BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Seven Number Two May 2021 ISSN: 1350-7265

CONTENTS
CARPENTIER, A. and VERZELEN, N. 727
Optimal sparsity testing in linear regression model

GENTIL, I. and ZUGMEYER, S. 751
A family of Beckner inequalities under various curvature-dimension conditions

BRUNEL, V.-E., KLUSOWSKI, J.M. and YANG, D. 772
Estimation of convex supports from noisy measurements

NA, S. and KOLAR, M. 794
High-dimensional index volatility models via Stein’s identity

DÜMBGEN, L., SAMWORTH, R.J. and WELLNER, J.A. 818
Bounding distributional errors via density ratios

GRANDITS, P. 853
Asymptotics of the hitting probability for a small sphere and a two dimensional Brownian motion with discontinuous anisotropic drift

WANG, C., YANG, S. and ZHANG, T. 866
Reflected Brownian motion with singular drift

BRIAND, P., GEISS, C., GEISS, S. and LABART, C. 899
Donsker-type theorem for BSDEs: Rate of convergence

ZHAO, P., WANG, L. and SHAO, J. 930
Sufficient dimension reduction and instrument search for data with nonignorable nonresponse

QIAO, W. 946
Asymptotic confidence regions for density ridges

DJEHICHE, B., MAZHAR, O. and ROJAS, C.R. 976
Finite impulse response models: A non-asymptotic analysis of the least squares estimator

KUPPER, M. and ZAPATA, J.M. 1001
Large deviations built on max-stability

JAKUBOWSKI, A., RODIONOV, I. and SOJA-KUKIELA, N. 1028
Directional phantom distribution functions for stationary random fields

LI, Y. and XU, Y. 1057
On fluctuations of global and mesoscopic linear statistics of generalized Wigner matrices

BOURGUIN, S., DIEZ, C.-P. and TUDOR, C.A. 1077
Limiting behavior of large correlated Wishart matrices with chaotic entries

(continued)

The papers published in Bernoulli are indexed or abstracted in Current Index to Statistics, Mathematical Reviews, Statistical Theory and Method Abstracts-Zentralblatt (STMA-Z), and Zentralblatt für Mathematik (also avalaible on the MATH via STN database and Compact MATH CD-ROM). A list of forthcoming papers can be found online at http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers
CONTENTS

(continued)

CHAKRABORTY, A. and PANARETOS, V.M. 1103
Functional registration and local variations: Identifiability, rank, and tuning

RÖCKNER, M. and ZHANG, X. 1131
Well-posedness of distribution dependent SDEs with singular drifts

BOGERD, K., CASTRO, R.M., VAN DER HOFSTAD, R. and VERZELEN, N. 1159
Detecting a planted community in an inhomogeneous random graph

ZHU, C. and SHAO, X. 1189
Interpoint distance based two sample tests in high dimension

YANO, K., KANEKO, R. and KOMAKI, F. 1212
Minimax predictive density for sparse count data

DREES, H. and NEBLUNG, S. 1239
Asymptotics for sliding blocks estimators of rare events

FAN, A. and KARAGULYAN, D. 1270
On $\mu$-Dvoretzky random covering of the circle

FOUCART, C., LI, P.-S. and ZHOU, X. 1291
Time-changed spectrally positive Lévy processes started from infinity

MIKOSCH, T. and RODIONOV, I. 1319
Precise large deviations for dependent subexponential variables

ARMSTRONG, T.B. 1348
Adaptation bounds for confidence bands under self-similarity

BASRAK, B. and PLANINIĆ, H. 1371
Compound Poisson approximation for regularly varying fields with application to sequence alignment

JENTSCH, C. and KULIK, R. 1409
Bootstrapping Hill estimator and tail array sums for regularly varying time series
Aims and Scope
BERNOULLI is the journal of the Bernoulli Society for Mathematical Statistics and Probability, issued four times per year. The journal provides a comprehensive account of important developments in the fields of statistics and probability, offering an international forum for both theoretical and applied work.

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
The Society is headed by an Executive Committee. As of February 2020 the Executive Committee consists of: President: Claudia Klüppelberg (Germany); President Elect: Adam Jakubowski (Poland); Past President: Susan Murphy (USA); Treasurer: Geoffrey Grimmett (UK); Scientific Secretary: Song Xi Chen (China); Membership Secretary: Sebastian Engelke (Switzerland); Publicity Secretary: Leonardo Rolla (Argentina); Publication Secretary: Herold Dehling (Germany); ISI Director: Ada van Krimpen (Netherlands). Further, the Society has a twelve member Council and a number of standing committees to carry out the tasks outlined above. Final authority is the general assembly of members of the Society, meeting at least biennially at the ISI World Statistics Congresses.

The papers published in Bernoulli are indexed or abstracted in Current Index to Statistics, Mathematical Reviews, Statistical Theory and Method Abstracts-Zentralblatt (STMA-Z), Thomson Scientific and Zentralblatt für Mathematik (also available on the MATH via STN database and Compact MATH CD-ROM). A list of forthcoming papers can be found online at http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

©2021 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 2021 Bernoulli consists of 4 issues published in February, May, August and November.
Optimal sparsity testing in linear regression model

ALEXANDRA CARPENTIER¹ and NICOLAS VERZELEN²

¹Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik (FMA), Institut für Mathematische Stochastik (IMST), Universitätsplatz 2, 39106 Magdeburg, Germany.
E-mail: alexandra.carpentier@ovgu.de
²MISTEA, Université Montpellier, INRAE, Institut Agro, 34060 Montpellier cedex 1, France.
E-mail: nicolas.verzelen@inrae.fr

We consider the problem of sparsity testing in the high-dimensional linear regression model. The problem is to test whether the number of non-zero components (aka the sparsity) of the regression parameter \( \theta^* \) is less than or equal to \( k_0 \). We pinpoint the minimax separation distances for this problem, which amounts to quantifying how far a \( k_1 \)-sparse vector \( \theta^* \) has to be from the set of \( k_0 \)-sparse vectors so that a test is able to reject the null hypothesis with high probability. Two scenarios are considered. In the independent scenario, the covariates are i.i.d. normally distributed and the noise level is known. In the general scenario, both the covariance matrix of the covariates and the noise level are unknown. Although the minimax separation distances differ in these two scenarios, both of them actually depend on \( k_0 \) and \( k_1 \) illustrating that for this composite-composite testing problem both the size of the null and of the alternative hypotheses play a key role.

Keywords: High dimensional regression; minimax composite-composite testing; model testing

References


A family of Beckner inequalities under various curvature-dimension conditions

IVAN GENTIL * and SIMON ZUGMEYER †

Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France.
E-mail: * gentil@math.univ-lyon1.fr; † zugmeyer@math.univ-lyon1.fr

In this paper, we offer a proof for a family of functional inequalities interpolating between the Poincaré and the logarithmic Sobolev (standard and weighted) inequalities. The proofs rely both on entropy flows and on a CD(\(\rho, n\)) condition, either with \(\rho = 0\) and \(n > 0\), or with \(\rho > 0\) and \(n \in \mathbb{R}\). As such, results are valid in the case of a Riemannian manifold, which constitutes a generalization of what was previously proved.

Keywords: Beckner inequalities; curvature-dimension condition; entropy flows; Poincaré inequality

References


Estimation of convex supports from noisy measurements

VICTOR-EMMANUEL BRUNEL\textsuperscript{1}, JASON M. KLUSOWSKI\textsuperscript{2} and DANA YANG\textsuperscript{3}

\textsuperscript{1}Department of Statistics, CREST UMR 9194, ENSAE, 5 avenue Henry Le Chatelier, TSA 96642, 91764 Palaiseau cedex, France.
E-mail: victor.emmanuel.brunel@ensae.fr
\textsuperscript{2}Department of Statistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA.
E-mail: jason.klusowski@rutgers.edu
\textsuperscript{3}The Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, USA.
E-mail: xiaoqian.yang@duke.edu

A popular class of problems in statistics deals with estimating the support of a density from \(n\) observations drawn at random from a \(d\)-dimensional distribution. In the one-dimensional case, if the support is an interval, the problem reduces to estimating its end points. In practice, an experimenter may only have access to a noisy version of the original data. Therefore, a more realistic model allows for the observations to be contaminated with additive noise.

In this paper, we consider estimation of convex bodies when the additive noise is distributed according to a multivariate Gaussian (or nearly Gaussian) distribution, even though our techniques could easily be adapted to other noise distributions. Unlike standard methods in deconvolution that are implemented by thresholding a kernel density estimate, our method avoids tuning parameters and Fourier transforms altogether. We show that our estimator, computable in \(O((\log n)^{d-1}/2)\) time, converges at a rate of \(O_d((\log \log n/\sqrt{\log n})\) in Hausdorff distance, in accordance with the polylogarithmic rates encountered in Gaussian deconvolution problems. Part of our analysis also involves the optimality of the proposed estimator. We provide a lower bound for the minimax rate of estimation in Hausdorff distance that is \(\Omega_1(d(1/\log^2 n))\).

\textbf{Keywords:} Convex bodies; order statistics; support estimation; support function

\textbf{References}


High-dimensional index volatility models via Stein’s identity

SEN NA\textsuperscript{1} and MLADEN KOLAR\textsuperscript{2}

\textsuperscript{1}Department of Statistics, University of Chicago, Chicago, U.S.A.
E-mail: senna@uchicago.edu
\textsuperscript{2}Booth School of Business, University of Chicago, Chicago, U.S.A.
E-mail: mkolar@chicagobooth.edu

We study the estimation of the parametric components of single and multiple index volatility models. Using the first- and second-order Stein’s identities, we develop methods that are applicable for the estimation of the variance index in the high-dimensional setting requiring finite moment condition, which allows for heavy-tailed data. Our approach complements the existing literature in the low-dimensional setting, while relaxing the conditions on estimation, and provides a novel approach in the high-dimensional setting. We prove that the statistical rate of convergence of our variance index estimators consists of a parametric rate and a nonparametric rate, where the latter appears from the estimation of the mean link function. However, under standard assumptions, the parametric rate dominates the rate of convergence and our results match the minimax optimal rate for the mean index estimation. Simulation results illustrate finite sample properties of our methodology and back our theoretical conclusions.

Keywords: High-dimensional estimation; index volatility model; Stein’s identity

References


Bounding distributional errors via density ratios

LUTZ DÜMBGEN¹, RICHARD J. SAMWORTH² and JON A. WELLNER³

¹Department of Mathematics and Statistics, University of Bern, Bern, Switzerland.
E-mail: lutz.duembgen@stat.unibe.ch

²Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, UK.
E-mail: r.samworth@statslab.cam.ac.uk

³Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, USA.
E-mail: jaw@stat.washington.edu

We present some new and explicit error bounds for the approximation of distributions. The approximation error is quantified by the maximal density ratio of the distribution \( Q \) to be approximated and its proxy \( P \). This non-symmetric measure is more informative than and implies bounds for the total variation distance.

Explicit approximation problems include, among others, hypergeometric by binomial distributions, binomial by Poisson distributions, and beta by gamma distributions. In many cases, we provide both upper and (matching) lower bounds.

Keywords: Binomial distribution; hypergeometric distribution; Poisson approximation; relative errors; total variation distance

References


Asymptotics of the hitting probability for a small sphere and a two dimensional Brownian motion with discontinuous anisotropic drift

PETER GRANDITS

Institut für Stochastik und Wirtschaftsmathematik, TU Wien, Wiedner Hauptstraße 8-10, A-1040 Wien, Austria.
E-mail: pgrand@fam.tuwien.ac.at

We provide an approximation of the hitting probability for a small sphere for the following two dimensional process: In $x$-direction it is just a Brownian motion with positive constant drift, whereas in $y$-direction the process $Y_t$ is a Brownian motion with drift given by a negative constant times the sign of $Y_t$. This process can be seen as the solution of a certain stochastic optimal control problem. It turns out that the approximating function can be expressed as the sum of a term involving a modified Bessel function and an ordinary Lebesgue integral.

Keywords: Discontinuous drift; hitting probabilities; optimal control problem; two-dimensional Brownian motion

References


Reflected Brownian motion with singular drift

CHEN WANG\textsuperscript{1,*}, SAISAI YANG\textsuperscript{2} and TUSHENG ZHANG\textsuperscript{1,†}

\textsuperscript{1}School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, England, U.K.
E-mail: *chen.wang-9@postgrad.manchester.ac.uk; †tusheng.zhang@manchester.ac.uk
\textsuperscript{2}School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, China.
E-mail: yangss@ustc.edu.cn

In this article, we show that there exists a unique weak solution to the reflected Brownian motion with singular drift $\mu$, where $\mu$ is a vector-valued Kato class measure on $\mathbb{R}^d$. Furthermore, we obtain some Gaussian type estimates of the transition density function of the solution.

Keywords: Stochastic differential equation; reflected Brownian motion; local time; singular drift; weak solution; transition density function

References


Donsker-type theorem for BSDEs: 
Rate of convergence

PHILIPPE BRIAND1,*, CHRISTEL GEISS2,‡, STEFAN GEISS2,§ and CÉLINE LABART1,†

1 Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA, 73000 Chambéry, France.
E-mail: * philippe.briand@univ-smb.fr; † celine.labart@univ-smb.fr
2 Department of Mathematics and Statistics, P.O.Box 35 (MaD), FI-40014 University of Jyväskyla, Finland.
E-mail: ‡ christel.geiss@jyu.fi; § stefan.geiss@jyu.fi

In this paper, we study in the Markovian case the rate of convergence in Wasserstein distance when the solution to a BSDE is approximated by a solution to a BSDE driven by a scaled random walk as introduced in Briand, Delyon and Mémin (Electron. Commun. Probab. 6 (2001) Art. ID 1). This is related to the approximation of solutions to semilinear second order parabolic PDEs by solutions to their associated finite difference schemes and the speed of convergence.

Keywords: Backward stochastic differential equations; convergence rate; Donsker’s theorem; finite difference scheme; scaled random walk; Wasserstein distance

References


Consider a response variable subject to nonignorable nonresponse and a fully observed covariate vector. The purpose of our study is threefold. First, we study how to extend nonparametric sufficient dimension reduction to data with nonignorable nonresponse. Second, we utilize sufficient dimension reduction to search an instrument, a linear function of covariates that is related to the response variable but can be excluded from the propensity of nonignorable nonresponse, for the purpose of identifying unknown parameters in a semiparametric propensity and a nonparametric distribution of response variable and covariates. Third, we establish asymptotic results for parameter estimators based on sufficient dimension reduction and instrument search, and investigate the effect on the limiting distribution of parameter estimators due to instrument search. We evaluate the performance of proposed estimators in a Monte Carlo study and illustrate our method in an application to AIDS Clinical Trials Group Protocol 175 data.

Keywords: Covariate dimension reduction; estimation; identifiability; instrument; nonparametric kernel regression; semiparametric propensity

References


Asymptotic confidence regions for density ridges

WANLI QIAO

Department of Statistics, George Mason University, Fairfax, VA 22030, USA.
E-mail: wqiao@gmu.edu

We develop large sample theory including nonparametric confidence regions for $r$-dimensional ridges of probability density functions on $\mathbb{R}^d$, where $1 \leq r < d$. We view ridges as the intersections of level sets of some special functions. The vertical variation of the plug-in kernel estimators for these functions constrained on the ridges is used as the measure of maximal deviation for ridge estimation. Our confidence regions for the ridges are determined by the asymptotic distribution of this maximal deviation, which is established by utilizing the extreme value distribution of nonstationary $\chi$-fields indexed by manifolds.

Keywords: Ridges; intersections; level sets; extreme value distribution; kernel density estimation

References


Finite impulse response models:
A non-asymptotic analysis of the least squares estimator

BOUALEM DJEHICHE1,*, OTHMANE MAZHAR1,† and CRISTIAN R. ROJAS2

1Department of Mathematics, School of Engineering Sciences, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden.
E-mail: * boualem@kth.se; † othmane@kth.se
2Division of Decision And Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.
E-mail: cristian.rojas@ee.kth.se

We consider a finite impulse response system with centered independent sub-Gaussian design covariates and noise components that are not necessarily identically distributed. We derive non-asymptotic near-optimal estimation and prediction bounds for the least squares estimator of the parameters. Our results are based on two concentration inequalities on the norm of sums of dependent covariate vectors and on the singular values of their covariance operator that are of independent value on their own and where the dependence arises from the time shift structure of the time series. These results generalize the known bounds for the independent case.

Keywords: Finite impulse response; least squares; non-asymptotic estimation; shifted random vector; random covariance Toeplitz matrix; concentration inequality

References

Large deviations built on max-stability

MICHAEL KUPPER * and JOSÉ MIGUEL ZAPATA †

Department of Mathematics and Statistics, University of Konstanz, Konstanz, Germany.
E-mail: * kupper@uni-konstanz.de; † jmzg1@um.es

In this paper, we show that the basic results in large deviations theory hold for general monetary risk measures, which satisfy the crucial property of max-stability. A max-stable monetary risk measure fulfills a lattice homomorphism property, and satisfies under a suitable tightness condition the Laplace Principle (LP), that is, admits a dual representation with affine convex conjugate. By replacing asymptotic concentration of probability by concentration of risk, we formulate a Large Deviation Principle (LDP) for max-stable monetary risk measures, and show its equivalence to the LP. In particular, the special case of the asymptotic entropic risk measure corresponds to the classical Varadhan–Bryc equivalence between the LDP and LP. The main results are illustrated by the asymptotic shortfall risk measure.

Keywords: Large deviations; max-stable monetary risk measures; Large Deviation Principle; Laplace Principle; concentration function; asymptotic shortfall risk

References

Directional phantom distribution functions for stationary random fields

ADAM JAKUBOWSKI1,*, IGOR RODIONOV2 and NATALIA SOJA-KUKIEŁA1,†

1Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland.
E-mail: *adjakubo@mat.umk.pl; †natas@mat.umk.pl
2Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation.
E-mail: vecsell@gmail.com

We give necessary and sufficient conditions for the existence of a phantom distribution function for a stationary random field on a regular lattice. We also introduce a less demanding notion of a directional phantom distribution, with potentially broader area of applicability. Such approach leads to sectorial limit properties, a phenomenon well-known in limit theorems for random fields. An example of a stationary Gaussian random field is provided showing that the two notions do not coincide. Criteria for the existence of the corresponding notions of the extremal index and the sectorial extremal index are also given.

Keywords: Stationary random fields; extreme value limit theory; phantom distribution function; extremal index; Gaussian random fields

References


On fluctuations of global and mesoscopic linear statistics of generalized Wigner matrices

YITING LI$^1$ and YUANYUAN XU$^2$

$^1$Aix-Marseille Université, CNRS, Marseille, France.
E-mail: yiting.li@univ-amu.fr

$^2$KTH Royal Institute of Technology, Stockholm, Sweden.
E-mail: yuax@kth.se

We consider an $N$ by $N$ real or complex generalized Wigner matrix $H_N$, whose entries are independent centered random variables with uniformly bounded moments. We assume that the variance profile, $s_{ij} := \mathbb{E}|H_{ij}|^2$, satisfies $\sum_{i=1}^{N} s_{ij} = 1$, for all $1 \leq j \leq N$ and $c^{-1} \leq Ns_{ij} \leq c$ for all $1 \leq i, j \leq N$ with some constant $c \geq 1$. We establish Gaussian fluctuations for the linear eigenvalue statistics of $H_N$ on global scales, as well as on all mesoscopic scales up to the spectral edges, with the expectation and variance formulated in terms of the variance profile. We subsequently obtain the universal mesoscopic central limit theorems for the linear eigenvalue statistics inside the bulk and at the edges, respectively.

Keywords: Central limit theorem; linear eigenvalue statistics; generalized Wigner matrix

References


Limiting behavior of large correlated Wishart matrices with chaotic entries

SOLESNE BOURGUIN\textsuperscript{1}, CHARLES-PHILIPPE DIEZ\textsuperscript{2,*} and CIPRIAN A. TUDOR\textsuperscript{2,†}

\textsuperscript{1}Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, MA 02215, USA. E-mail: bourguin@math.bu.edu
\textsuperscript{2}CNRS, Laboratoire Paul Painlevé, Université de Lille, UMR 8524, F-59655 Villeneuve d’Ascq, France. E-mail: \textsuperscript{*}charles-philippe.diez@univ-lille.fr; \textsuperscript{†}ciprian.tudor@math.univ-lille.fr

We study the fluctuations, as \(d, n \to \infty\), of the Wishart matrix \(W_{n,d} = \frac{1}{d} \mathcal{X}_{n,d} \mathcal{X}_{n,d}^T\) associated to a \(n \times d\) random matrix \(\mathcal{X}_{n,d}\) with non-Gaussian entries. We analyze the limiting behavior in distribution of \(W_{n,d}\) in two situations: when the entries of \(\mathcal{X}_{n,d}\) are independent elements of a Wiener chaos of arbitrary order and when the entries are partially correlated and belong to the second Wiener chaos. In the first case, we show that the (suitably normalized) Wishart matrix converges in distribution to a Gaussian matrix while in the correlated case, we obtain its convergence in law to a diagonal non-Gaussian matrix. In both cases, we derive the rate of convergence in the Wasserstein distance via Malliavin calculus and analysis on Wiener space.

Keywords: Wishart matrix; multiple stochastic integrals; Malliavin calculus; Stein’s method; Rosenblatt process; fractional Brownian motion; high-dimensional regime

References


Functional registration and local variations: Identifiability, rank, and tuning

ANIRVAN CHAKRABORTY\textsuperscript{1} and VICTOR M. PANARETOS\textsuperscript{2}
\textsuperscript{1} Indian Institute of Science Education \\& Research (IISER) Kolkata, India. \\E-mail: anirvan.c@iiserkol.ac.in
\textsuperscript{2} Institut de Mathématiques, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. \\E-mail: victor.panaretos@epfl.ch

We develop theory and methodology for the problem of nonparametric registration of functional data that have been subjected to random deformation (warping) of their time scale. The separation of this phase variation (“horizontal” variation) from the amplitude variation (“vertical” variation) is crucial in order to properly conduct further analyses, which otherwise can be severely distorted. We determine precise nonparametric conditions under which the two forms of variation are identifiable. These show that the identifiability delicately depends on the underlying rank. By means of several counterexamples, we demonstrate that our conditions are sharp if one wishes a genuinely nonparametric setup; and in doing so we caution that popular remedies such as structural assumptions or roughness penalties can easily fail. We then propose a nonparametric registration method based on a “local variation measure”, the main element in elucidating identifiability. A key advantage of the method is that it is free of any tuning or penalisation parameters regulating the amount of alignment, thus circumventing the problem of over/under-registration often encountered in practice. We provide asymptotic theory for the resulting estimators under the identifiable regime, but also under mild departures from identifiability, quantifying the resulting bias in terms of the amplitude variation’s spectral gap.

Keywords: Functional data analysis; phase variation; synchronisation; warping

References

Well-posedness of distribution dependent SDEs with singular drifts

MICHAEL RÖCKNER$^{1,2}$ and XICHENG ZHANG$^3$

$^1$Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany
$^2$Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, 100190, P.R. China. E-mail: roeckner@math.uni-bielefeld.de
$^3$School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China. E-mail: XichengZhang@gmail.com

Consider the following distribution dependent SDE:

$$dX_t = \sigma_t(X_t, \mu_{X_t}) \, dW_t + b_t(X_t, \mu_{X_t}) \, dt,$$

where $\mu_{X_t}$ stands for the distribution of $X_t$. In this paper for non-degenerate $\sigma$, we show the strong well-posedness of the above SDE under some integrability assumptions in the spatial variable and Lipschitz continuity in $\mu$ about $b$ and $\sigma$. In particular, we extend the results of Krylov–Röckner (Probab. Theory Related Fields 131 (2005) 154–196) to the distribution dependent case.

Keywords: Distribution dependent SDEs; Zvonkin’s transformation; singular drifts; superposition principle; McKean–Vlasov system

References


Detecting a planted community in an inhomogeneous random graph

KAY BOGERD¹,* , RUI M. CASTRO¹,†, REMCO VAN DER HOFSTAD¹,‡ and NICOLAS VERZELEN²

¹Eindhoven University of Technology, Eindhoven, The Netherlands.
E-mail: *k.m.bogerd@tue.nl; †rmcastro@tue.nl; ‡r.w.v.d.hofstad@tue.nl
²INRAE, Montpellier SupAgro, MISTEA, Univ. Montpellier, France.
E-mail: nicolas.verzelen@inrae.fr

We study the problem of detecting whether an inhomogeneous random graph contains a planted community. Specifically, we observe a single realization of a graph. Under the null hypothesis, this graph is a sample from an inhomogeneous random graph, whereas under the alternative, there exists a small subgraph where the edge probabilities are increased by a multiplicative scaling factor. We present a scan test that is able to detect the presence of such a planted community, even when this community is very small and the underlying graph is inhomogeneous. We also derive an information theoretic lower bound for this problem which shows that in some regimes the scan test is almost asymptotically optimal. We illustrate our results through examples and numerical experiments.

Keywords: Community detection; inhomogeneous random graphs; scan statistics; minimax hypothesis testing

References


In this paper, we study a class of two sample test statistics based on inter-point distances in the high dimensional and low/medium sample size setting. Our test statistics include the well-known energy distance and maximum mean discrepancy with Gaussian and Laplacian kernels, and the critical values are obtained via permutations. We show that all these tests are inconsistent when the two high dimensional distributions correspond to the same marginal distributions but differ in other aspects of the distributions. The tests based on energy distance and maximum mean discrepancy mainly target the differences between marginal means and variances, whereas the test based on $L^1$-distance can capture the difference in marginal distributions. Our theory sheds new light on the limitation of inter-point distance based tests, the impact of different distance metrics, and the behavior of permutation tests in high dimension. Some simulation results and a real data illustration are also presented to corroborate our theoretical findings.

Keywords: Two sample test; high dimensionality; permutation test; power analysis

References


Minimax predictive density for sparse count data

KEISUKE YANO\(^1\), RYOYA KANEKO\(^2\) and FUMIYASU KOMAKI\(^3,4\)

\(^1\)The Institute of Statistical Mathematics, 10-3 Midori cho, Tachikawa City, Tokyo, 190-8562, Japan.
E-mail: yano@ism.ac.jp

\(^2\)Tokyo Marine Holdings, Inc., 1-2-1 Marunouchi, Chiyoda-ku, Tokyo, 100-8050, Japan.
E-mail: ryykaneko@gmail.com

\(^3\)Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan.
E-mail: komaki@g.ecc.u-tokyo.ac.jp

\(^4\)RIKEN Center for Brain Science, 2-1 Hirosawa, Wako City, Saitama, 351-0198, Japan

This paper discusses predictive densities under the Kullback–Leibler loss for high-dimensional Poisson sequence models under sparsity constraints. Sparsity in count data implies zero-inflation. We present a class of Bayes predictive densities that attain asymptotic minimaxity in sparse Poisson sequence models. We also show that our class with an estimator of unknown sparsity level plugged-in is adaptive in the asymptotically minimax sense. For application, we extend our results to settings with quasi-sparsity and with missing-completely-at-random observations. The simulation studies as well as application to real data illustrate the efficiency of the proposed Bayes predictive densities.

**Keywords:** Adaptation; high dimension; Kullback–Leibler divergence; missing at random; Poisson model; zero inflation

**References**


Asymptotics for sliding blocks estimators of rare events

HOLGER DREES* and SEBASTIAN NEBLUNG†

Department of Mathematics, SPST, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany.
E-mail: *drees@math.uni-hamburg.de; †sebastian.neblung@uni-hamburg.de

Drees and Rootzén (Ann. Statist. 38 (2010) 2145–2186) have established limit theorems for a general class of empirical processes of statistics that are useful for the extreme value analysis of time series, but do not apply to statistics of sliding blocks, including so-called runs estimators. We generalize these results to empirical processes which cover both the class considered by Drees and Rootzén (Ann. Statist. 38 (2010) 2145–2186) and processes of sliding blocks statistics. Using this approach, one can analyze different types of statistics in a unified framework. We show that statistics based on sliding blocks are asymptotically normal with an asymptotic variance which, under rather mild conditions, is smaller than or equal to the asymptotic variance of the corresponding estimator based on disjoint blocks. Finally, the general theory is applied to three well-known estimators of the extremal index. It turns out that they all have the same limit distribution, a fact which has so far been overlooked in the literature.

Keywords: Asymptotic efficiency; empirical processes; extremal index; extreme value analysis; sliding vs disjoint blocks; time series; uniform central limit theorems

References


On $\mu$-Dvoretzky random covering of the circle

AIHUA FAN$^{1,2}$ and DAVIT KARAGULYAN$^3$

$^1$LAMFA, UMR 7352, CNRS, University of Picardie, 33 rue Saint Leu, 80039 Amiens CEDEX 1, France. E-mail: ai-hua.fan@u-picardie.fr
$^2$Department of Mathematics, University of Maryland, College Park 20742, USA.
$^3$School of Mathematics and Statistics, Central China Normal University, 152 Luoyu road, Wuhan 430077, China. E-mail: dkaragul@umd.edu

In this paper, we study the Dvoretzky covering problem with non-uniformly distributed centers. When the probability law of the centers is absolutely continuous w.r.t. Lebesgue measure and satisfies a regularity condition on the set of essential infimum points, we give a necessary and sufficient condition for covering the circle. When the lengths of covering intervals are of the form $\ell_n = \frac{c}{n}$, we give a necessary and sufficient condition for covering the circle, without imposing any regularity on the density function.

Keywords: Random covering; non-uniform densities

References


Consider a spectrally positive Lévy process $Z$ with log-Laplace exponent $\Psi_1$ and a positive continuous function $R$ on $(0, \infty)$. We investigate the entrance from infinity of the process $X$ obtained by changing time in $Z$ with the inverse of the additive functional $\eta(t) = \int_0^t R(Z_s) \, ds$. We provide a necessary and sufficient condition for infinity to be an entrance boundary of the process $X$. Under this condition, the process can start from infinity and we study its speed of coming down from infinity. When the Lévy process has a negative drift $\delta := -\gamma < 0$, sufficient conditions over $R$ and $\Psi_1$ are found for the process to come down from infinity along the deterministic function $(x_t, t \geq 0)$ solution to $dx_t = -\gamma R(x_t) \, dt$ with $x_0 = \infty$. If $\Psi(\lambda) \sim c\lambda^\alpha$ as $\lambda \to 0$, $\alpha \in (1, 2]$, $c > 0$ and $R$ is regularly varying at $\infty$ with index $\theta > \alpha$, the process comes down from infinity and we find a renormalisation in law of its running infimum at small times.

**Keywords:** Coming down from infinity; entrance boundary; hitting time; continuous-state non-linear branching process; regularly varying function; spectrally positive Lévy process; time-change; weighted occupation time

**References**


Precise large deviations for dependent subexponential variables

THOMAS MIKOSCH\textsuperscript{1} and IGOR RODIONOV\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.
E-mail: mikosch@math.ku.dk

\textsuperscript{2}V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Profsoyuznaya ulitsa 65, 117997, Moscow, Russia.
E-mail: vecsell@gmail.com

In this paper, we study precise large deviations for the partial sums of a stationary sequence with a subexponential marginal distribution. Our main focus is on distributions which either have a regularly varying or a lognormal-type tail. We apply the results to prove limit theory for the maxima of the entries large sample covariance matrices.

\textbf{Keywords:} Large deviation probability; subexponential distribution; maximum domain of attraction; Gumbel distribution; Fréchet distribution; regular variation; stationary sequence

\textbf{References}


Adaptation bounds for confidence bands under self-similarity

TIMOTHY B. ARMSTRONG

Department of Economics, Yale University, 30 Hillhouse Ave., New Haven, CT 06511, USA.
E-mail: timothy.armstrong@yale.edu

We derive bounds on the scope for a confidence band to adapt to the unknown regularity of a nonparametric function that is observed with noise, such as a regression function or density, under the self-similarity condition proposed by Giné and Nickl (Ann. Statist. 38 (2010) 1122–1170). We find that adaptation can only be achieved up to a term that depends on the choice of the constant used to define self-similarity, and that this term becomes arbitrarily large for conservative choices of the self-similarity constant. We construct a confidence band that achieves this bound, up to a constant term that does not depend on the self-similarity constant. Our results suggest that care must be taken in choosing and interpreting the constant that defines self-similarity, since the dependence of adaptive confidence bands on this constant cannot be made to disappear asymptotically.

Keywords: Honest confidence interval; adaptation; self-similarity

References

Compound Poisson approximation for regularly varying fields with application to sequence alignment

BOJAN BASRAK* and HRVOJE PLANINIĆ†

Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia.
E-mail: *bbasrak@math.hr; †planinic@math.hr

The article determines the asymptotic shape of the extremal clusters in stationary regularly varying random fields. To deduce this result, we present a general framework for the Poisson approximation of point processes on Polish spaces which appears to be of independent interest. We further introduce a novel and convenient concept of anchoring of the extremal clusters for regularly varying sequences and fields. Together with the Poissonian approximation theory, this allows for a concise description of the limiting behavior of random fields in this setting. We apply this theory to shed entirely new light on the classical problem of evaluating local alignments of biological sequences.

Keywords: Compound Poisson approximation; random fields; regular variation; tail process; point process; local sequence alignment; Gumbel distribution

References


Bootstrapping Hill estimator and tail array sums for regularly varying time series

CARSTEN JENTSCH¹ and RAFAŁ KULIK²

¹Faculty of Statistics, TU Dortmund University, D-44221 Dortmund, Germany
²Department of Mathematics and Statistics, University of Ottawa.
E-mail: rkulik@uottawa.ca

In the extreme value analysis of stationary regularly varying time series, tail array sums form a broad class of statistics suitable to analyze their extremal behavior. This class includes for example, the Hill estimator or estimators of the extremogram and the tail dependence coefficient.


However, the resulting limiting distributions turn out to be very complex and cumbersome to estimate as they usually depend on the whole extremal dependence structure of the time series. Hence, a suitable bootstrap procedure is desired, but available bootstrap consistency results for tail array sums are scarce. In this paper, following Drees (Drees (2015)), we consider a multiplier block bootstrap to estimate the limiting distribution of tail array sums. We prove that, conditionally on the data, an appropriately constructed multiplier block bootstrap statistic converges to the correct limiting distribution. Interestingly, in contrast, it turns out that an apparently natural, but naïve application of the multiplier block bootstrap scheme does not yield the correct limit.

In simulations, we provide numerical evidence of our theoretical findings and illustrate the superiority of the proposed multiplier block bootstrap over some obvious competitors. The proposed bootstrap scheme proves to be computationally efficient in comparison to other approaches.

Keywords: Heavy tails; Hill estimator; multiplier bootstrap; regular variation; stationary time series; tail empirical process; tail array sums

References


Bernoulli Forthcoming Papers

GNEDIN, A. and SEKSENBAEV, A.
Asymptotics and renewal approximation in the online selection of increasing sub-sequence

KRAMPE, J., KREISS, J.-P. and PAPARODITIS, E.
Bootstrap based inference for sparse high-dimensional time series models

LAW, M. and RITOV, Y.
Inference without compatibility: Using exponential weighting for inference on a parameter of a linear model

KRAAIJ, R.C.
Flux large deviations of weakly interacting jump processes via well-posedness of an associated Hamilton–Jacobi equation

BETZ, V., MÜHLBAUER, J., SCHÄFER, H. and ZEINDLER, D.
Precise asymptotics of longest cycles in random permutations without macroscopic cycles

CATOR, E. and DON, H.
Explicit bounds for critical infection rates and expected extinction times of the contact process on finite random graphs

ABI JABER, E.
Weak existence and uniqueness for affine stochastic Volterra equations with $L^1$-kernels

AGAPIOU, S., DASHTI, M. and HELIN, T.
Rates of contraction of posterior distributions based on $p$-exponential priors

CARDONA-TOBÓN, N. and PALAU, S.
Yaglom’s limit for critical Galton–Watson processes in varying environment: A probabilistic approach

KREISS, A.
Correlation bounds, mixing and $m$-dependence under random time-varying network distances with an application to Cox-processes

KREBS, J.
On the law of the iterated logarithm and strong invariance principles in stochastic geometry

FONTES, L.R., GOMES, P.A. and SANCHIS, R.
Contact process under heavy-tailed renewals on finite graphs

GARINO, V., NOURDIN, I., NUALART, D. and SALAMAT, M.
Limit theorems for integral functionals of Hermite-driven processes

KOVCHEGOV, Y. and ZALIAPIN, I.
Invariance and attraction properties of Galton–Watson trees

BOBBIA, B., DOMBRY, C. and VARRON, D.
The coupling method in extreme value theory

Continues
JAKUBOWSKI, J. and WIŚNIEWOLSKI, M.
A convolution formula for the local time of an Itô diffusion reflecting at 0 and a
generalized Stroock–Williams equation

FAN, W.-T.L.
Stochastic PDEs on graphs as scaling limits of discrete interacting systems

MA, Y.-A., CHATTERJI, N.S., CHENG, X., FLAMMARION, N., BARTLETT,
P.L. and JORDAN, M.I.
Is there an analog of Nesterov acceleration for gradient-based MCMC?

BREHMER, J.R. and GNEITING, T.
Scoring interval forecasts: Equal-tailed, shortest, and modal interval

CÉNAC, P., CHAUVIN, B., NOÛS, C., PACCAUT, F. and POUYANNE, N.
Variable Length Memory Chains: Characterization of stationary probability mea-
sures

HUANG, L.-J., MAJKA, M.B. and WANG, J.
Approximation of heavy-tailed distributions via stable-driven SDEs

KUTOYANTS, Y.A.
On multi-step estimation of delay for SDE

NICOLUSSI, F. and CAZZARO, M.
Context-specific independencies in stratified chain regression graphical models

KABLUCHKO, Z., PROCHNO, J. and THÄLE, C.
A new look at random projections of the cube and general product measures

GUHA, A., HO, N. and NGUYEN, X.L.
On posterior contraction of parameters and interpretability in Bayesian mixture
modeling

HUANG, D. and TROPP, J.A.
From Poincaré inequalities to nonlinear matrix concentration

BAHMANI, S.
Nearly optimal robust mean estimation via empirical characteristic function

BABIC, S., GELBGRAS, L., HALLIN, M. and LEY, C.
Optimal tests for elliptical symmetry: Specified and unspecified location

KASPER, T., FUCHS, S. and TRUTSCHNIG, W.
On weak conditional convergence of bivariate Archimedean and Extreme Value
copulas, and consequences to nonparametric estimation

PAGE, S. and GRUNEWALDER, S.
The Goldenshluger–Lepski method for constrained least-squares estimators over
RKHSs

KLEBANOV, I., SPRUNGK, B. and SULLIVAN, T.J.
The linear conditional expectation in Hilbert space

Continues
Bernoulli Forthcoming Papers—Continued

QIU, H., LUEDTKE, A. and CARONE, M.
Universal sieve-based strategies for efficient estimation using machine learning tools

JIANG, D. and BAI, Z.
Partial generalized four moment theorem revisited. A note on “Generalized four moment theorem and its application to CLT for spiked eigenvalues of high-dimensional covariance matrices”

PETE, G. and TIMAR, A.
Finite-energy infinite clusters without anchored expansion

SCHMISSER, E. and KRELL, N.
Nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes

CHZHEN, E., DENIS, C. and HEBIRI, M.
Minimax semi-supervised set-valued approach to multi-class classification

CAZARES, J.G. and IVANOVA, J.
Recovering Brownian and jump parts from high-frequency observations of a Levy process

LEDOIT, O. and WOLF, M.
Quadratic shrinkage for large covariance matrices

BARBER, R.F. and SAMWORTH, R.J.
Local continuity of log-concave projection, with applications to estimation under model misspecification

VERGARA, R.C., ALLARD, D. and DESASSIS, N.
A general framework for SPDE-based stationary random fields

GANESAN, G.
Minimum spanning trees of random geometric graphs with location dependent weights

BHUDISAKSANG, T. and CARTEA, A.
Online drift estimation for jump-diffusion processes

CHAUDHURI, R., JAIN, V. and PILLAI, N.S.
Universality and least singular values of random matrix products: A simplified approach

NING, N., IONIDES, E. and RITOV, Y.
Scalable Monte Carlo inference and rescaled local asymptotic normality

WANG, D. and TANG, C.-F.
Testing against uniform stochastic ordering with paired observations

KOHLER, M. and KRZYZAK, A.
Over-parametrized deep neural networks do not generalize well

Continues
BORDA, B.
Equidistribution of random walks on compact groups II. The Wasserstein metric

CERNY, A. and RUŽ, J.
Pure-jump semimartingales

DUVAL, C. and MARIUCCI, E.
Spectral-free estimation of Lévy densities in high-frequency regime

DENIS, C., DION, C. and MARTINEZ, M.
A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation