Biometrika, ISSN 0006-3444
Volume 102, number 2 (june 2015)

Abstract
Model organisms and human studies have yielded increasing empirical evidence that interactions among genes contribute broadly to genetic variation of complex traits. In the presence of gene-gene interactions, the dimensionality of the feature space becomes extremely high relative to the sample size. This poses a significant methodological challenge in the identification of gene-gene interactions. In this paper, by using a Gaussian graphical model framework, we translate the problem of identifying gene-gene interactions associated with a binary trait D into an inference problem on the difference of two high-dimensional precision matrices that summarize the conditional dependence network structures of the genes. We propose a procedure for testing the differential network globally, which is particularly powerful against sparse alternatives. In addition, a multiple testing procedure with false discovery rate control is developed to infer the specific structure of the differential network. Theoretical justification is provided to ensure the validity of the proposed tests, and optimality results are derived under sparsity assumptions. Through a simulation study we demonstrate that the proposed tests maintain the desired error rates under the null hypothesis and have good power under the alternative hypothesis. The methods are applied to a breast cancer gene expression study.

Abstract
We study how to separate signals from noisy data accurately and determine the patterns of the selected signals. Controlling the inflation of false positive errors is important in large-scale simultaneous inference but has not been addressed in the pattern recognition literature. We develop a decision-theoretic framework and formulate the sparse pattern recognition problem as a simultaneous inference problem with multiple decision trees. Oracle and adaptive classifiers are proposed for maximizing the expected number of true positives subject to a constraint on the overall false positive rate. Existing results on multiple testing are extended by allowing more than two states of nature, hierarchical decision-making and new error rate concepts.

Abstract
Meta-analysis is widely used to compare and combine the results of multiple independent studies. To account for between-study heterogeneity, investigators often employ random-effects models, under which the effect sizes of interest are assumed to follow a normal distribution. It is common to estimate the mean effect size by a weighted linear combination of study-specific estimators, with the weight for each study being inversely proportional to the sum of the variance of the effect-size estimator and the estimated variance component of the random-effects distribution. Because the estimator of the variance

component involved in the weights is random and correlated with study-specific effect-size estimators, the commonly adopted asymptotic normal approximation to the meta-analysis estimator is grossly inaccurate unless the number of studies is large. When individual participant data are available, one can also estimate the mean effect size by maximizing the joint likelihood. We establish the asymptotic properties of the meta-analysis estimator and the joint maximum likelihood estimator when the number of studies is either fixed or increases at a slower rate than the study sizes and we discover a surprising result: the former estimator is always at least as efficient as the latter. We also develop a novel resampling technique that improves the accuracy of statistical inference. We demonstrate the benefits of the proposed inference procedures using simulated and empirical data.

Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator P. 295-313

A. Doucet, M. K. Pitt, G. Deligiannidis, R. Kohn

Abstract

When an unbiased estimator of the likelihood is used within a Metropolis–Hastings chain, it is necessary to trade off the number of Monte Carlo samples used to construct this estimator against the asymptotic variances of the averages computed under this chain. Using many Monte Carlo samples will typically result in Metropolis–Hastings averages with lower asymptotic variances than the corresponding averages that use fewer samples; however, the computing time required to construct the likelihood estimator increases with the number of samples. Under the assumption that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with variance inversely proportional to the number of samples and independent of the parameter value at which it is evaluated, we provide guidelines on the number of samples to select. We illustrate our results by considering a stochastic volatility model applied to stock index returns.

A useful variant of the Davis–Kahan theorem for statisticians P. 315-323

Y. Yu, T. Wang, R. J. Samworth

Abstract

The Davis–Kahan theorem is used in the analysis of many statistical procedures to bound the distance between subspaces spanned by population eigenvectors and their sample versions. It relies on an eigenvalue separation condition between certain population and sample eigenvalues. We present a variant of this result that depends only on a population eigenvalue separation condition, making it more natural and convenient for direct application in statistical contexts, and provide an improvement in many cases to the usual bound in the statistical literature. We also give an extension to situations where the matrices under study may be asymmetric or even non-square, and where interest is in the distance between subspaces spanned by corresponding singular vectors.

Information-theoretic optimality of observation-driven time series models for continuous responses P. 325-343

F. Blasques, S. J. Koopman, A. Lucas

Abstract

We investigate information-theoretic optimality properties of the score function of the predictive likelihood as a device for updating a real-valued time-varying parameter in a univariate observation-driven model with continuous responses. We restrict our attention to models with updates of one lag order. The results provide theoretical justification for a class of score-driven models which includes the generalized autoregressive conditional heteroskedasticity model as a special case. Our main contribution is to show that only parameter updates based on the score will always reduce the local Kullback–Leibler divergence between the true conditional density and the model-implied conditional density. This result holds irrespective of the severity of model misspecification. We also show that use of the score leads to a considerably smaller global Kullback–Leibler divergence in empirically relevant settings. We illustrate the theory with an application to time-varying volatility models. We show that the reduction in Kullback–Leibler divergence across a range of different

settings can be substantial compared to updates based on, for example, squared lagged observations.

On the dependence structure of bivariate recurrent event processes: inference and estimation P. 345-358

Jing Ning, Yong Chen, Chunyan Cai, Xuelin Huang, Mei-Cheng Wang

Abstract

Bivariate or multivariate recurrent event processes are often encountered in longitudinal studies in which more than one type of event is of interest. There has been much research on regression analysis for such data, but little has been done to measure the dependence between recurrent event processes. We propose a time-dependent measure, termed the rate ratio, to assess the local dependence between two types of recurrent event processes. We model the rate ratio as a parametric function of time, and leave unspecified all other aspects of the distribution. We develop a composite likelihood procedure for model fitting and parameter estimation. We show that the proposed estimator is consistent and asymptotically normal. Its finite sample performance is evaluated by simulation and illustrated by an application to a soft tissue sarcoma study.

A Möbius transformation-induced distribution on the torus P. 359-370

Shogo Kato, Arthur Pewsey

Abstract

We propose a five-parameter bivariate wrapped Cauchy distribution as a unimodal model for toroidal data. It is highly tractable, displays numerous desirable properties, including marginal and conditional distributions that are all wrapped Cauchy, and arises as an appealing submodel of a six-parameter distribution obtained by applying Möbius transformation to a pre-existing bivariate circular model. Method of moments and maximum likelihood estimation of its parameters are fast, and tests for independence and goodness-of-fit are available. An analysis involving dihedral angles of the proteinogenic amino acid Tyrosine illustrates the distribution's application. A Markov process for circular data is also explored.

Maximum projection designs for computer experiments P. 371-380

V. Roshan Joseph, Evren Gul, Shan Ba

Abstract

Space-filling properties are important in designing computer experiments. The traditional maximin and minimax distance designs consider only space-filling in the full-dimensional space; this can result in poor projections onto lower-dimensional spaces, which is undesirable when only a few factors are active. Restricting maximin distance design to the class of Latin hypercubes can improve one-dimensional projections but cannot guarantee good space-filling properties in larger subspaces. We propose designs that maximize space-filling properties on projections to all subsets of factors. We call our designs maximum projection designs. Our design criterion can be computed at no more cost than a design criterion that ignores projection properties.

Automatic structure recovery for additive models P. 381-395

Yichao Wu, Leonard A. Stefanski

Abstract

We propose an automatic structure recovery method for additive models, based on a backfitting algorithm coupled with local polynomial smoothing, in conjunction with a new kernel-based variable selection strategy. Our method produces estimates of the set of noise predictors, the sets of predictors that contribute polynomially at different degrees up to a specified degree M, and the set of predictors that contribute beyond polynomially of degree M. We prove consistency of the proposed method, and describe an extension to partially linear models. Finite-sample performance of the method is illustrated via Monte Carlo studies and a real-data example.

Jump information criterion for statistical inference in estimating discontinuous curves P. 397-408

Zhiming Xia, Peihua Qiu

Abstract

Nonparametric regression analysis when the regression function is discontinuous has many applications. Existing methods for estimating a discontinuous regression curve usually assume that the number of jumps in the regression curve is known beforehand, which is unrealistic in some situations. Although there has been research on estimation of a discontinuous regression curve when the number of jumps is unknown, the problem remains mostly open because such research often requires assumptions on other related quantities, such as a known minimum jump size. In this paper we propose a jump information criterion which consists of a term measuring the fidelity of the estimated regression curve to the observed data and a penalty related to the number of jumps and the jump sizes. The number of jumps can then be determined by minimizing our criterion. Theoretical and numerical studies show that our method works well.

A validated information criterion to determine the structural dimension in dimension reduction models P. 409-420

Yanyuan Ma, Xinyu Zhang

Abstract

A crucial component of performing sufficient dimension reduction is to determine the structural dimension of the reduction model. We propose a novel information criterion-based method for this purpose, a special feature of which is that when examining the goodness-of-fit of the current model, one needs to perform model evaluation by using an enlarged candidate model. Although the procedure does not require estimation under the enlarged model of dimension $k+1$, the decision as to how well the current model of dimension k fits relies on the validation provided by the enlarged model; thus we call this procedure the validated information criterion, vic(k). Our method is different from existing information criterion-based model selection methods; it breaks free from dependence on the connection between dimension reduction models and their corresponding matrix eigenstructures, which relies heavily on a linearity condition that we no longer assume. We prove consistency of the proposed method, and its finite-sample performance is demonstrated numerically.

Effective dimension reduction for sparse functional data P. 421-437

F. Yao, E. Lei, Y. Wu

We propose a method of effective dimension reduction for functional data, emphasizing the sparse design where one observes only a few noisy and irregular measurements for some or all of the subjects. The proposed method borrows strength across the entire sample and provides a way to characterize the effective dimension reduction space, via functional cumulative slicing. Our theoretical study reveals a bias-variance trade-off associated with the regularizing truncation and decaying structures of the predictor process and the effective dimension reduction space. A simulation study and an application illustrate the superior finite-sample performance of the method.

Envelopes and reduced-rank regression P. 439-456

R. Dennis Cook, Liliana Forzani, Xin Zhang

Abstract

We incorporate the nascent idea of envelopes (Cook et al., *Statist. Sinica* **20**, 927–1010) into reduced-rank regression by proposing a reduced-rank envelope model, which is a hybrid of reduced-rank and envelope regressions. The proposed model has total number of parameters no more than either of reduced-rank regression or envelope regression. The resulting estimator is at least as efficient as both existing estimators. The methodology of this paper can be adapted to other envelope models, such as partial envelopes (Su & Cook, *Biometrika* **98**, 133–46) and envelopes in predictor space (Cook et al., *J. R. Statist. Soc.* B **75**, 851–77).

On the degrees of freedom of reduced-rank estimators in multivariate regression P. 457-477

A. Mukherjee, K. Chen, N. Wang, J. Zhu

Abstract

We study the effective degrees of freedom of a general class of reduced-rank estimators for multivariate regression in the framework of Stein's unbiased risk estimation. A finite-sample exact unbiased estimator is derived that admits a closed-form expression in terms of the thresholded singular values of the least-squares solution and hence is readily computable. The results continue to hold in the high-dimensional setting where both the predictor and the response dimensions may be larger than the sample size. The derived analytical form facilitates the investigation of theoretical properties and provides new insights into the empirical behaviour of the degrees of freedom. In particular, we examine the differences and connections between the proposed estimator and a commonly-used naive estimator. The use of the proposed estimator leads to efficient and accurate prediction risk estimation and model selection, as demonstrated by simulation studies and a data example.

Effective degrees of freedom: a flawed metaphor P. 479-485

Lucas Janson, William Fithian, Trevor J. Hastie

Abstract

To most applied statisticians, a fitting procedure's degrees of freedom is synonymous with its model complexity, or its capacity for overfitting to data. In particular, the degrees of freedom is often used to parameterize the bias-variance trade-off in model selection. We argue that, on the contrary, model complexity and degrees of freedom may correspond very poorly. We exhibit and theoretically explore various fitting procedures for which the degrees of freedom is not monotonic in the model complexity parameter and can exceed the total dimension of the ambient space even in very simple settings. We show that the degrees of freedom for any nonconvex projection method can be unbounded.

Semiparametric exponential families for heavy-tailed data P. 486-493

William Fithian, Stefan Wager

Abstract

We propose a semiparametric method for fitting the tail of a heavy-tailed population given a relatively small sample from that population and a larger sample from a related background population. We model the tail of the small sample as an exponential tilt of the better-observed large-sample tail, using a robust sufficient statistic motivated by extreme value theory. In particular, our method induces an estimator of the small-population mean, and we give theoretical and empirical evidence that this estimator outperforms methods that do not use the background sample. We demonstrate substantial efficiency gains over competing methods in simulation and on data from a large controlled experiment conducted by Facebook.

Optimum designs for two treatments with unequal variances in the presence of covariates P. 494-499

A. C. Atkinson

Abstract

Optimum designs are described for two treatments with different variances when covariates are included in the model. The designs, a generalization of Neyman allocation, are required in personalized medicine to model the effect of covariates on the choice of treatment. The use of the designs in clinical trials is indicated. D-optimality of the designs is established using results from Kiefer's general equivalence theorem. The results are obtained with the use of surprisingly elementary algebra.