analysts. Several researchers have attempted to address this decision problem, but their results have been based only on specific degradation models. Therefore, they lack a unified approach toward general degradation models. This study proposes a class of exponential dispersion (ED) degradation models to overcome this difficulty. Assuming that the underlying degradation path comes from the ED class, we analytically derive the optimum allocation rules (by minimizing the asymptotic variance of the estimated q quantile of product's lifetime) for two-level and three-level ADT allocation problems whether the testing stress levels are prefixed or not. For a three-level allocation problem, we show that all test units should be allocated into two out of three stresses, depending on certain specific conditions. Two examples are used to illustrate the proposed procedure. Furthermore, the penalties of using nonoptimum allocation rules are also addressed. This study demonstrates that a three-level compromise plan with small proportion allocation

in the middle stress, in general, is a good strategy for ADT allocation. Supplementary materials for this article are available online.

Comparing the Slack-Variable Mixture Model With Other Alternatives P. 255-268

Lulu Kang, Javier Cruz Salgado & William A. Brenneman

Abstract

There have been many linear regression models proposed to analyze mixture experiments including the Scheffé model, the slack-variable model, and the Kronecker model. The use of the slack-variable model is somewhat controversial within the mixture experiment research community. However, in situations that the slack-variable ingredient is used to fill in the formulation and the remaining ingredients have constraints such that they can be chosen independently of one another, the slack-variable model is extremely popular by practitioners mainly due to the ease of interpretation. In this article, we advocate that for some mixture experiments the slack-variable model has appealing properties including numerical stability and better prediction accuracy when model-term selection is performed. We also explain how the effects of the slack-variable model components should be interpreted and how easy it is for practitioners to understand the components effects. We also investigate how to choose the slack-variable component, what transformation should be used to reduce collinearity, and under what circumstances the slack-variable model should be preferred. Both simulation and practical examples are provided to support the conclusions.

Optimal Experimental Designs in the Flow Rate of Particles P. 269-276

Mariano Amo-Salas, Elvira Delgado-Márquez & Jesús López-Fidalgo

Abstract

This article focuses on analyzing the process of jam formation during the discharge by gravity of granular material stored in a two-dimensional silo. The aim of the article is two-fold. First, optimal experimental designs are computed, in which four approaches are considered: D-optimality, a combination of D-optimality and a cost/gain function, Bayesian D-optimality, and sequential designing. These results reveal that the efficiency of the design used by the experimenters can be improved dramatically. A sensitivity analysis with respect to the most important parameter is also performed. Second, estimation of the unknown parameters is done using least squares, that is, assuming normality, and also via maximum likelihood assuming the exponential distribution. Simulations for the designs considered in this article show that the variance, the mean squared error, and the bias of the estimators using maximum likelihood are in most cases lower than those using least squares. Supplementary materials for this article are available online.