CONTENTS

BELLEC, P.C. and ZHANG, C.-H.  
De-biasing the lasso with degrees-of-freedom adjustment  713

WANG, Y. and POLITIS, D.N.  
Model-free bootstrap for a general class of stationary time series  744

MAÏDA, M., NGUYEN, T.D., PHAM NGOC, T.M., RIVOIRARD, V. and TRAN, V.C.  
Statistical deconvolution of the free Fokker-Planck equation at fixed time  771

GAÏGALL, D., GERSTENBERG, J. and TRINH, T.T.H.  
Empirical process of concomitants for partly categorial data and applications in statistics  803

JOURDAIN, B. and MARGHERITI, W.  
Martingale Wasserstein inequality for probability measures in the convex order  830

QIN, Q. and JONES, G.L.  
Convergence rates of two-component MCMC samplers  859

ALT, J. and KRÜGER, T.  
Local elliptic law  886

MICHEL, B. and NOUY, A.  
Learning with tree tensor networks: Complexity estimates and model selection  910

LU, J., TAN, Y. and XU, L.  
Central limit theorem and self-normalized Cramér-type moderate deviation for Euler-Maruyama scheme  937

PAÎNDAVEINE, D.  
On the measure of anchored Gaussian simplices, with applications to multivariate medians  965

BING, X., BUNEA, F. and WEGKAMP, M.  
Inference in latent factor regression with clusterable features  997

PADOAN, S.A. and STUPFLER, G.  
Joint inference on extreme expectiles for multivariate heavy-tailed distributions  1021

TANG, M., CAPE, J. and PRIEBE, C.E.  
Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral estimator  1049

CHATTERJI, N.S., BARTLETT, P.L. and LONG, P.M.  
Oracle lower bounds for stochastic gradient sampling algorithms  1074

CHAKRABORTY, M. and GHOSAL, S.  
Rates and coverage for monotone densities using projection-posterior  1093

GOLDENSHLUGER, A. and LEPSKI, O.V.  
Minimax estimation of norms of a probability density: I. Lower bounds  1120

A list of forthcoming papers can be found online at http://www.bernoullisociety.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers
CONTENTS

(continued)

GOLDENSHLAGER, A. and LEPSKI, O.V.
Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures
1155

LAM-WEIL, J., CARPENTIER, A. and SRIPERUMBUDUR, B.K.
Local minimax rates for closeness testing of discrete distributions
1179

ALISHAHI, K. and BARZEGAR, M.
Paving property for real stable polynomials and strongly Rayleigh processes
1198

ZHANG, L., ZHOU, W. and WANG, H.
Non-asymptotic properties of spectral decomposition of large Gram-type matrices and applications
1224

KURISU, D.
Nonparametric regression for locally stationary random fields under stochastic sampling design
1250

BERGER, D. and LINDNER, A.
A Cramér–Wold device for infinite divisibility of $\mathbb{Z}^d$-valued distributions
1276

NAULET, Z.
Adaptive Bayesian density estimation in sup-norm
1284

BEHME, A. and SIDERIS, A.
Markov-modulated generalized Ornstein-Uhlenbeck processes and an application in risk theory
1309

FRANCESCHINI, C., GONÇALVES, P. and SAU, F.
Symmetric inclusion process with slow boundary: Hydrodynamics and hydrostatics
1340

BELOMESTNY, D., IOSIPOI, L., PARIS, Q. and ZHIVOTOVSKIY, N.
Empirical variance minimization with applications in variance reduction and optimal control
1382

KERSTING, G. and MINUESA, C.
Defective Galton-Watson processes in a varying environment
1408

HILÁRIO, M. and UNGARETTI, D.
A note on the phase transition for independent alignment percolation
1432

NAZAROV, A. and NIKITIN, Y.
Spectral equivalence of Gaussian random functions: Operator approach
1448

HART, S. and RINOTT, Y.
Posterior probabilities: Nonmonotonicity, asymptotic rates, log-concavity, and Turán’s inequality
1461

BARNES, C.
Convergence of jump processes with stochastic intensity to Brownian motion with inert drift
1491
BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Aims and Scope
Issued four times per year, BERNOULLI is the flagship journal of the Bernoulli Society for Mathematical Statistics and Probability. The journal aims at publishing original research contributions of the highest quality in all subfields of Mathematical Statistics and Probability. The main emphasis of Bernoulli is on theoretical work, yet discussion of interesting applications in relation to the proposed methodology is also welcome.

Bernoulli Society for Mathematical Statistics and Probability
The Bernoulli Society was founded in 1973. It is an autonomous Association of the International Statistical Institute, ISI. According to its statutes, the object of the Bernoulli Society is the advancement, through international contacts, of the sciences of probability (including the theory of stochastic processes) and mathematical statistics and of their applications to all those aspects of human endeavour which are directed towards the increase of natural knowledge and the welfare of mankind.

Meetings: http://www.bernoulli-society.org/index.php/meetings
The Society holds a World Congress every four years; more frequent meetings, coordinated by the Society’s standing committees and often organised in collaboration with other organisations, are the European Meeting of Statisticians, the Conference on Stochastic Processes and their Applications, the CLAPEM meeting (Latin-American Congress on Probability and Mathematical Statistics), the European Young Statisticians Meeting, and various meetings on special topics – in the physical sciences in particular. The Society, as an association of the ISI, also collaborates with other ISI associations in the organization of the biennial ISI World Statistics Congresses (formerly ISI Sessions).

Executive Committee
Detailed information about the members of the Executive Committee can be found on http://www.bernoulli-society.org/who-is-who

The papers published in Bernoulli are indexed or abstracted in Mathematical Reviews (MathSciNet), Zentralblatt MATH (zbMATH Open), Science Citation Index Expanded (Web of Science), SCOPUS and Google Scholar.

©2022 International Statistical Institute/Bernoulli Society

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the Publisher. In 2022 Bernoulli consists of 4 issues published in February, May, August and November.
De-biasing the lasso with degrees-of-freedom adjustment

PIERRE C. BELLEC* and CUN-HUI ZHANG†

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08854, USA.
E-mail: * pierre.bellec@rutgers.edu; † czhang@stat.rutgers.edu

This paper studies schemes to de-bias the Lasso in sparse linear regression with Gaussian design where the goal is to estimate and construct confidence intervals for a low-dimensional projection of the unknown coefficient vector in a preconceived direction $a_0$. Our analysis reveals that previously analyzed propositions to de-bias the Lasso require a modification in order to enjoy nominal coverage and asymptotic efficiency in a full range of the level of sparsity. This modification takes the form of a degrees-of-freedom adjustment that accounts for the dimension of the model selected by the Lasso. The degrees-of-freedom adjustment (a) preserves the success of de-biasing methodologies in regimes where previous proposals were successful, and (b) repairs the nominal coverage and provides efficiency in regimes where previous proposals produce spurious inferences and provably fail to achieve the nominal coverage. Hence our theoretical and simulation results call for the implementation of this degrees-of-freedom adjustment in de-biasing methodologies.

Let $s_0$ denote the number of nonzero coefficients of the true coefficient vector and $\Sigma$ the population Gram matrix. The unadjusted de-biasing scheme may fail to achieve the nominal coverage as soon as $s_0 \gg n^{2/3}$ if $\Sigma$ is known. If $\Sigma$ is unknown, the degrees-of-freedom adjustment grants efficiency for the contrast in a general direction $a_0$ when

$$\frac{s_0 \log p}{n} + \min \left\{ \frac{s_0 \log p}{n}, \frac{\| \Sigma^{-1} a_0 \|_1 \sqrt{\log p}}{\frac{\| \Sigma^{-1/2} a_0 \|_2 \sqrt{n}}{1}} \right\} + \frac{\min(s_0, s_0 \log p)}{\sqrt{n}} \rightarrow 0,$$

where $s_0 = \| \Sigma^{-1} a_0 \|_0$. The dependence in $s_0$, $s_0 \log p$ and $\| \Sigma^{-1} a_0 \|_1$ is optimal and closes a gap in previous upper and lower bounds. Our construction of the estimated score vector provides a novel methodology to handle dense directions $a_0$.

Beyond the degrees-of-freedom adjustment, our proof techniques yield a sharp $\ell_\infty$ error bound for the Lasso which is of independent interest.

Keywords: Statistical inference; Lasso; semiparametric model; Fisher information; efficiency; confidence interval; p-value; regression; high-dimensional data

References


Model-free bootstrap for a general class of stationary time series

YIREN WANG\textsuperscript{1} and DIMITRIS N. POLITIS\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, University of California, San Diego, La Jolla, CA, USA. E-mail: yiw518@ucsd.edu

\textsuperscript{2}Department of Mathematics and Halicioğlu Data Science Institute, University of California, San Diego, La Jolla, CA, USA. E-mail: dpolitis@ucsd.edu

A model-free bootstrap procedure for a general class of stationary time series is introduced. The theoretical framework is established, showing asymptotic validity of bootstrap confidence intervals for many statistics of interest. In addition, asymptotic validity of one-step ahead bootstrap prediction intervals is also demonstrated. Finite-sample experiments are conducted to empirically confirm the performance of the new method, and to compare with popular methods such as the block bootstrap and the autoregressive (AR)-sieve bootstrap.

Keywords: m-approximation; bootstrap validity; transformation function; confidence intervals

References


We are interested in reconstructing the initial condition of a non-linear partial differential equation (PDE), namely the Fokker-Planck equation, from the observation of a Dyson Brownian motion at a given time $t > 0$. The Fokker-Planck equation describes the evolution of electrostatic repulsive particle systems, and can be seen as the large particle limit of correctly renormalized Dyson Brownian motions. The solution of the Fokker-Planck equation can be written as the free convolution of the initial condition and the semi-circular distribution. We propose a non-parametric estimator for the initial condition obtained by performing the free deconvolution via the subordination functions method. This statistical estimator is original as it involves the resolution of a fixed point equation, and a classical deconvolution by a Cauchy distribution. This is due to the fact that, in free probability, the analogue of the Fourier transform is the R-transform, related to the Cauchy transform. In past literature, there has been a focus on the estimation of the initial conditions of linear PDEs such as the heat equation, but to the best of our knowledge, this is the first time that the problem is tackled for a non-linear PDE. The convergence of the estimator is proved and the integrated mean square error is computed, providing rates of convergence similar to the ones known for non-parametric deconvolution methods. Finally, a simulation study illustrates the good performances of our estimator.

Keywords: 35Q62; 65M32; 62G05; 46L53; 35R30; 60B20; 46L54

References


Empirical process of concomitants for partly categorial data and applications in statistics

DANIEL GAIGALL 1, JULIAN GERSTENBERG 1,2,* and THI THU HA TRINH 1

1 Institute of Probability and Statistics, Leibniz University Hannover, Welfengarten 1, 30167 Hannover, Germany.
2 Institute of Mathematics, Goethe University Frankfurt, Robert-Mayer-Straße 10, 60325 Frankfurt am Main, Germany.
E-mail: * gerstenb@math.uni-frankfurt.de

On the basis of independent and identically distributed bivariate random vectors, where the components are categorial and continuous variables, respectively, the related concomitants, also called induced order statistic, are considered. The main theoretical result is a functional central limit theorem for the empirical process of the concomitants in a triangular array setting. A natural application is hypothesis testing. An independence test and a two-sample test are investigated in detail. The fairly general setting enables limit results under local alternatives and bootstrap samples. For the comparison with existing tests from the literature simulation studies are conducted. The empirical results obtained confirm the theoretical findings.

Keywords: Categorial variable; concomitant; induced order statistic; empirical process; independence test; two-sample test; bootstrap; triangular array; local alternatives

References


Martingale Wasserstein inequality for probability measures in the convex order

BENJAMIN JOURDAIN * and WILLIAM MARGHERITI †

CERMICS, Ecole des Ponts, INRIA, Marne-la-Vallée, France.
E-mail: *benjamin.jourdain@enpc.fr; †william.margheriti@enpc.fr

It was shown by the authors that two one-dimensional probability measures in the convex order admit a martingale coupling with respect to which the integral of \(|x - y|\) is smaller than twice their \(W_1\)-distance (Wasserstein distance with index 1). We showed that replacing \(|x - y|\) and \(W_1\) respectively with \(|x - y|^{\rho}\) and \(W_\rho\) does not lead to a finite multiplicative constant. We show here that a finite constant is recovered when replacing \(W_\rho\) with the product of \(W_\rho\) times the centred \(\rho\)-th moment of the second marginal to the power \(\rho - 1\). Then we study the generalisation of this new martingale Wasserstein inequality to higher dimension.

Keywords: Convex order; martingale optimal transport; Wasserstein distance; martingale couplings

References


Convergence rates of two-component MCMC samplers

QIAN QIN* and GALIN L. JONES†

School of Statistics, University of Minnesota, Minneapolis, USA.
E-mail: * qqin@umn.edu; † galin@umn.edu

Component-wise MCMC algorithms, including Gibbs and conditional Metropolis-Hastings samplers, are commonly used for sampling from multivariate probability distributions. A long-standing question regarding Gibbs algorithms is whether a deterministic-scan (systematic-scan) sampler converges faster than its random-scan counterpart. We answer this question when the samplers involve two components by establishing an exact quantitative relationship between the $L^2$ convergence rates of the two samplers. The relationship shows that the deterministic-scan sampler converges faster. We also establish qualitative relations among the convergence rates of two-component Gibbs samplers and some conditional Metropolis-Hastings variants. For instance, it is shown that if some two-component conditional Metropolis-Hastings samplers are geometrically ergodic, then so are the associated Gibbs samplers.

Keywords: Deterministic-scan; geometric ergodicity; Gibbs; Metropolis-within-Gibbs; random-scan

References


Local elliptic law

JOHANNES ALT$^{1,2}$ and TORBEN KRÜGER$^3$

$^1$Section of Mathematics, University of Geneva, 24, rue du Général Dufour, Case postale 64, 1211 Genève 4, Switzerland.
E-mail: johannes.alt@unige.ch

$^2$Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012-1110, United States of America.
E-mail: johannes.alt@unige.ch

$^3$Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 København, Denmark.
E-mail: tk@math.ku.dk

The empirical eigenvalue distribution of the elliptic random matrix ensemble tends to the uniform measure on an ellipse in the complex plane as its dimension tends to infinity. We show this convergence on all mesoscopic scales slightly above the typical eigenvalue spacing in the bulk spectrum with an optimal convergence rate. As a corollary we obtain complete delocalisation for the corresponding eigenvectors in any basis.

Keywords: Local law; eigenvector delocalisation; elliptic ensemble; matrix Dyson equation

References


Learning with tree tensor networks: Complexity estimates and model selection

BERTRAND MICHEL * and ANTHONY NOUY †

Centrale Nantes, Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, France.
E-mail: * bertrand.michel@ec-nantes.fr; † anthony.nouy@ec-nantes.fr

In this paper, we propose and analyze a model selection method for tree tensor networks in an empirical risk minimization framework and analyze its performance over a wide range of smoothness classes. Tree tensor networks, or tree-based tensor formats, are prominent model classes for the approximation of high-dimensional functions in numerical analysis and data science. They correspond to sum-product neural networks with a sparse connectivity associated with a dimension partition tree $T$, widths given by a tuple $r$ of tensor ranks, and multilinear activation functions (or units). The approximation power of these model classes has been proved to be optimal (or near to optimal) for classical smoothness classes. However, in an empirical risk minimization framework with a limited number of observations, the dimension tree $T$ and ranks $r$ should be selected carefully to balance estimation and approximation errors. In this paper, we propose a complexity-based model selection strategy à la Barron, Birgé, Massart. Given a family of model classes associated with different trees, ranks, tensor product feature spaces and sparsity patterns for sparse tensor networks, a model is selected by minimizing a penalized empirical risk, with a penalty depending on the complexity of the model class. After deriving bounds of the metric entropy of tree tensor networks with bounded parameters, we deduce a form of the penalty from bounds on suprema of empirical processes. This choice of penalty yields a risk bound for the predictor associated with the selected model. In a least-squares setting, after deriving fast rates of convergence of the risk, we show that the proposed strategy is (near to) minimax adaptive to a wide range of smoothness classes including Sobolev or Besov spaces (with isotropic, anisotropic or mixed dominating smoothness) and analytic functions. We discuss the role of sparsity of the tensor network for obtaining optimal performance in several regimes. In practice, the amplitude of the penalty is calibrated with a slope heuristics method. Numerical experiments in a least-squares regression setting illustrate the performance of the strategy for the approximation of multivariate functions and univariate functions identified with tensors by tensorization (quantization).

Keywords: Tensor networks; statistical learning; metric entropy; model selection; minimax adaptive

References


Central limit theorem and self-normalized Cramér-type moderate deviation for Euler-Maruyama scheme

JIANYA LU$^{1,2,*}$, YUZHEN TAN$^3$ and LIHU XU$^{1,2,†}$

$^1$Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, China. E-mail: *jianya.lu@connect.um.edu.mo; †lihuxu@um.edu.mo
$^2$UM Zhuhai Research Institute, Hengqin New District, Zhuhai, China
$^3$Zhongtai Securities Institute for Financial Studies, Shandong University, 250100, Jinan, China. E-mail: tanyuzhensdu@gmail.com

We consider a stochastic differential equation and its Euler-Maruyama (EM) scheme, under some appropriate conditions, they both admit a unique invariant measure, denoted by $\pi$ and $\pi_\eta$ respectively ($\eta$ is the step size of the EM scheme). We construct an empirical measure $\Pi_\eta$ of the EM scheme as a statistic of $\pi_\eta$, and use Stein’s method developed in Fang, Shao and Xu (Probab. Theory Related Fields 174 (2019) 945–979) to prove a central limit theorem of $\Pi_\eta$. The proof of the self-normalized Cramér-type moderate deviation (SNCMD) is based on a standard decomposition on Markov chain, splitting $\eta^{-1/2}(\Pi_\eta(\cdot) - \pi(\cdot))$ into a martingale difference series sum $\mathcal{H}_\eta$ and a negligible remainder $\mathcal{R}_\eta$. We handle $\mathcal{H}_\eta$ by the time-change technique for martingale, while prove that $\mathcal{R}_\eta$ is exponentially negligible by concentration inequalities, which have their independent interest. Moreover, we show that SNCMD holds for $x = o(\eta^{-1/6})$, which has the same order as that of the classical result in Shao (J. Theoret. Probab. 12 (1999) 385–398), Jing, Shao and Wang (Ann. Probab. 31 (2003) 2167–2215).

Keywords: Stochastic differential equation; Euler-Maruyama scheme; central limit theorem; self-normalized Cramér-type moderate deviation; Stein’s method

References


On the measure of anchored Gaussian simplices, with applications to multivariate medians

DAVY PAINDAVEINE\textsuperscript{1,2}

\textsuperscript{1}ECARES and Département de Mathématique, Université libre de Bruxelles, Avenue F.D. Roosevelt, 50, ECARES, CP114/04, B-1050, Brussels, Belgium. E-mail: dpaindav@ulb.ac.be

\textsuperscript{2}Toulouse School of Economics, Université Toulouse Capitole, 1, Esplanade de l’Université, 31080 Toulouse Cedex 06, France

We consider anchored Gaussian $\ell$-simplices in the $d$-dimensional Euclidean space, that is, simplices with one fixed vertex $y \in \mathbb{R}^d$ and the remaining vertices $X_1, \ldots, X_\ell$ randomly sampled from the $d$-variate standard normal distribution. We determine the distribution of the measure of such simplices for any $d$, any $\ell$, and any anchor point $y$, which is of interest, e.g., when studying the asymptotic behaviour of U-statistics based on such simplex measures. We provide two proofs of the results. The first one is short but is not self-contained as it crucially relies on a technical result for non-central Wishart distributions. The second one is a simple and self-contained proof, that also provides some geometric insight on the results. Quite nicely, variations on this second argument reveal intriguing distributional identities on products of central and non-central chi-square distributions with Beta-distributed non-centrality parameters. We independently establish these distributional identities by making use of Mellin transforms. Beyond the aforementioned use to study the asymptotic behaviour of some U-statistics, our results do find natural applications in the context of robust location estimation, as we illustrate by considering a class of simplex-based multivariate medians that contains the celebrated spatial median and Oja median as special cases. Throughout, our results are confirmed by numerical experiments.

Keywords: Distributional identities; Mellin transforms; Multivariate medians; Oja median; Random simplices; Spatial median; Stochastic geometry

References


Inference in latent factor regression with clusterable features

XIN BING\textsuperscript{1,*}, FLORENTINA BUNEA\textsuperscript{1,†} and MARTEN WEGKAMP\textsuperscript{1,2}

\textsuperscript{1}Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA.
E-mail: \textsuperscript{*}xb43@cornell.edu; \textsuperscript{†}fb238@cornell.edu
\textsuperscript{2}Department of Mathematics, Cornell University, Ithaca, New York, USA.
E-mail: mhw73@cornell.edu

Regression models, in which the observed features $X \in \mathbb{R}^p$ and the response $Y \in \mathbb{R}$ depend, jointly, on a lower dimensional, unobserved, latent vector $Z \in \mathbb{R}^K$, with $K \ll p$, are popular in a large array of applications, and mainly used for predicting a response from correlated features. In contrast, methodology and theory for inference on the regression coefficient $\beta \in \mathbb{R}^K$ relating $Y$ to $Z$ are scarce, since typically the un-observable factor $Z$ is hard to interpret. Furthermore, the determination of the asymptotic variance of an estimator of $\beta$ is a long-standing problem, with solutions known only in a few particular cases.

To address some of these outstanding questions, we develop inferential tools for $\beta$ in a class of factor regression models in which the observed features are signed mixtures of the latent factors. The model specifications are both practically desirable, in a large array of applications, render interpretability to the components of $Z$, and are sufficient for parameter identifiability.

Without assuming that the number of latent factors $K$ or the structure of the mixture is known in advance, we construct computationally efficient estimators of $\beta$, along with estimators of other important model parameters. We benchmark the rate of convergence of $\beta$ by first establishing its $\ell_2$-norm minimax lower bound, and show that our proposed estimator $\hat{\beta}$ is minimax-rate adaptive. Our main contribution is the provision of a unified analysis of the component-wise Gaussian asymptotic distribution of $\hat{\beta}$ and, especially, the derivation of a closed form expression of its asymptotic variance, together with consistent variance estimators. The resulting inferential tools can be used when both $K$ and $p$ are independent of the sample size $n$, and also when both, or either, $p$ and $K$ vary with $n$, while allowing for $p > n$. This complements the only asymptotic normality results obtained for a particular case of the model under consideration, in the regime $K = O(1)$ and $p \to \infty$, but without a variance estimate.

As an application, we provide, within our model specifications, a statistical platform for inference in regression on latent cluster centers, thereby increasing the scope of our theoretical results.

We benchmark the newly developed methodology on a recently collected data set for the study of the effectiveness of a new SIV vaccine. Our analysis enables the determination of the top latent antibody-centric mechanisms associated with the vaccine response.

\textbf{Keywords:} High dimensional regression; latent factor model; identification; uniform inference; minimax estimation; pure variables; post clustering inference/regression; adaptive estimation

\textbf{References}


Joint inference on extreme expectiles for multivariate heavy-tailed distributions

SIMONE A. PADOAN and GILLES STUPFLER

1Department of Decision Sciences, Bocconi University, via Roentgen 1, 20136 Milano, Italy. E-mail: simone.padoan@unibocconi.it
2Univ Rennes, Ensai, CNRS, CREST - UMR 9194, Rennes, France. E-mail: gilles.stupfler@ensai.fr

Expectiles induce a law-invariant, coherent and elicitable risk measure that has received substantial attention in actuarial and financial risk management contexts. A number of recent papers have focused on the behaviour and estimation of extreme expectile-based risk measures and their potential for risk assessment was highlighted in financial and actuarial real data applications. Joint inference of several extreme expectiles has however been left untouched; in fact, even the inference about a marginal extreme expectile turns out to be a difficult problem in finite samples, even though an accurate idea of estimation uncertainty is crucial for the construction of confidence intervals in applications to risk management. We investigate the joint estimation of extreme marginal expectiles of a random vector with heavy-tailed marginal distributions, in a general extremal dependence model. We use these results to derive corrected confidence regions for extreme expectiles, as well as a test for the equality of tail expectiles. The methods are showcased in a finite-sample simulation study and on real financial data.

Keywords: Expectiles; extremal dependence; heavy tails; joint convergence; joint inference; tail copula; testing

References

Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral estimator

MINH TANG¹, JOSHUA CAPE² and CAREY E. PRIEBE³

¹Department of Statistics, North Carolina State University, Raleigh, NC, USA.
E-mail: mtang8@ncsu.edu
²Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA.
E-mail: joshua.cape@pitt.edu
³Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA.
E-mail: cep@jhu.edu

We establish asymptotic normality results for estimation of the block probability matrix $B$ in stochastic blockmodel graphs using spectral embedding when the average degrees grows at the rate of $\omega(\sqrt{n})$ in $n$, the number of vertices. As a corollary, we show that when $B$ is of full-rank, estimates of $B$ obtained from spectral embedding are asymptotically efficient. When $B$ is singular the estimates obtained from spectral embedding can have smaller mean square error than those obtained from maximizing the log-likelihood under no rank assumption, and furthermore, can be almost as efficient as the true MLE that assumes the rank of $B$ is known. Our results indicate, in the context of stochastic blockmodel graphs, that spectral embedding is not just computationally tractable, but that the resulting estimates are also admissible, even when compared to the purportedly optimal but computationally intractable maximum likelihood estimation under no rank assumption.

Keywords: Asymptotic efficiency; random dot product graph; stochastic blockmodels; asymptotic normality; spectral embedding

References


Oracle lower bounds for stochastic gradient sampling algorithms

NILADRI S. CHATTERJI¹, PETER L. BARTLETT² and PHILIP M. LONG³

¹Department of Computer Science, Stanford University, 353 Jane Stanford Way, Stanford, CA 94305, USA. E-mail: niladri@cs.stanford.edu
²University of California, Berkeley & Google, 367 Evans Hall #3860, Berkeley, CA 94720-3860, USA. E-mail: peter@berkeley.edu
³Google, 1600 Amphitheatre Parkway, Mountain View, CA, 94043, USA. E-mail: plong@google.com

We consider the problem of sampling from a strongly log-concave density in $\mathbb{R}^d$, and prove an information theoretic lower bound on the number of stochastic gradient queries of the log density needed. Several popular sampling algorithms (including many Markov chain Monte Carlo methods) operate by using stochastic gradients of the log density to generate a sample; our results establish an information theoretic limit for all these algorithms.

We show that for every algorithm, there exists a well-conditioned strongly log-concave target density for which the distribution of points generated by the algorithm would be at least $\varepsilon$ away from the target in total variation distance if the number of gradient queries is less than $\Omega\left(\frac{\sigma^2d}{\varepsilon^2}\right)$, where $\sigma^2d$ is the variance of the stochastic gradient. Our lower bound follows by combining the ideas of Le Cam deficiency routinely used in the comparison of statistical experiments along with standard information theoretic tools used in lower bounding Bayes risk functions. To the best of our knowledge our results provide the first nontrivial dimension-dependent lower bound for this problem.

Keywords: Sampling lower bounds; information theoretic lower bounds; Markov chain Monte Carlo; stochastic gradient Monte Carlo

References

Rates and coverage for monotone densities using projection-posterior

MOUMITA CHAKRABORTY* and SUBHASHIS GHOSAL†

Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695-8203, USA.
E-mail: *mchakra@ncsu.edu; †sghosal@ncsu.edu

We consider Bayesian inference for a monotone density on the unit interval and study the resulting asymptotic properties. We consider a “projection-posterior” approach, where we construct a prior on density functions through random histograms without imposing the monotonicity constraint, but induce a random distribution by projecting a sample from the posterior on the space of monotone functions. The approach allows us to retain posterior conjugacy, allowing explicit expressions extremely useful for studying asymptotic properties. We show that the projection-posterior contracts at the optimal $n^{-1/3}$-rate. We then construct a consistent test based on the posterior distribution for testing the hypothesis of monotonicity. Finally, we obtain the limiting coverage of a projection-posterior credible interval for the value of the function at an interior point. Interestingly, the limiting coverage turns out to be higher than the nominal credibility level, the opposite of the undercoverage phenomenon observed in a smoothness regime. Moreover, we show that a recalibration method using a lower credibility level gives an intended limiting coverage. We also discuss extensions of the obtained results for densities on the half-line. We conduct a simulation study to demonstrate the accuracy of the asymptotic results in finite samples.

Keywords: Monotone density; contraction rate; Bayesian test for monotonicity; credible interval; coverage

References


Minimax estimation of norms of a probability density: I. Lower bounds

ALEXANDER GOLDENSHLUGER 1 and OLEG V. LEPSKI 2

1 Department of Statistics, University of Haifa, Haifa, Israel. E-mail: goldensh@stat.haifa.ac.il
2 Institut de Mathématique de Marseille, Aix-Marseille Université, Marseille, France. E-mail: oleg.lepski@univ-amu.fr

The paper deals with the problem of nonparametric estimating the $L_p$–norm, $p \in (1, \infty)$, of a probability density on $\mathbb{R}^d$, $d \geq 1$ from independent observations. The unknown density is assumed to belong to a ball in the anisotropic Nikolskii’s space. We adopt the minimax approach, and derive lower bounds on the minimax risk. In particular, we demonstrate that accuracy of estimation procedures essentially depends on whether $p$ is integer or not. Moreover, we develop a general technique for derivation of lower bounds on the minimax risk in the problems of estimating nonlinear functionals. The proposed technique is applicable for a broad class of nonlinear functionals, and it is used for derivation of the lower bounds in the $L_p$–norm estimation.

Keywords: Estimation of nonlinear functionals; minimax estimation; minimax risk; anisotropic Nikolskii’s class; best approximation

References

In this paper we develop rate–optimal estimation procedures in the problem of estimating the $L_p$–norm, $p \in (1, \infty)$ of a probability density from independent observations. The density is assumed to be defined on $\mathbb{R}^d$, $d \geq 1$ and to belong to a ball in the anisotropic Nikol’skii space. We adopt the minimax approach and construct rate–optimal estimators in the case of integer $p \geq 2$. We demonstrate that, depending on the parameters of the Nikol’skii class and the norm index $p$, the minimax rates of convergence may vary from inconsistency to the parametric $\sqrt{n}$–estimation. The results in this paper complement the minimax lower bounds derived in the companion paper (Goldenshluger and Lepski (2020)).

**Keywords:** Density estimation; minimax risk; $L_p$-norm; $U$-statistics; anisotropic Nikol’skii class

**References**


Local minimax rates for closeness testing of discrete distributions

JOSEPH LAM-WEIL$^1$,* , ALEXANDRA CARPENTIER$^1,†$ and BHARATH K. SRIPERUMBUDUR$^2$

$^1$Magdeburg University, Universitaetsplatz 2, 39106 Magdeburg, Germany. E-mail: *joseph.lam@ovgu.de; †alexandra.carpentier@ovgu.de
$^2$Pennsylvania State University, 201 Old Main, Pennsylvania 16802, USA. E-mail: bks18@psu.edu

We consider the closeness testing problem for discrete distributions. The goal is to distinguish whether two samples are drawn from the same unspecified distribution, or whether their respective distributions are separated in $L_1$-norm. In this paper, we focus on adapting the rate to the shape of the underlying distributions, i.e. we consider a local minimax setting. We provide, to the best of our knowledge, the first local minimax rate for the separation distance up to logarithmic factors, together with a test that achieves it. In view of the rate, closeness testing turns out to be substantially harder than the related one-sample testing problem over a wide range of cases.

Keywords: Local minimax optimality; closeness testing; two-sample; instance optimal; discrete distributions; hypothesis testing; composite-composite testing

References

Paving property for real stable polynomials and strongly Rayleigh processes

KASRA ALISHAHI* and MILAD BARZEGAR†

Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran.
E-mail: *alishahi@sharif.edu; †milad.barzegar@sharif.edu

One of the equivalent formulations of the Kadison–Singer problem which was resolved in 2013 by Marcus, Spielman and Srivastava, is the “paving conjecture”. In this paper, we first extend this result to real stable polynomials. We prove that for every multi-affine real stable polynomial satisfying a simple condition, it is possible to partition its set of variables to a small number of subsets such that the “restriction” of the polynomial to each subset has small roots. Then, we derive a probabilistic interpretation of this result. We show that there exists a partition of the underlying space of every strongly Rayleigh process into a small number of sets such that the restriction of the point process to each set has “almost independent” points. This result implies that the dependence structure of strongly Rayleigh processes is constrained—a phenomenon that is to be expected in negatively dependent measures. To prove this result, we introduce and study the notion of kernel polynomial for strongly Rayleigh processes. This notion is a natural generalization of the kernel of determinantal processes. We also derive an entropy bound for strongly Rayleigh processes in terms of the roots of the kernel polynomial which is interesting on its own.

Keywords: Paving property; real stable polynomials; strongly Rayleigh processes

References


Non-asymptotic properties of spectral decomposition of large Gram-type matrices and applications

LYOU ZHANG¹, WEN ZHOU²,*, and HAONAN WANG²,†

¹School of Statistics and Management, Shanghai University of Finance of Economics, 777 Guoding Road, Shanghai, 200433, P.R. China. E-mail: zhanglyou@mail.shufe.edu.cn

²Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA. E-mail: *riczw@stat.colostate.edu; †wanghn@stat.colostate.edu

Gram-type matrices and their spectral decomposition are of central importance for numerous problems in statistics, applied mathematics, physics, and machine learning. In this paper, we carefully study the non-asymptotic properties of spectral decomposition of large Gram-type matrices when data are not necessarily independent. Specifically, we derive the exponential tail bounds for the deviation between eigenvectors of the right Gram matrix to their population counterparts as well as the Berry-Esseen type bound for these deviations. We also obtain the non-asymptotic tail bound of the ratio between eigenvalues of the left Gram matrix, namely the sample covariance matrix, and their population counterparts regardless of the size of the data matrix. The documented non-asymptotic properties are further demonstrated in a suite of applications, including the non-asymptotic characterization of the estimated number of latent factors in factor models and relate machine learning problems, the estimation and forecasting of high-dimensional time series, the spectral properties of large sample covariance matrix such as perturbation bounds and inference on the spectral projectors, and low-rank matrix denoising using dependent data.

Keywords: Approximate factor model; Gram-type matrices; high-dimensional time series; non-asymptotic analysis; principal component analysis; spectral decomposition

References


Nonparametric regression for locally stationary random fields under stochastic sampling design

DAISUKE KURISU

Department of Industrial Engineering and Economics, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan.
E-mail: kurisu.d.aa@m.titech.ac.jp

In this study, we develop an asymptotic theory of nonparametric regression for locally stationary random fields (LSRFs) \( \{ X_s, A_n \} : s \in \mathbb{R}^p \) observed at irregularly spaced locations in \( \mathbb{R}^d \). We first derive the uniform convergence rate of general kernel estimators, followed by the asymptotic normality of an estimator for the mean function of the model. Moreover, we consider additive models to avoid the curse of dimensionality arising from the dependence of the convergence rate of estimators on the number of covariates. Subsequently, we derive the uniform convergence rate and joint asymptotic normality of the estimators for additive functions. We also introduce approximately \( m_n \)-dependent RFs to provide examples of LSRFs. We find that these RFs include a wide class of Lévy-driven moving average RFs.

Keywords: Nonparametric regression; locally stationary random field; irregularly spaced data; additive model; Lévy-driven moving average random field

References


We show that a Cramér–Wold device holds for infinite divisibility of $\mathbb{Z}^d$-valued distributions, i.e. that the distribution of a $\mathbb{Z}^d$-valued random vector $X$ is infinitely divisible if and only if the distribution of $a^TX$ is infinitely divisible for all $a \in \mathbb{R}^d$, and that this in turn is equivalent to infinite divisibility of the distribution of $a^TX$ for all $a \in \mathbb{N}_0^d$. A key tool for proving this is a Lévy–Khintchine type representation with a signed Lévy measure for the characteristic function of a $\mathbb{Z}^d$-valued distribution, provided the characteristic function is zero-free.

Keywords: Cramér–Wold device; infinitely divisible distribution; quasi-infinitely divisible distribution; signed Lévy measure

References


Adaptive Bayesian density estimation in sup-norm

ZACHARIE NAULET

Laboratoire de mathématiques d'Orsay, CNRS, Université Paris-Saclay, 91405, Orsay, France.
E-mail: zacharie.naulet@universite-paris-saclay.fr

We investigate the problem of deriving adaptive posterior rates of contraction on $L_\infty$ balls in density estimation. Although it is known that log-density priors can achieve optimal rates when the true density is sufficiently smooth, adaptive rates were still to be proven. Here we establish that the so-called spike-and-slab prior can achieve adaptive and optimal posterior contraction rates. Along the way, we prove a generic $L_\infty$ contraction result for log-density priors with independent wavelet coefficients. Interestingly, our approach is different from previous works on $L_\infty$ contraction and is reminiscent of the classical test-based approach used in Bayesian nonparametrics. Moreover, we require no lower bound on the smoothness of the true density, albeit the rates are deteriorated by an extra $\log(n)$ factor in the case of low smoothness.

Keywords: Bayesian density estimation; supremum norm; adaptation

References


Markov-modulated generalized Ornstein-Uhlenbeck processes and an application in risk theory

ANITA BEHME* and APOSTOLOS SIDERIS†

Technische Universität Dresden, Institute of Mathematical Stochastics, 01062 Dresden, Germany.
E-mail: *anita.behme@tu-dresden.de; †apostolos.sideris@tu-dresden.de

By embedding a Markov-modulated random recurrence equation in continuous time, we derive the Markov-modulated generalized Ornstein-Uhlenbeck process. This process turns out to be the unique solution of a stochastic differential equation driven by a bivariate Markov-additive process. We present this stochastic differential equation as well as its solution explicitly in terms of the driving Markov-additive process. Moreover, we give necessary and sufficient conditions for strict stationarity of the Markov-modulated generalized Ornstein-Uhlenbeck process, and prove that its stationary distribution is given by the distribution of a certain exponential functional of Markov-additive processes. Finally, we propose a Markov-modulated risk model with investment that generalizes Paulsen’s risk process to a Markov-switching environment, and derive a formula for the ruin probability in this risk model.

Keywords: Exponential functional; generalized Ornstein-Uhlenbeck process; Lévy process; Markov additive process; Markov-modulated random recurrence equation; Markov-switching model; risk theory; ruin probability; stationary process

References

Symmetric inclusion process with slow boundary: Hydrodynamics and hydrostatics

CHIARA FRANCESCHINI 1,*, PATRÍCIA GONÇALVES 1,† and FEDERICO SAU 2

1 Center for Mathematical Analysis, Geometry and Dynamical Systems, IST, Universidade de Lisboa, 1049-001 Lisboa, Portugal.
E-mail: *chiara.franceschini@tecnico.ulisboa.pt; †pgoncalves@tecnico.ulisboa.pt
2 IST Austria, Am Campus 1, 3400, Klosterneuburg, Austria.
E-mail: federico.sau@ist.ac.at

We study the hydrodynamic and hydrostatic limits of the one-dimensional open symmetric inclusion process with slow boundary. Depending on the value of the parameter tuning the interaction rate of the bulk of the system with the boundary, we obtain a linear heat equation with either Dirichlet, Robin or Neumann boundary conditions as hydrodynamic equation. In our approach, we combine duality and first-second class particle techniques to reduce the scaling limit of the inclusion process to the limiting behavior of a single, non-interacting, particle.

Keywords: Symmetric inclusion process; slow boundary; hydrodynamic limit; hydrostatic limit; non-equilibrium steady state; duality for Markov processes; second class particles

References


[41] Mitoma, I. (1983). Tightness of probabilities on $C([0, 1]; \mathbb{S}'_1)$ and $D([0, 1]; \mathbb{S}')$. *Ann. Probab.* **11** 989–999. MR0714961


Empirical variance minimization with applications in variance reduction and optimal control

DENIS BELOMESTNY\textsuperscript{1,2,*}, LEONID IOSIPOI\textsuperscript{2,†}, QUENTIN PARIS\textsuperscript{2,‡} and NIKITA ZHIVOTOVSKIY\textsuperscript{3}

\textsuperscript{1}Faculty of Mathematics, Duisburg-Essen University, Essen, Germany. E-mail: *denis.belomestny@uni-due.de
\textsuperscript{2}Faculty of Computer Science, HSE University, Moscow, Russia. E-mail: †liosipoi@hse.ru; ‡qparis@hse.ru
\textsuperscript{3}Department of Mathematics, ETH, Zürich, Switzerland. E-mail: nikita.zhivotovskii@math.ethz.ch

We study the problem of empirical minimization for variance-type functionals over functional classes. Sharp non-asymptotic bounds for the excess variance are derived under mild conditions. In particular, it is shown that under some restrictions imposed on the functional class fast convergence rates can be achieved including the optimal non-parametric rates for expressive classes in the non-Donsker regime under some additional assumptions. Our main applications include variance reduction and optimal control.

Keywords: Empirical variance minimization; variance reduction; control variates; optimal control

References

Defective Galton-Watson processes in a varying environment

GÖTZ KERSTING and CARMEN MINUESA

1 Institut für Mathematik, Goethe Universität, Robert-Mayer-Strasse, 6-10, 60325 Frankfurt am Main, Germany. E-mail: kersting@math.uni-frankfurt.de
2 Department of Mathematics, Autonomous University of Madrid, Ciudad Universitaria de Cantoblanco, s/n, 28049 Madrid, Spain. E-mail: carmen.minuesa@uam.es, 0000-0002-8858-3145
3 Department of Mathematics, University of Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain. E-mail: cminuesaa@unex.es

We study an extension of the so-called defective Galton-Watson processes obtained by allowing the offspring distribution to change over the generations. Thus, in these processes, the individuals reproduce independently of the others and in accordance to some possibly defective offspring distribution depending on the generation. Moreover, the defect $1 - f_n(1)$ of the offspring distribution at generation $n$ represents the probability that the process hits an absorbing state $\Delta$ at that generation. We focus on the asymptotic behaviour of these processes. We establish the almost sure convergence of the process to a random variable with values in $\mathbb{N}_0 \cup \{\Delta\}$ and we provide two characterisations of the duality extinction-absorption at $\Delta$. We also state some results on the absorption time and the properties of the process conditioned upon its non-absorption, some of which require us to introduce the notion of defective branching trees in varying environment.

Keywords: Branching process; varying environment; defective distribution; absorption; family tree

References


A note on the phase transition for independent alignment percolation

MARCELO HILÁRIO 1 and DANIEL UNGARETTI 1,2,*

1 ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.
E-mail: mhilario@mat.ufmg.com
2 IME, Universidade de São Paulo, São Paulo, Brazil.
E-mail: *danielungaretti@gmail.com

We study the independent alignment percolation model on \( \mathbb{Z}^d \) introduced by Beaton, Grimmett and Holmes. It is a model for random intersecting line segments defined as follows. First the sites of \( \mathbb{Z}^d \) are independently declared occupied with probability \( p \) and vacant otherwise. Conditional on the configuration of occupied vertices, consider the set of all line segments that are parallel to the coordinate axis, whose extremes are occupied vertices and that do not traverse any other occupied vertex. Declare independently the segments on this set open with probability \( \lambda \) and closed otherwise. All the edges that lie on open segments are also declared open giving rise to a bond percolation model in \( \mathbb{Z}^d \). We show that for any \( d \geq 2 \) and \( p \in (0, 1] \) the critical value for \( \lambda \) satisfies \( \lambda_c(p) < 1 \) completing the proof that the phase transition is non-trivial over the whole interval \( (0, 1] \). We also show that the critical curve \( p \mapsto \lambda_c(p) \) is continuous at \( p = 1 \).

Keywords: Percolation; renormalization; phase transition

References

Spectral equivalence of Gaussian random functions: Operator approach

ALEXANDER NAZAROV\textsuperscript{1,2} and YAKOV NIKITIN\textsuperscript{2,3,†}

\textsuperscript{1}St.Petersburg Department of Steklov Mathematical Institute of Russian Academy of Science, St.Petersburg, Russia.
E-mail: al.il.nazarov@gmail.com

\textsuperscript{2}Mathematics and Mechanics Faculty, St.Petersburg State University, St.Petersburg, Russia

\textsuperscript{3}National Research University — Higher School of Economics, St.Petersburg, Russia.
E-mail: y.nikitin@spbu.ru

We introduce a new approach to the spectral equivalence of Gaussian processes and fields, based on the methods of operator theory in Hilbert space. Besides several new results including identities in law of quadratic norms for integrated and multiply integrated Gaussian random functions we give an application to goodness-of-fit testing.

\textit{Keywords:} Gaussian random functions; identity in law; spectral equivalence; tensor product; Brownian sheet

References


In the standard Bayesian framework data are assumed to be generated by a distribution parametrized by $\theta$ in a parameter space $\Theta$, over which a prior distribution $\pi$ is given. A Bayesian statistician quantifies the belief that the true parameter is $\theta_0$ in $\Theta$ by its posterior probability given the observed data. We investigate the behavior of the posterior belief in $\theta_0$ when the data are generated under some parameter $\theta_1$, which may or may not be the same as $\theta_0$. Starting from stochastic orders, specifically, likelihood ratio dominance, that obtain for resulting distributions of posteriors, we consider monotonicity properties of the posterior probabilities as a function of the sample size when data arrive sequentially. While the $\theta_0$-posterior is monotonically increasing (i.e., it is a submartingale) when the data are generated under that same $\theta_0$, it need not be monotonically decreasing in general, not even in terms of its overall expectation, when the data are generated under a different $\theta_1$. In fact, it may keep going up and down many times, even in simple cases such as iid coin tosses. We obtain precise asymptotic rates when the data come from the wide class of exponential families of distributions; these rates imply in particular that the expectation of the $\theta_0$-posterior under $\theta_1 \neq \theta_0$ is eventually strictly decreasing. Finally, we show that in a number of interesting cases this expectation is a log-concave function of the sample size, and thus unimodal. In the Bernoulli case we obtain this result by developing an inequality that is related to Turán’s inequality for Legendre polynomials.

Keywords: Bayesian analysis; stochastic and likelihood ratio orders; sequential observations; expected posteriors; unimodality; Legendre polynomials; exponential families

References


Convergence of jump processes with stochastic intensity to Brownian motion with inert drift

CLAYTON BARNES

Technion-Israel’s Institute of Technology.
E-mail: cbarnes@campus.technion.ac.il

Consider a random walker on the nonnegative lattice, moving in continuous time, whose positive transition intensity is proportional to the time the walker spends at the origin. In this way, the walker is a jump process with a stochastic and adapted jump intensity. We show that, upon Brownian scaling, the sequence of such processes converges to Brownian motion with inert drift (BMID). BMID was introduced by Frank Knight in 2001 and generalized by White in 2007. This confirms a conjecture of Burdzy and White in 2008 in the one-dimensional setting.

Keywords: Brownian motion; discrete approximation; random walk; local time

References