Testing for stationarity of functional time series in the frequency domain
ALEXANDER AUE AND ANNE VAN DELFT 2505
On spike and slab empirical Bayes multiple testing
ISMAEL CASTILLO AND ÉTIENNE ROQUAIN 2548
Theoretical and computational guarantees of mean field variational inference for
community detection ........... ANDERSON Y. ZHANG AND HARRISON H. ZHOU 2575
Minimax optimal sequential hypothesis tests for Markov processes
MICHAEL FAUSS, ABDELHAK M. ZOUBIR AND H. VINCENT POOR 2599
Test of significance for high-dimensional longitudinal data
ETHAN X. FANG, YANG NING AND RUNZE LI 2622
Geometrizing rates of convergence under local differential privacy constraints
ANGELIKA ROHDE AND LUKAS STEINBERGER 2646
Additive regression with Hilbertian responses
JEONG MIN JEON AND BYEONG U. PARK 2671
Nonparametric Bayesian estimation for multivariate Hawkes processes
SOPHIE DONNET, VINCENT RIVOIRARD AND JUDITH ROUSSEAU 2698
Hypothesis testing for high-dimensional time series via self-normalization
RUNMIN WANG AND XIAOFENG SHAO 2728
Variational analysis of constrained M-estimators
JOHANNES O. ROYSET AND ROGER J-B WETS 2759
Which bridge estimator is the best for variable selection?
SHUAI WANG, HAOLEI WENG AND ARIAN MALEKI 2791
Permutation methods for factor analysis and PCA ................. EDGAR DOBRIBAN 2824
A general framework for Bayes structured linear models
CHAO GAO, AAD W. VAN DER VAART AND HARRISON H. ZHOU 2848
Asymptotic distribution and detection thresholds for two-sample tests based on geometric
graphs ........................................ BHASWAR B. BHATTACHARYA 2879
Controlled sequential Monte Carlo ............... JEREMY HENG, ADRIAN N. BISHOP,
GEORGE DELIGIANNIDIS AND ARNAUD DOUCET 2904
A framework for adaptive MCMC targeting multimodal distributions
EMILIA POMPE, CHRIS HOLMES AND KRZYSZTOF ŁATUSZYŃSKI 2930
Valid post-selection inference in model-free linear regression
ARUN K. KUCHIBHOTLA, LAWRENCE D. BROWN, ANDREAS BUJA,
JUNHUI CAI, EDWARD I. GEORGE AND LINDA H. ZHAO 2953
Inference for spherical location under high concentration
DAVY PAINDAINE AND THOMAS VERDEBOUT 2982
Semiparametric Bayesian causal inference . . KOLYAN RAY AND AAD VAN DER VAART 2999
Relaxing the assumptions of knockoffs by conditioning
DONGMING HUANG AND LUCAS JANSON 3021
Analytical nonlinear shrinkage of large-dimensional covariance matrices
OLIVIER LEDOIT AND MICHAEL WOLF 3043
Coupled conditional backward sampling particle filter
ANTHONY LEE, SUMEETPAL S. SINGH AND MATTI VIHOLA 3066
Asymptotic risk and phase transition of $l_1$-penalized robust estimator . . HANWEN HUANG 3090
The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

President-Elect: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

Past President: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

The Annals of Statistics. Editors: Richard J. Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK. Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027, USA

The Annals of Applied Statistics. Editor-in-Chief: Karen Kafadar, Department of Statistics, University of Virginia, Heidelberg Institute for Theoretical Studies, Charlottesville, VA 22904-4135, USA.

The Annals of Probability. Editor: Amir Dembo, Department of Statistics and Department of Mathematics, Stanford University, Stanford, California 94305, USA


Statistical Science. Editor: Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. Editor: Vlada Limic, UMR 7501 de l’Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

The Annals of Statistics [ISSN 0090-5364 (print); ISSN 2168-8966 (online)], Volume 48, Number 5, October 2020. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to The Annals of Statistics, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

Copyright © 2020 by the Institute of Mathematical Statistics

Printed in the United States of America
TESTING FOR STATIONARITY OF FUNCTIONAL TIME SERIES IN THE FREQUENCY DOMAIN

BY ALEXANDER AUE1 AND ANNE VAN DELFT2

1Department of Statistics, University of California, aaue@ucdavis.edu
2Fakultät für Mathematik, Ruhr-Universität Bochum, Anne.vanDelft@rub.de

Interest in functional time series has spiked in the recent past with papers covering both methodology and applications being published at a much increased pace. This article contributes to the research in this area by proposing a new stationarity test for functional time series based on frequency domain methods. The proposed test statistics is based on joint dimension reduction via functional principal components analysis across the spectral density operators at all Fourier frequencies, explicitly allowing for frequency-dependent levels of truncation to adapt to the dynamics of the underlying functional time series. The properties of the test are derived both under the null hypothesis of stationary functional time series and under the smooth alternative of locally stationary functional time series. The methodology is theoretically justified through asymptotic results. Evidence from simulation studies and an application to annual temperature curves suggests that the test works well in finite samples.

REFERENCES


Key words and phrases. Frequency domain methods, functional data analysis, locally stationary processes, spectral analysis.
ON SPIKE AND SLAB EMPIRICAL BAYES MULTIPLE TESTING

BY ISMAËL CASTILLO * AND ÉTIENNE ROQUAIN **

Laboratoire de Probabilités, Statistique et Modélisation, LPSM, Sorbonne Université, *ismael.castillo@upmc.fr; **etienne.roquain@upmc.fr

This paper explores a connection between empirical Bayes posterior distributions and false discovery rate (FDR) control. In the Gaussian sequence model this work shows that empirical Bayes-calibrated spike and slab posterior distributions allow a correct FDR control under sparsity. Doing so, it offers a frequentist theoretical validation of empirical Bayes methods in the context of multiple testing. Our theoretical results are illustrated with numerical experiments.

REFERENCES


MSC2020 subject classifications. 62C12, 62G10.

Key words and phrases. Frequentist properties of Bayesian procedures, false discovery rate, sparsity, multiple testing.


THEORETICAL AND COMPUTATIONAL GUARANTEES OF MEAN FIELD VARIATIONAL INFERENCE FOR COMMUNITY DETECTION

BY ANDERSON Y. ZHANG¹ AND HARRISON H. ZHOU²

¹Department of Statistics, The Wharton School, University of Pennsylvania, ayz@wharton.upenn.edu
²Department of Statistics, Yale University, huibin.zhou@yale.edu

The mean field variational Bayes method is becoming increasingly popular in statistics and machine learning. Its iterative coordinate ascent variational inference algorithm has been widely applied to large scale Bayesian inference. See Blei et al. (2017) for a recent comprehensive review. Despite the popularity of the mean field method, there exist remarkably little fundamental theoretical justifications. To the best of our knowledge, the iterative algorithm has never been investigated for any high-dimensional and complex model. In this paper, we study the mean field method for community detection under the stochastic block model. For an iterative batch coordinate ascent variational inference algorithm, we show that it has a linear convergence rate and converges to the minimax rate within \( \log n \) iterations. This complements the results of Bickel et al. (2013) which studied the global minimum of the mean field variational Bayes and obtained asymptotic normal estimation of global model parameters. In addition, we obtain similar optimality results for Gibbs sampling and an iterative procedure to calculate maximum likelihood estimation, which can be of independent interest.

REFERENCES


MSC2020 subject classifications. 60G05.
Key words and phrases. Mean field, variational inference, Bayesian, community detection, stochastic block model.
Under mild Markov assumptions, sufficient conditions for strict minimax optimality of sequential tests for multiple hypotheses under distributional uncertainty are derived. First, the design of optimal sequential tests for simple hypotheses is revisited, and it is shown that the partial derivatives of the corresponding cost function are closely related to the performance metrics of the underlying sequential test. Second, an implicit characterization of the least favorable distributions for a given testing policy is stated. By combining the results on optimal sequential tests and least favorable distributions, sufficient conditions for a sequential test to be minimax optimal under general distributional uncertainties are obtained. The cost function of the minimax optimal test is further identified as a generalized $f$-dissimilarity and the least favorable distributions as those that are most similar with respect to this dissimilarity. Numerical examples for minimax optimal sequential tests under different uncertainties illustrate the theoretical results.

REFERENCES


MSC2020 subject classifications. Primary 62L10; secondary 62C20.
Key words and phrases. Sequential analysis, minimax procedures, distributional uncertainty, robust hypothesis testing, multiple hypothesis testing.
TEST OF SIGNIFICANCE FOR HIGH-DIMENSIONAL LONGITUDINAL DATA

BY ETHAN X. FANG¹,* YANG NING² AND RUNZE LI¹,**

¹Department of Statistics, Pennsylvania State University. *xxf13@psu.edu; **rzli@psu.edu
²Department of Statistics and Data Science, Cornell University, yn265@cornell.edu

This paper concerns statistical inference for longitudinal data with ultrahigh dimensional covariates. We first study the problem of constructing confidence intervals and hypothesis tests for a low-dimensional parameter of interest. The major challenge is how to construct a powerful test statistic in the presence of high-dimensional nuisance parameters and sophisticated within-subject correlation of longitudinal data. To deal with the challenge, we propose a new quadratic decorrelated inference function approach which simultaneously removes the impact of nuisance parameters and incorporates the correlation to enhance the efficiency of the estimation procedure. When the parameter of interest is of fixed dimension, we prove that the proposed estimator is asymptotically normal and attains the semiparametric information bound, based on which we can construct an optimal Wald test statistic. We further extend this result and establish the limiting distribution of the estimator under the setting with the dimension of the parameter of interest growing with the sample size at a polynomial rate. Finally, we study how to control the false discovery rate (FDR) when a vector of high-dimensional regression parameters is of interest. We prove that applying the Storey (J. R. Stat. Soc. Ser. B. Stat. Methodol. 64 (2002) 479–498) procedure to the proposed test statistics for each regression parameter controls FDR asymptotically in longitudinal data. We conduct simulation studies to assess the finite sample performance of the proposed procedures. Our simulation results imply that the newly proposed procedure can control both Type I error for testing a low-dimensional parameter of interest and the FDR in the multiple testing problem. We also apply the proposed procedure to a real data example.

REFERENCES


MSC2020 subject classifications. Primary 62F03; secondary 62F05.

Key words and phrases. False discovery rate, generalized estimating equation, quadratic inference function.
GEOMETRIZING RATES OF CONVERGENCE UNDER LOCAL DIFFERENTIAL PRIVACY CONSTRAINTS

BY ANGELIKA ROHDE\textsuperscript{1} AND LUKAS STEINBERGER\textsuperscript{2}

\textsuperscript{1}University of Freiburg, angelika.rohde@stochastik.uni-freiburg.de
\textsuperscript{2}Department of Statistics and OR, University of Vienna, lukas.steinberger@univie.ac.at

We study the problem of estimating a functional $\theta(P)$ of an unknown probability distribution $P \in \mathcal{P}$ in which the original iid sample $X_1, \ldots, X_n$ is kept private even from the statistician via an $\alpha$-local differential privacy constraint. Let $\omega_{TV}$ denote the modulus of continuity of the functional $\theta$ over $\mathcal{P}$ with respect to total variation distance. For a large class of loss functions $l$ and a fixed privacy level $\alpha$, we prove that the privatized minimax risk is equivalent to $l(\omega_{TV}(n^{-1/2}))$ to within constants, under regularity conditions that are satisfied, in particular, if $\theta$ is linear and $\mathcal{P}$ is convex. Our results complement the theory developed by Donoho and Liu (1991) with the nowadays highly relevant case of privatized data. Somewhat surprisingly, the difficulty of the estimation problem in the private case is characterized by $\omega_{TV}$, whereas, it is characterized by the Hellinger modulus of continuity if the original data $X_1, \ldots, X_n$ are available. We also find that for locally private estimation of linear functionals over a convex model a simple sample mean estimator, based on independently and binary privatized observations, always achieves the minimax rate. We further provide a general recipe for choosing the functional parameter in the optimal binary privatization mechanisms and illustrate the general theory in numerous examples. Our theory allows us to quantify the price to be paid for local differential privacy in a large class of estimation problems. This price appears to be highly problem specific.

REFERENCES


MSC2020 subject classifications. Primary 62G05; secondary 62C20.

Key words and phrases. Local differential privacy, minimax estimation, rate of convergence, moduli of continuity, nonparametric estimation.


ADDITIVE REGRESSION WITH HILBERTIAN RESPONSES

BY JEONG MIN JEON* AND BYEONG U. PARK**

Department of Statistics, Seoul National University, * jeongmin.jeon@kuleuven.be; ** bupark@stats.snu.ac.kr

This paper develops a foundation of methodology and theory for the estimation of structured nonparametric regression models with Hilbertian responses. Our method and theory are focused on the additive model, while the main ideas may be adapted to other structured models. For this, the notion of Bochner integration is introduced for Banach-space-valued maps as a generalization of Lebesgue integration. Several statistical properties of Bochner integrals, relevant for our method and theory and also of importance in their own right, are presented for the first time. Our theory is complete. The existence of our estimators and the convergence of a practical algorithm that evaluates the estimators are established. These results are nonasymptotic as well as asymptotic. Furthermore, it is proved that the estimators achieve the univariate rates in pointwise, $L^2$ and uniform convergence, and that the estimators of the component maps converge jointly in distribution to Gaussian random elements. Our numerical examples include the cases of functional, density-valued and simplex-valued responses, demonstrating the validity of our approach.

REFERENCES


MSC2020 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Additive models, smooth backfitting, Bochner integral, non-Euclidean data, infinite-dimensional spaces, Hilbert spaces, functional responses.


NONPARAMETRIC BAYESIAN ESTIMATION FOR MULTIVARIATE HAWKES PROCESSES

By Sophie Donnet¹, Vincent Rivoirard²,* and Judith Rousseau²,**

¹INRA, Université Paris Saclay, sophie.donnet@inra.fr
²CEREMADE, CNRS, UMR 7534, Université Paris–Dauphine, *Vincent.Rivoirard@dauphine.fr; **Judith.Rousseau@stats.ox.ac.uk

This paper studies nonparametric estimation of parameters of multivariate Hawkes processes. We consider the Bayesian setting and derive posterior concentration rates. First, rates are derived for $L_1$-metrics for stochastic intensities of the Hawkes process. We then deduce rates for the $L_1$-norm of interactions functions of the process. Our results are exemplified by using priors based on piecewise constant functions, with regular or random partitions and priors based on mixtures of Betas distributions. We also present a simulation study to illustrate our results and to study empirically the inference on functional connectivity graphs of neurons

REFERENCES


MSC2020 subject classifications. Primary 62G20, 60G55; secondary 62G05.
Key words and phrases. Multivariate counting process, Hawkes processes, nonparametric Bayesian estimation, posterior concentration rates.


HYPOTHESIS TESTING FOR HIGH-DIMENSIONAL TIME SERIES VIA SELF-NORMALIZATION

BY RUNMIN WANG* AND XIAOFENG SHAO**

Department of Statistics, University of Illinois at Urbana-Champaign, *rwang52@illinois.edu; **xshao@illinois.edu

Self-normalization has attracted considerable attention in the recent literature of time series analysis, but its scope of applicability has been limited to low-/fixed-dimensional parameters for low-dimensional time series. In this article, we propose a new formulation of self-normalization for inference about the mean of high-dimensional stationary processes. Our original test statistic is a U-statistic with a trimming parameter to remove the bias caused by weak dependence. Under the framework of nonlinear causal processes, we show the asymptotic normality of our U-statistic with the convergence rate dependent upon the order of the Frobenius norm of the long-run covariance matrix. The self-normalized test statistic is then constructed on the basis of recursive subsampled U-statistics and its limiting null distribution is shown to be a functional of time-changed Brownian motion, which differs from the pivotal limit used in the low-dimensional setting. An interesting phenomenon associated with self-normalization is that it works in the high-dimensional context even if the convergence rate of original test statistic is unknown. We also present applications to testing for bandedness of the covariance matrix and testing for white noise for high-dimensional stationary time series and compare the finite sample performance with existing methods in simulation studies. At the root of our theoretical arguments, we extend the martingale approximation to the high-dimensional setting, which could be of independent theoretical interest.

REFERENCES


Ann. Statist. 43 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315


Key words and phrases. High-dimensional inference, nonlinear time series, pivotal, self-normalization, U-statistic.


VARIATIONAL ANALYSIS OF CONSTRAINED M-ESTIMATORS

BY JOHANNES O. ROYSET\textsuperscript{1} AND ROGER J-B WETS\textsuperscript{2}

\begin{flushleft}
\textsuperscript{1}Operations Research Department, Naval Postgraduate School, joroyset@nps.edu
\textsuperscript{2}Department of Mathematics, University of California, Davis, rjwets@ucdavis.edu
\end{flushleft}

We propose a unified framework for establishing existence of nonparametric $M$-estimators, computing the corresponding estimates, and proving their strong consistency when the class of functions is exceptionally rich. In particular, the framework addresses situations where the class of functions is complex involving information and assumptions about shape, pointwise bounds, location of modes, height at modes, location of level-sets, values of moments, size of subgradients, continuity, distance to a “prior” function, multivariate total positivity and any combination of the above. The class might be engineered to perform well in a specific setting even in the presence of little data. The framework views the class of functions as a subset of a particular metric space of upper semicontinuous functions under the Attouch–Wets distance. In addition to allowing a systematic treatment of numerous $M$-estimators, the framework yields consistency of plug-in estimators of modes of densities, maximizers of regression functions, level-sets of classifiers and related quantities, and also enables computation by means of approximating parametric classes. We establish consistency through a one-sided law of large numbers, here extended to sieves, that relaxes assumptions of uniform laws, while ensuring global approximations even under model misspecification.

REFERENCES


MSC2020 subject classifications. 62G07.
Key words and phrases. Shape-constrained estimation, variational approximations.


WHICH BRIDGE ESTIMATOR IS THE BEST FOR VARIABLE SELECTION?

BY SHUAIWEN WANG¹,*, HAOLEI WENG² AND ARIAN MALEKI¹,†

¹Department of Statistics, Columbia University; *sw2853@columbia.edu; †arian@stat.columbia.edu
²Department of Statistics and Probability, Michigan State University, wenghaol@msu.edu

We study the problem of variable selection for linear models under the high-dimensional asymptotic setting, where the number of observations $n$ grows at the same rate as the number of predictors $p$. We consider two-stage variable selection techniques (TVS) in which the first stage uses bridge estimators to obtain an estimate of the regression coefficients, and the second stage simply thresholds this estimate to select the “important” predictors. The asymptotic false discovery proportion (AFDP) and true positive proportion (ATPP) of these TVS are evaluated. We prove that for a fixed ATPP, in order to obtain a smaller AFDP, one should pick a bridge estimator with smaller asymptotic mean square error in the first stage of TVS. Based on such principled discovery, we present a sharp comparison of different TVS, via an in-depth investigation of the estimation properties of bridge estimators. Rather than “orderwise” error bounds with loose constants, our analysis focuses on precise error characterization. Various interesting signal-to-noise ratio and sparsity settings are studied. Our results offer new and thorough insights into high-dimensional variable selection. For instance, we prove that a TVS with Ridge in its first stage outperforms TVS with other bridge estimators in large noise settings; two-stage LASSO becomes inferior when the signal is rare and weak. As a by-product, we show that two-stage methods outperform some standard variable selection techniques, such as LASSO and Sure Independence Screening, under certain conditions.

REFERENCES


MSC2020 subject classifications. 62J05, 62J07.

Key words and phrases. Variable selection, high dimension, bridge regression, two-stage methods, false discovery proportion, true positive proportion, rare signal, large noise, large sample, debiasing.


Researchers often have datasets measuring features $x_{ij}$ of samples, such as test scores of students. In factor analysis and PCA, these features are thought to be influenced by unobserved factors, such as skills. Can we determine how many components affect the data? This is an important problem, because decisions made here have a large impact on all downstream data analysis. Consequently, many approaches have been developed. *Parallel Analysis* is a popular permutation method: it randomly scrambles each feature of the data. It selects components if their singular values are larger than those of the permuted data. Despite widespread use, as well as empirical evidence for its accuracy, it currently has no theoretical justification.

In this paper, we show that parallel analysis (or permutation methods) consistently select the large components in certain high-dimensional factor models. However, when the signals are too large, the smaller components are not selected. The intuition is that permutations keep the noise invariant, while “destroying” the low-rank signal. This provides justification for permutation methods. Our work also uncovers drawbacks of permutation methods, and paves the way to improvements.

REFERENCES


**MSC2020 subject classifications.** Primary 62H25; secondary 62H12.

**Key words and phrases.** Factor analysis, PCA, parallel analysis, permutation methods, high-dimensional asymptotics.


A GENERAL FRAMEWORK FOR BAYES STRUCTURED LINEAR MODELS

BY CHAO GAO\(^1\), AAD W. VAN DER VAART\(^2\) AND HARRISON H. ZHOU\(^3\)

\(^1\)Department of Statistics, University of Chicago, chaogao@galton.uchicago.edu
\(^2\)Mathematical Institute, Faculty of Science, Leiden University, avdvaart@math.leidenuniv.nl
\(^3\)Department of Statistics, Yale University, huibin.zhou@yale.edu

High dimensional statistics deals with the challenge of extracting structured information from complex model settings. Compared with a large number of frequentist methodologies, there are rather few theoretically optimal Bayes methods for high dimensional models. This paper provides a unified approach to both Bayes high dimensional statistics and Bayes nonparametrics in a general framework of structured linear models. With a proposed two-step prior, we prove a general oracle inequality for posterior contraction under an abstract setting that allows model misspecification. The general result can be used to derive new results on optimal posterior contraction under many complex model settings including recent works for stochastic block model, graphon estimation and dictionary learning. It can also be used to improve upon posterior contraction results in literature including sparse linear regression and nonparametric aggregation. The key of the success lies in the novel two-step prior distribution: one for model structure, that is, model selection, and the other one for model parameters. The prior on the parameters of a model is an elliptical Laplace distribution that is capable of modeling signals with large magnitude, and the prior on the model structure involves a factor that compensates the effect of the normalizing constant of the elliptical Laplace distribution, which is important to attain rate-optimal posterior contraction.

REFERENCES


MSC2020 subject classifications. Primary 62C10; secondary 62F15.
Key words and phrases. Oracle inequality, stochastic block model, graphon, sparse linear regression, aggregation, dictionary learning, posterior contraction.


ASYMPTOTIC DISTRIBUTION AND DETECTION_THRESHOLDS FOR TWO-SAMPLE TESTS BASED ON GEOMETRIC GRAPHS

BY BHASWAR B. BHATTACHARYA

Department of Statistics, The Wharton School, University of Pennsylvania, bhaswar@wharton.upenn.edu

In this paper, we consider the problem of testing the equality of two multivariate distributions based on geometric graphs constructed using the inter-point distances between the observations. These include the tests based on the minimum spanning tree and the \(K\)-nearest neighbor (NN) graphs, among others. These tests are asymptotically distribution-free, universally consistent and computationally efficient, making them particularly useful in modern applications. However, very little is known about the power properties of these tests. In this paper, using the theory of stabilizing geometric graphs, we derive the asymptotic distribution of these tests under general alternatives, in the Poissonized setting. Using this, the detection threshold and the limiting local power of the test based on the \(K\)-NN graph are obtained, where interesting exponents depending on dimension emerge. This provides a way to compare and justify the performance of these tests in different examples.

REFERENCES


MSC2020 subject classifications. 62F07, 62G10, 60D05, 60F05, 60C05.

Key words and phrases. Efficiency, local power, geometric probability, nearest-neighbor graphs, nonparametric hypothesis testing.


CONTROLLED SEQUENTIAL MONTE CARLO

BY JEREMY HENG\(^1\), ADRIAN N. BISHOP\(^2\), GEORGE DELIGIANNIDIS\(^3,\ast\) and ARNAUD DOUCET\(^3,\ast\ast\)

\(^1\)ESSEC Business School, heng@essec.edu
\(^2\)CSIRO, adrian.bishop@uts.edu.au
\(^3\)University of Oxford, \(*\) deligian@stats.ox.ac.uk; \(**\) doucet@stats.ox.ac.uk

Sequential Monte Carlo methods, also known as particle methods, are a popular set of techniques for approximating high-dimensional probability distributions and their normalizing constants. These methods have found numerous applications in statistics and related fields; for example, for inference in nonlinear non-Gaussian state space models, and in complex static models. Like many Monte Carlo sampling schemes, they rely on proposal distributions which crucially impact their performance. We introduce here a class of controlled sequential Monte Carlo algorithms, where the proposal distributions are determined by approximating the solution to an associated optimal control problem using an iterative scheme. This method builds upon a number of existing algorithms in econometrics, physics and statistics for inference in state space models, and generalizes these methods so as to accommodate complex static models. We provide a theoretical analysis concerning the fluctuation and stability of this methodology that also provides insight into the properties of related algorithms. We demonstrate significant gains over state-of-the-art methods at a fixed computational complexity on a variety of applications.

REFERENCES


**MSC2020 subject classifications.** Primary 62M05; secondary 62F12, 62M10.

**Key words and phrases.** State space models, annealed importance sampling, normalizing constants, optimal control, approximate dynamic programming, reinforcement learning.
A FRAMEWORK FOR ADAPTIVE MCMC TARGETING MULTIMODAL DISTRIBUTIONS

By Emilia Pompe1,∗, Chris Holmes1,† and Krzysztof Łatuszyński2

1Department of Statistics, University of Oxford, ∗emilia.pompe@stats.ox.ac.uk; †cholmes@stats.ox.ac.uk
2Department of Statistics, University of Warwick, K.G.Latuszynski@warwick.ac.uk

We propose a new Monte Carlo method for sampling from multimodal distributions. The idea of this technique is based on splitting the task into two: finding the modes of a target distribution π and sampling, given the knowledge of the locations of the modes. The sampling algorithm relies on steps of two types: local ones, preserving the mode; and jumps to regions associated with different modes. Besides, the method learns the optimal parameters of the algorithm, while it runs, without requiring user intervention. Our technique should be considered as a flexible framework, in which the design of moves can follow various strategies known from the broad MCMC literature.

In order to design an adaptive scheme that facilitates both local and jump moves, we introduce an auxiliary variable representing each mode, and we define a new target distribution ̃π on an augmented state space X × I, where X is the original state space of π and I is the set of the modes. As the algorithm runs and updates its parameters, the target distribution ̃π also keeps being modified. This motivates a new class of algorithms, Auxiliary Variable Adaptive MCMC. We prove general ergodic results for the whole class before specialising to the case of our algorithm.

REFERENCES


MSC2020 subject classifications. Primary 60J05, 65C05; secondary 62F15.
Key words and phrases. Multimodal distribution, adaptive MCMC, ergodicity.


VALID POST-SELECTION INFEERENCE IN MODEL-FREE LINEAR REGRESSION

BY ARUN K. KUCHIBHOTLA *, LAWRENCE D. BROWN, ANDREAS BUJA, JUNHUI CAI, EDWARD I. GEORGE AND LINDA H. ZHAO

Department of Statistics, The Wharton School, University of Pennsylvania, *arunku@wharton.upenn.edu

Modern data-driven approaches to modeling make extensive use of covariate/model selection. Such selection incurs a cost: it invalidates classical statistical inference. A conservative remedy to the problem was proposed by Berk et al. (Ann. Statist. 41 (2013) 802–837) and further extended by Bachoc, Preinerstorfer and Steinberger (2016). These proposals, labeled “PoSI methods,” provide valid inference after arbitrary model selection. They are computationally NP-hard and have limitations in their theoretical justifications. We therefore propose computationally efficient confidence regions, named “UPoSI” and prove large-p asymptotics for them. We do this for linear OLS regression allowing misspecification of the normal linear model, for both fixed and random covariates, and for independent as well as some types of dependent data. We start by proving a general equivalence result for the post-selection inference problem and a simultaneous inference problem in a setting that strips inessential features still present in a related result of Berk et al. (Ann. Statist. 41 (2013) 802–837). We then construct valid PoSI confidence regions that are the first to have vastly improved computational efficiency in that the required computation times grow only quadratically rather than exponentially with the total number of covariates. These are also the first PoSI confidence regions with guaranteed asymptotic validity when the total number of covariates diverges (almost exponentially) with the sample size. Under standard tail assumptions, we only require \((\log p)^7 = o(n)\) and \(k = o(\sqrt{n}/\log p)\) where \(k (\leq p)\) is the largest number of covariates (model size) considered for selection. We study various properties of these confidence regions, including their Lebesgue measures, and compare them theoretically with those proposed previously.

REFERENCES


Key words and phrases. Simultaneous inference, multiplier bootstrap, uniform consistency, high-dimensional linear regression, concentration inequalities, Orlicz norms, model selection.
INFEREN CE FOR SPHERICAL LOCATION UNDER HIGH CONCENTRATION

BY DAVY PAINDAVEINE1,2,* AND THOMAS VERDEBOUT1,**

1Université libre de Bruxelles. * dpaindav@ulb.ac.be; ** tverdebo@ulb.ac.be
2Université Toulouse I Capitole

Motivated by the fact that circular or spherical data are often much concentrated around a location $\theta$, we consider inference about $\theta$ under high concentration asymptotic scenarios for which the probability of any fixed spherical cap centered at $\theta$ converges to one as the sample size $n$ diverges to infinity. Rather than restricting to Fisher–von Mises–Langevin distributions, we consider a much broader, semiparametric, class of rotationally symmetric distributions indexed by the location parameter $\theta$, a scalar concentration parameter $\kappa$ and a functional nuisance $f$. We determine the class of distributions for which high concentration is obtained as $\kappa$ diverges to infinity. For such distributions, we then consider inference (point estimation, confidence zone estimation, hypothesis testing) on $\theta$ in asymptotic scenarios where $\kappa n$ diverges to infinity at an arbitrary rate with the sample size $n$. Our asymptotic investigation reveals that, interestingly, optimal inference procedures on $\theta$ show consistency rates that depend on $f$. Using asymptotics “à la Le Cam,” we show that the spherical mean is, at any $f$, a parametrically superefficient estimator of $\theta$ and that the Watson and Wald tests for $H_0 : \theta = \theta_0$ enjoy similar, nonstandard, optimality properties. We illustrate our results through simulations and treat a real data example. On a technical point of view, our asymptotic derivations require challenging expansions of rotationally symmetric functionals for large arguments of $f$.

REFERENCES


Key words and phrases. Concentrated distributions, directional statistics, Le Cam’s asymptotic theory of statistical experiments, local asymptotic normality, superefficiency.


SEMIPARAMETRIC BAYESIAN CAUSAL INFERENCE

BY KOLYAN RAY\textsuperscript{1} AND AAD VAN DER VAART\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, Imperial College London, kolyan.ray@imperial.ac.uk
\textsuperscript{2}Mathematical Institute, Leiden University, avdvaart@math.leidenuniv.nl

We develop a semiparametric Bayesian approach for estimating the mean response in a missing data model with binary outcomes and a nonparametrically modelled propensity score. Equivalently, we estimate the causal effect of a treatment, correcting nonparametrically for confounding. We show that standard Gaussian process priors satisfy a semiparametric Bernstein–von Mises theorem under smoothness conditions. We further propose a novel propensity score-dependent prior that provides efficient inference under strictly weaker conditions. We also show that it is theoretically preferable to model the covariate distribution with a Dirichlet process or Bayesian bootstrap, rather than modelling its density.

REFERENCES


MSC2020 subject classifications. Primary 62G20; secondary 62G15, 62G08.

Key words and phrases. Bernstein–von Mises, Gaussian processes, propensity score-dependent priors, causal inference, Dirichlet process.


RELAXING THE ASSUMPTIONS OF KNOCKOFFS BY CONDITIONING

BY DONGMING HUANG* AND LUCAS JANSON**

Department of Statistics, Harvard University; * dhuang01@g.harvard.edu; ** ljanson@fas.harvard.edu

The recent paper Candès et al. (J. R. Stat. Soc. Ser. B. Stat. Methodol. 80 (2018) 551–577) introduced model-X knockoffs, a method for variable selection that provably and nonasymptotically controls the false discovery rate with no restrictions or assumptions on the dimensionality of the data or the conditional distribution of the response given the covariates. The one requirement for the procedure is that the covariate samples are drawn independently and identically from a precisely-known (but arbitrary) distribution. The present paper shows that the exact same guarantees can be made without knowing the covariate distribution fully, but instead knowing it only up to a parametric model with as many as \( \Omega(n^* p) \) parameters, where \( p \) is the dimension and \( n^* \) is the number of covariate samples (which may exceed the usual sample size \( n \) of labeled samples when unlabeled samples are also available). The key is to treat the covariates as if they are drawn conditionally on their observed value for a sufficient statistic of the model. Although this idea is simple, even in Gaussian models conditioning on a sufficient statistic leads to a distribution supported on a set of zero Lebesgue measure, requiring techniques from topological measure theory to establish valid algorithms. We demonstrate how to do this for three models of interest, with simulations showing the new approach remains powerful under the weaker assumptions.

REFERENCES


MSC2020 subject classifications. Primary 62G10; secondary 62B05, 62J02.

Key words and phrases. High-dimensional inference, knockoffs, model-X, sufficient statistic, false discovery rate (FDR), topological measure, graphical model.


ANALYTICAL NONLINEAR SHRINKAGE OF LARGE-DIMENSIONAL COVARIANCE MATRICES

BY OLIVIER LEDOIT* AND MICHAEL WOLF**

Department of Economics, University of Zurich, *olivier.ledoit@econ.uzh.ch; **michael.wolf@econ.uzh.ch

This paper establishes the first analytical formula for nonlinear shrinkage estimation of large-dimensional covariance matrices. We achieve this by identifying and mathematically exploiting a deep connection between nonlinear shrinkage and nonparametric estimation of the Hilbert transform of the sample spectral density. Previous nonlinear shrinkage methods were of numerical nature: QuEST requires numerical inversion of a complex equation from random matrix theory whereas NERCOME is based on a sample-splitting scheme. The new analytical method is more elegant and also has more potential to accommodate future variations or extensions. Immediate benefits are (i) that it is typically 1000 times faster with basically the same accuracy as QuEST and (ii) that it accommodates covariance matrices of dimension up to 10,000 and more. The difficult case where the matrix dimension exceeds the sample size is also covered.

REFERENCES


DAWSON, H. G. (1898). On the Numerical Value of \[ \int_0^h \text{exp}(x^2) \, dx \]


MSC2020 subject classifications. Primary 62H12; secondary 62G20, 15A52.

Key words and phrases. Hilbert transform, kernel estimation, rotation equivariance.
COUPLED CONDITIONAL BACKWARD SAMPLING PARTICLE FILTER

BY ANTHONY LEE¹, SUMEETPAL S. SINGH² AND MATTI VIHOLA³

¹School of Mathematics, University of Bristol, anthony.lee@bristol.ac.uk
²Department of Engineering, University of Cambridge
³Department of Mathematics and Statistics, University of Jyväskylä

The conditional particle filter (CPF) is a promising algorithm for general hidden Markov model smoothing. Empirical evidence suggests that the variant of CPF with backward sampling (CBPF) performs well even with long time series. Previous theoretical results have not been able to demonstrate the improvement brought by backward sampling, whereas we provide rates showing that CBPF can remain effective with a fixed number of particles independent of the time horizon. Our result is based on analysis of a new coupling of two CBPFs, the coupled conditional backward sampling particle filter (CCBPF). We show that CCBPF has good stability properties in the sense that with fixed number of particles, the coupling time in terms of iterations increases only linearly with respect to the time horizon under a general (strong mixing) condition. The CCBPF is useful not only as a theoretical tool, but also as a practical method that allows for unbiased estimation of smoothing expectations, following the recent developments by Jacob, Lindsten and Schön (2020). Unbiased estimation has many advantages, such as enabling the construction of asymptotically exact confidence intervals and straightforward parallelisation.

REFERENCES


MSC2020 subject classifications. Primary 65C40; secondary 65C05, 65C35, 65C60.

Key words and phrases. Backward sampling, convergence rate, coupling, conditional particle filter, unbiased.
Mean square error (MSE) of the estimator can be used to evaluate the performance of a regression model. In this paper, we derive the asymptotic MSE of \( l_1 \)-penalized robust estimators in the limit of both sample size \( n \) and dimension \( p \) going to infinity with fixed ratio \( n/p \to \delta \). We focus on the \( l_1 \)-penalized least absolute deviation and \( l_1 \)-penalized Huber’s regressions. Our analytic study shows the appearance of a sharp phase transition in the two-dimensional sparsity-undersampling phase space. We derive the explicit formula of the phase boundary. Remarkably, the phase boundary is identical to the phase transition curve of LASSO which is also identical to the previously known Donoho–Tanner phase transition for sparse recovery. Our derivation is based on the asymptotic analysis of the generalized approximation passing (GAMP) algorithm. We establish the asymptotic MSE of the \( l_1 \)-penalized robust estimator by connecting it to the asymptotic MSE of the corresponding GAMP estimator. Our results provide some theoretical insight into the high-dimensional regression methods. Extensive computational experiments have been conducted to validate the correctness of our analytic results. We obtain fairly good agreement between theoretical prediction and numerical simulations on finite-size systems.

REFERENCES


Key words and phrases. Mean square error, minimax, penalized, phase transition, robust.


BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE
STATISTICAL
INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE
The Skew-Normal and Related Families
Adelchi Azzalini
in collaboration with Antonella Capitanio

Interest in the skew-normal and related families of distributions has grown enormously over recent years, as theory has advanced, challenges of data have grown, and computational tools have made substantial progress. This comprehensive treatment, blending theory and practice, will be the standard resource for statisticians and applied researchers. Assuming only basic knowledge of (non-measure-theoretic) probability and statistical inference, the book is accessible to the wide range of researchers who use statistical modelling techniques. Guiding readers through the main concepts and results, it covers both the probability and the statistics sides of the subject, in the univariate and multivariate settings. The theoretical development is complemented by numerous illustrations and applications to a range of fields including quantitative finance, medical statistics, environmental risk studies, and industrial and business efficiency.

The author’s freely available R package sn, available from CRAN, equips readers to put the methods into action with their own data.