Preface: Section of memorial articles for Willem van Zwet

Richard J. Samworth and Ming Yuan 2431

Willem van Zwet’s contributions to the profession

Nicholas I. Fisher and Adrian F. M. Smith 2432

Willem van Zwet’s research

Peter Bickel, Marta Fiocco, Mathisca de Gunst and Friedrich Götzte 2439

Willem van Zwet, teacher and thesis advisor

Sara van de Geer and Chris A. J. Klaassen 2448

Optimal rates for independence testing via $U$-statistic permutation tests

Thomas B. Berrett, Ioannis Kontoyiannis and Richard J. Samworth 2457

Variable selection consistency of Gaussian process regression

Sheng Jiang and Surya T. Tokdar 2491

Optimality of spectral clustering in the Gaussian mixture model

Matthias Löffler, Anderson Y. Zhang and Harrison H. Zhou 2506

Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees

Sabyasachi Chatterjee and Subhajit Goswami 2531

Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes

Michael Lalancette, Sebastian Engelke and Stanislav Volgushev 2552

Estimation of smooth functionals in normal models: Bias reduction and asymptotic efficiency

Vladimir Koltchinskii and Mayya Zhiyova 2577

Additive regression for non-Euclidean responses and predictors

Jeong Min Jeon, Byeong U. Park and Ingrid Van Keilegom 2611

Set structured global empirical risk minimizers are rate optimal in general dimensions

Qiyang Han 2642

Empirical tail copulas for functional data

John H. J. Einmahl and Johan Segers 2672

Inference for a two-stage enrichment design

Zhantao Lin, Nancy Flournoy and William F. Rosenberger 2697

Existence and uniqueness of the Kronecker covariance MLE

Mathias Drton, Satoshi Kuriki and Peter Hoff 2721

Prediction bounds for higher order total variation regularized least squares

Francesco Orteelli and Sara van de Geer 2755

Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain

Suhasini Subba Rao and Junho Yang 2774

Integrative methods for post-selection inference under convex constraints

Snigdha Panigrahi, Jonathan Taylor and Asaf Weinstein 2803

The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy

T. Tony Cai, Yichen Wang and Linjun Zhang 2825
Distributed statistical inference for massive data . . . SONG XI CHEN AND LIUHUA PENG 2851
Construction of mixed orthogonal arrays with high strength
SHANQI PANG, JING WANG, DENNIS K. J. LIN AND MIN-QIAN LIU 2870
Foundations of structural causal models with cycles and latent variables
STEPHAN BONGERS, PATRICK FORRÊ, JONAS PETERS AND JORIS M. MOOIJ 2885
Measuring dependence in the Wasserstein distance for Bayesian nonparametric models
MARTA CATALANO, ANTONIO LIJOI AND IGOR PRÜNSTER 2916
Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data . . . . . . . . . . . . YUXIN CHEN, JIANQING FAN, CONG MA AND YULING YAN 2948
Efficiency of delayed-acceptance random walk Metropolis algorithms
CHRIS SHERLOCK, ALEXANDRE H. THIERY AND ANDREW GOLIGHTLY 2972
Semiparametric optimal estimation with nonignorable nonresponse data
KOSUKE MORIKAWA AND JAE KWANG KIM 2991
Two-level parallel flats designs . . . . . . . . . . . . CHUNYAN WANG AND ROBERT W. MEE 3015
PREFACE: SECTION OF MEMORIAL ARTICLES FOR WILLEM VAN ZWET

BY RICHARD J. SAMWORTH¹ AND MING YUAN

¹Statistical Laboratory, University of Cambridge, r.samworth@statslab.cam.ac.uk
²Department of Statistics, Columbia University, ming.yuan@columbia.edu

REFERENCES


WILLEM VAN ZWET’S CONTRIBUTIONS TO THE PROFESSION

BY NICHOLAS I. FISHER\textsuperscript{1} AND ADRIAN F. M. SMITH\textsuperscript{2}

\textsuperscript{1}School of Mathematics & Statistics, University of Sydney, Nicholas.Fisher@sydney.edu.au
\textsuperscript{2}The Alan Turing Institute, InstituteDirector@turing.ac.uk

Bill van Zwet’s contributions to his profession are, in their own way, comparable to his very substantial research, research supervision and teaching contributions. He had several leadership roles in professional societies, edited leading theoretical journals, started up or revived important conference series, founded a research center and played a huge role in providing colleagues in Central and Eastern Europe access to the western world of probability and statistics. And to each of these several activities, he brought farsightedness, energy and wit, qualities that shone through when he was interviewed about what he’d done (\textit{Statist. Sci.} \textbf{24} (2009) 87–115). In fact, because Bill provided a wealth of detail in “the interview,” this article focuses more on the recollections of some of the people with whom he interacted.

REFERENCES


Willem van Zwet made deep and influential contributions to probability and statistics, which we review in this paper. Bickel and Götze collaborated with him on his major contributions to higher order asymptotics of nonlinear statistics and on resampling and the bootstrap. We relate this work to his remarkable development of the properties of the Hoeffding expansion for symmetric statistics as well as Fourier analytic tools. Fiocco and De Gunst were his students. We describe how in their theses and subsequent papers with him they developed statistical inference in two subtle stochastic models, the contact process and cell population development under a plausible regime. We also touch on his solutions of intriguing problems not related directly to his main interests.

REFERENCES


MSC2020 subject classifications. Primary 01A70; secondary 62-03.

Key words and phrases. Van Zwet, asymptotic approximation, resampling, contact process.


WILLEM VAN ZWET, TEACHER AND THESIS ADVISOR

EDITED BY SARA VAN DE GEER¹ AND CHRIS A. J. KLAASSEN²

¹Department of Mathematics, ETH Zürich, sara.vandegeer@stat.math.ethz.ch
²Korteweg-de Vries Institute for Mathematics, University of Amsterdam, c.a.j.klaassen@uva.nl

Willem van Zwet was supervisor of sixteen PhD students. All of them pursued academic careers and most of them became full professor. Below are some stories of PhD students Wim Albers, Cees Diks, Ronald Does, Marta Fiocco, Sara van de Geer, Mathisca de Gunst, Chris Klaassen, Hein Putter, Aad van der Vaart, Marten Wegkamp and Martien van Zuijlen with in addition a contribution by Nelly Litvak who was guided by Willem after her PhD.

MSC2020 subject classifications. Primary 01A70; secondary 62-03.

Key words and phrases. Obituary, research supervision, PhD students, teaching.
OPTIMAL RATES FOR INDEPENDENCE TESTING VIA U-STATISTIC PERMUTATION TESTS

BY THOMAS B. BERRETT1, IOANNIS KONTOYIANNIS2,* AND RICHARD J. SAMWORTH2,†

1Department of Statistics, University of Warwick, tom.berrett@warwick.ac.uk
2Statistical Laboratory, Centre for Mathematical Sciences, *yiannis@maths.cam.ac.uk; †r.samworth@statslab.cam.ac.uk

We study the problem of independence testing given independent and identically distributed pairs taking values in a σ-finite, separable measure space. Defining a natural measure of dependence \( D(f) \) as the squared \( L^2 \)-distance between a joint density \( f \) and the product of its marginals, we first show that there is no valid test of independence that is uniformly consistent against alternatives of the form \( \{ f : D(f) \geq \rho^2 \} \). We therefore restrict attention to alternatives that impose additional Sobolev-type smoothness constraints, and define a permutation test based on a basis expansion and a U-statistic estimator of \( D(f) \) that we prove is minimax optimal in terms of its separation rates in many instances. Finally, for the case of a Fourier basis on \([0,1]^2\), we provide an approximation to the power function that offers several additional insights. Our methodology is implemented in the R package USP.

REFERENCES


Key words and phrases. Independence testing, permutation tests, minimax separation rates, U-statistics, Stein’s method.
PITMAN, E. J. G. (1938). Significance tests which may be applied to samples from any populations: III. The analysis of variance test. Biometrika 29 322–335.


VARIABLE SELECTION CONSISTENCY OF GAUSSIAN PROCESS REGRESSION

BY SHENG JIANG* AND SURYA T. TOKDAR†

Department of Statistical Science, Duke University, *sheng.jiang@duke.edu; †surya.tokdar@duke.edu

Bayesian nonparametric regression under a rescaled Gaussian process prior offers smoothness-adaptive function estimation with near minimax-optimal error rates. Hierarchical extensions of this approach, equipped with stochastic variable selection, are known to also adapt to the unknown intrinsic dimension of a sparse true regression function. But it remains unclear if such extensions offer variable selection consistency, that is, if the true subset of important variables could be consistently learned from the data. It is shown here that variable consistency may indeed be achieved with such models at least when the true regression function has finite smoothness to induce a polynomially larger penalty on inclusion of false positive predictors. Our result covers the high-dimensional asymptotic setting where the predictor dimension is allowed to grow with the sample size. The proof utilizes Schwartz theory to establish that the posterior probability of wrong selection vanishes asymptotically. A necessary and challenging technical development involves providing sharp upper and lower bounds to small ball probabilities at all rescaling levels of the Gaussian process prior, a result that could be of independent interest.

REFERENCES


MSC2020 subject classifications. Primary 62G08, 62G20; secondary 60G05.

Key words and phrases. Gaussian process priors, high-dimensional regression, nonparametric variable selection, Bayesian inference, adaptive estimation.


OPTIMALITY OF SPECTRAL CLUSTERING IN THE GAUSSIAN MIXTURE MODEL

BY MATTHIAS LÖFFLER\textsuperscript{1}, ANDERSON Y. ZHANG\textsuperscript{2}, AND HARRISON H. ZHOU\textsuperscript{3}

\textsuperscript{1}Seminar for Statistics, ETH Zürich, matthias.oeffler@stat.math.ethz.ch
\textsuperscript{2}Department of Statistics, University of Pennsylvania, ayz@wharton.upenn.edu
\textsuperscript{3}Department of Statistics and Data Science, Yale University, huibin.zhou@yale.edu

Spectral clustering is one of the most popular algorithms to group high-dimensional data. It is easy to implement and computationally efficient. Despite its popularity and successful applications, its theoretical properties have not been fully understood. In this paper, we show that spectral clustering is minimax optimal in the Gaussian mixture model with isotropic covariance matrix, when the number of clusters is fixed and the signal-to-noise ratio is large enough. Spectral gap conditions are widely assumed in the literature to analyze spectral clustering. On the contrary, these conditions are not needed to establish optimality of spectral clustering in this paper.

REFERENCES


MSC2020 subject classifications. 62H30.

Key words and phrases. Spectral clustering, K-means, Gaussian mixture model, Spectral perturbation.


ADAPTIVE ESTIMATION OF MULTIVARIATE PIECEWISE POLYNOMIALS
AND BOUNDED VARIATION FUNCTIONS BY OPTIMAL DECISION TREES

BY SABYASACHI CHATTERJEE1 AND SUBHAJIT GOSWAMI2

1Department of Statistics, University of Illinois at Urbana–Champaign, sc1706@illinois.edu
2Tata Institute of Fundamental Research, goswami@math.tifr.res.in

Proposed by Donoho (Ann. Statist. 25 (1997) 1870–1911), Dyadic CART is a nonparametric regression method which computes a globally optimal dyadic decision tree and fits piecewise constant functions in two dimensions. In this article, we define and study Dyadic CART and a closely related estimator, namely Optimal Regression Tree (ORT), in the context of estimating piecewise smooth functions in general dimensions in the fixed design setup. More precisely, these optimal decision tree estimators fit piecewise polynomials of any given degree. Like Dyadic CART in two dimensions, we reason that these estimators can also be computed in polynomial time in the sample size $N$ via dynamic programming. We prove oracle inequalities for the finite sample risk of Dyadic CART and ORT, which imply tight risk bounds for several function classes of interest. First, they imply that the finite sample risk of ORT of order $r \geq 0$ is always bounded by $Ck \log \frac{N}{N}$ whenever the regression function is piecewise polynomial of degree $r$ on some reasonably regular axis aligned rectangular partition of the domain with at most $k$ rectangles. Beyond the univariate case, such guarantees are scarcely available in the literature for computationally efficient estimators. Second, our oracle inequalities uncover minimax rate optimality and adaptivity of the Dyadic CART estimator for function spaces with bounded variation. We consider two function spaces of recent interest where multivariate total variation denoising and univariate trend filtering are the state of the art methods. We show that Dyadic CART enjoys certain advantages over these estimators while still maintaining all their known guarantees.

REFERENCES


MSC2020 subject classifications. 62G05.
Key words and phrases. Optimal decision trees, Dyadic CART, piecewise polynomial fitting, oracle risk bounds, bounded variation function estimation.

RANK-BASED ESTIMATION UNDER ASYMPTOTIC DEPENDENCE AND INDEPENDENCE, WITH APPLICATIONS TO SPATIAL EXTREMES

BY Michaël Lalancette1,*, Sebastian Engelke2 and Stanislav Volgushev1,†

1Department of Statistical Sciences, University of Toronto, *lalancette@utstat.toronto.edu; †stanislav.volgushev@utoronto.ca
2Research Center for Statistics, University of Geneva, sebastian.engelke@unige.ch

Multivariate extreme value theory is concerned with modeling the joint tail behavior of several random variables. Existing work mostly focuses on asymptotic dependence, where the probability of observing a large value in one of the variables is of the same order as observing a large value in all variables simultaneously. However, there is growing evidence that asymptotic independence is equally important in real world applications. Available statistical methodology in the latter setting is scarce and not well understood theoretically. We revisit nonparametric estimation and introduce rank-based M-estimators for parametric models that simultaneously work under asymptotic dependence and asymptotic independence, without requiring prior knowledge on which of the two regimes applies. Asymptotic normality of the proposed estimators is established under weak regularity conditions. We further show how bivariate estimators can be leveraged to obtain parametric estimators in spatial tail models, and again provide a thorough theoretical justification for our approach.

REFERENCES


Key words and phrases. Multivariate extremes, asymptotic independence, inverted max-stable distribution, spatial process, M-estimation.
ESTIMATION OF SMOOTH FUNCTIONALS IN NORMAL MODELS: BIAS REDUCTION AND ASYMPTOTIC EFFICIENCY

BY VLADIMIR KOLTCHINSKII* AND MAYYA ZHILOVA†

School of Mathematics, Georgia Institute of Technology, *vlad@math.gatech.edu; †mzhilova@math.gatech.edu

Let $X_1, \ldots, X_n$ be i.i.d. random variables sampled from a normal distribution $N(\mu, \Sigma)$ in $\mathbb{R}^d$ with unknown parameter $\theta = (\mu, \Sigma) \in \Theta := \mathbb{R}^d \times C_+^d$, where $C_+^d$ is the cone of positively definite covariance operators in $\mathbb{R}^d$. Given a smooth functional $f : \Theta \mapsto \mathbb{R}$, the goal is to estimate $f(\theta)$ based on $X_1, \ldots, X_n$. Let

$$\Theta(a; d) := \mathbb{R}^d \times \{\Sigma \in C_+^d : \sigma(\Sigma) \subset [1/a, a]\}, \quad a \geq 1,$$

where $\sigma(\Sigma)$ is the spectrum of covariance $\Sigma$. Let $\hat{\theta} := (\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu}$ is the sample mean and $\hat{\Sigma}$ is the sample covariance, based on the observations $X_1, \ldots, X_n$. For an arbitrary functional $f \in C^s(\Theta)$, $s = k + 1 + \rho$, $k \geq 0$, $\rho \in (0, 1)$, we define a functional $f_k : \Theta \mapsto \mathbb{R}$ such that

$$\sup_{\theta \in \Theta(a; d)} \|f_k(\hat{\theta}) - f(\theta)\|_{L_2(\mathbb{P}_\theta)} \lesssim a^\beta \|f\|_{C^s(\Theta)} \left[\left(\frac{a}{\sqrt{n}}\right)^{d/s} \right] \land 1,$$

where $\beta = 1$ for $k = 0$ and $\beta > 1$ is arbitrary for $k \geq 1$. This error rate is minimax optimal and similar bounds hold for more general loss functions. If $d = d_0 \leq n^a$ for some $a \in (0, 1)$ and $s \geq \frac{1}{1-\alpha}$, the rate becomes $O(n^{-1/2})$. Moreover, for $s > \frac{1}{1-\alpha}$, the estimator $f_k(\hat{\theta})$ is shown to be asymptotically efficient. The crucial part of the construction of estimator $f_k(\hat{\theta})$ is a bias reduction method studied in the paper for more general statistical models than normal.

REFERENCES


Key words and phrases. Efficiency, smooth functionals, bias reduction, bootstrap chain, random homotopy, concentration.


Additive regression is studied in a very general setting where both the response and predictors are allowed to be non-Euclidean. The response takes values in a general separable Hilbert space, whereas the predictors take values in general semimetric spaces, which covers a very wide range of nonstandard response variables and predictors. A general framework of estimating additive models is presented for semimetric space-valued predictors. In particular, full details of implementation and the corresponding theory are given for predictors taking values in Hilbert spaces and/or Riemannian manifolds. The existence of the estimators, convergence of a backfitting algorithm, rates of convergence and asymptotic distributions of the estimators are discussed. The finite sample performance of the estimators is investigated by means of two simulation studies. Finally, three data sets covering several types of non-Euclidean data are analyzed to illustrate the usefulness of the proposed general approach.


SET STRUCTURED GLOBAL EMPIRICAL RISK MINIMIZERS ARE RATE OPTIMAL IN GENERAL DIMENSIONS

BY QIYANG HAN

Department of Statistics, Rutgers University, qh85@stat.rutgers.edu

Entropy integrals are widely used as a powerful empirical process tool to obtain upper bounds for the rates of convergence of global empirical risk minimizers (ERMs), in standard settings such as density estimation and regression. The upper bound for the convergence rates thus obtained typically matches the minimax lower bound when the entropy integral converges, but admits a strict gap compared to the lower bound when it diverges. Birgé and Massart (Probab. Theory Related Fields 97 (1993) 113–150) provided a striking example showing that such a gap is real with the entropy structure alone: for a variant of the natural Hölder class with low regularity, the global ERM actually converges at the rate predicted by the entropy integral that substantially deviates from the lower bound. The counter-example has spawned a long-standing negative position on the use of global ERMs in the regime where the entropy integral diverges, as they are heuristically believed to converge at a suboptimal rate in a variety of models.

The present paper demonstrates that this gap can be closed if the models admit certain degree of “set structures” in addition to the entropy structure. In other words, the global ERMs in such set structured models will indeed be rate-optimal, matching the lower bound even when the entropy integral diverges. The models with set structures we investigate include (i) image and edge estimation, (ii) binary classification, (iii) multiple isotonic regression, (iv) $s$-concave density estimation, all in general dimensions when the entropy integral diverges. Here, set structures are interpreted broadly in the sense that the complexity of the underlying models can be essentially captured by the size of the empirical process over certain class of measurable sets, for which matching upper and lower bounds are obtained to facilitate the derivation of sharp convergence rates for the associated global ERMs.

REFERENCES


MSC2020 subject classifications. Primary 60E15; secondary 62G05.

Key words and phrases. Empirical risk minimization, empirical process, classification, nonparametric regression, density estimation, non-Donsker.


EMPIRICAL TAIL COPULAS FOR FUNCTIONAL DATA

BY JOHN H. J. EINMAHL 1 AND JOHAN SEGERS 2

1 Department of Econometrics and OR and CentER, Tilburg University, j.h.j.einmahl@uvt.nl
2 LIDAM/ISBA, UCLouvain, johan.segers@uclouvain.be

For multivariate distributions in the domain of attraction of a max-stable distribution, the tail copula and the stable tail dependence function are equivalent ways to capture the dependence in the upper tail. The empirical versions of these functions are rank-based estimators whose inflated estimation errors are known to converge weakly to a Gaussian process that is similar in structure to the weak limit of the empirical copula process. We extend this multivariate result to continuous functional data by establishing the asymptotic normality of the estimators of the tail copula, uniformly over all finite subsets of at most D points (D fixed). An application for testing tail copula stationarity is presented. The main tool for deriving the result is the uniform asymptotic normality of all the D-variate tail empirical processes. The proof of the main result is nonstandard.

REFERENCES


Key words and phrases. Extreme value statistics, functional data, tail empirical process, tail dependence, tail copula estimation, uniform asymptotic normality.


INFERENCE FOR A TWO-STAGE ENRICHMENT DESIGN

BY ZHANTAO LIN\(^1,\,*\), NANCY FLOURNOY\(^2\) AND WILLIAM F. ROSENBERGER\(^1,\,\dagger\)

\(^1\)Department of Statistics, George Mason University, \(^*\)zlin4@masonlive.gmu.edu; \(^\dagger\)wrosenbe@gmu.edu

\(^2\)Department of Statistics, University of Missouri, Columbia, flournoyn@missouri.edu

Two-stage enrichment designs can be used to target the benefiting population in clinical trials based on patients’ biomarkers. In the case of continuous biomarkers, we show that using a bivariate model that treats biomarkers as random variables more accurately identifies a treatment-benefiting enriched population than assuming biomarkers are fixed. Additionally, we show that under the bivariate model, the maximum likelihood estimators (MLEs) follow a randomly scaled mixture of normal distributions. Using random normings, we obtain asymptotically standard normal MLEs and construct hypothesis tests. Finally, in a simulation study, we demonstrate that our proposed design is more powerful than a single stage design when outcomes and biomarkers are correlated; the model-based estimators have smaller bias and mean square error (MSE) than weighted average estimators.

REFERENCES


MSC2020 subject classifications. Primary 62L05, 62K99, 62F12; secondary 62E20, 60F05, 60G52.

Key words and phrases. Adaptive designs, design of experiments, inference for stochastic processes, precision medicine, random biomarker, threshold determination.
EXISTENCE AND UNIQUENESS OF THE KRONECKER COVARIANCE MLE

BY MATHIAS DRTON¹, SATOSHI KURIKI² AND PETER HOFF³

¹Department of Mathematics, Technical University of Munich, mathias.drton@tum.de
²The Institute of Statistical Mathematics, kuriki@ism.ac.jp
³Department of Statistical Science, Duke University, peter.hoff@duke.edu

In matrix-valued datasets the sampled matrices often exhibit correlations among both their rows and their columns. A useful and parsimonious model of such dependence is the matrix normal model, in which the covariances among the elements of a random matrix are parameterized in terms of the Kronecker product of two covariance matrices, one representing row covariances and one representing column covariance. An appealing feature of such a matrix normal model is that the Kronecker covariance structure allows for standard likelihood inference even when only a very small number of data matrices is available. For instance, in some cases a likelihood ratio test of dependence may be performed with a sample size of one. However, more generally the sample size required to ensure boundedness of the matrix normal likelihood or the existence of a unique maximizer depends in a complicated way on the matrix dimensions. This motivates the study of how large a sample size is needed to ensure that maximum likelihood estimators exist, and exist uniquely with probability one. Our main result gives precise sample size thresholds in the paradigm where the number of rows and the number of columns of the data matrices differ by at most a factor of two. Our proof uses invariance properties that allow us to consider data matrices in canonical form, as obtained from the Kronecker canonical form for matrix pencils.

REFERENCES


MSC2020 subject classifications. Primary 62H12; secondary 62R01.

Key words and phrases. Gaussian distribution, Kronecker canonical form, matrix normal model, maximum likelihood estimation, separable covariance.
PREDICTION BOUNDS FOR HIGHER ORDER TOTAL VARIATION REGULARIZED LEAST SQUARES

BY FRANCESCO ORTELLI* AND SARA VAN DE GEER†

Seminar for Statistics, ETH Zürich, *fortelli@ethz.ch; †vsara@ethz.ch

We establish adaptive results for trend filtering: least squares estimation with a penalty on the total variation of \((k - 1)\)th order differences. Our approach is based on combining a general oracle inequality for the \(\ell_1\)-penalized least squares estimator with “interpolating vectors” to upper bound the “effective sparsity.” This allows one to show that the \(\ell_1\)-penalty on the \(k\)th order differences leads to an estimator that can adapt to the number of jumps in the \((k - 1)\)th order differences of the underlying signal or an approximation thereof. We show the result for \(k \in \{1, 2, 3, 4\}\) and indicate how it could be derived for general \(k \in \mathbb{N}\).

REFERENCES


Key words and phrases. Oracle inequality, projection, compatibility, lasso, analysis, total variation regularization, minimax, Moore–Penrose pseudo inverse.


RECONCILING THE GAUSSIAN AND WHITTLE LIKELIHOOD WITH AN APPLICATION TO ESTIMATION IN THE FREQUENCY DOMAIN

BY SUHASINI SUBBA RAO1 AND JUNHO YANG2

1Department of Statistics, Texas A&M University, suhasini@stat.tamu.edu
2Institute of Statistical Science, Academia Sinica, junhoyang@stat.sinica.edu.tw

In time series analysis there is an apparent dichotomy between time and frequency domain methods. The aim of this paper is to draw connections between frequency and time domain methods. Our focus will be on reconciling the Gaussian likelihood and the Whittle likelihood. We derive an exact, interpretable, bound between the Gaussian and Whittle likelihood of a second order stationary time series. The derivation is based on obtaining the transformation which is biorthogonal to the discrete Fourier transform of the time series. Such a transformation yields a new decomposition for the inverse of a Toeplitz matrix and enables the representation of the Gaussian likelihood within the frequency domain. We show that the difference between the Gaussian and Whittle likelihood is due to the omission of the best linear predictions outside the domain of observation in the periodogram associated with the Whittle likelihood. Based on this result, we obtain an approximation for the difference between the Gaussian and Whittle likelihoods in terms of the best fitting, finite order autoregressive parameters. These approximations are used to define two new frequency domain quasi-likelihood criteria. We show that these new criteria can yield a better approximation of the spectral divergence criterion, as compared to both the Gaussian and Whittle likelihoods. In simulations, we show that the proposed estimators have satisfactory finite sample properties.

REFERENCES


Key words and phrases. Biorthogonal transforms, discrete Fourier transform, periodogram, quasi-likelihoods, second order stationary time series, Toeplitz inverse.


INTEGRATIVE METHODS FOR POST-SELECTION INFERENCE UNDER CONVEX CONSTRAINTS

BY SNIGDHA PANIGRAHI1, JONATHAN TAYLOR2 AND ASAF WEINSTEIN3

1Department of Statistics, University of Michigan, psnigdha@umich.edu
2Department of Statistics, Stanford University, jtaylo@stanford.edu
3Department of Statistics, Hebrew University of Jerusalem, asaf.weinstein@mail.huji.ac.il

Inference after model selection has been an active research topic in the past few years, with numerous works offering different approaches to addressing the perils of the reuse of data. In particular, major progress has been made recently on large and useful classes of problems by harnessing general theory of hypothesis testing in exponential families, but these methods have their limitations. Perhaps most immediate is the gap between theory and practice: implementing the exact theoretical prescription in realistic situations—for example, when new data arrives and inference needs to be adjusted accordingly—may be a prohibitive task. In this paper, we propose a Bayesian framework for carrying out inference after variable selection. Our framework is very flexible in the sense that it naturally accommodates different models for the data instead of requiring a case-by-case treatment. This flexibility is achieved by considering the full selective likelihood function where, crucially, we propose a novel and nontrivial approximation to the exact but intractable expression. The advantages of our methods in practical data analysis are demonstrated in an application to HIV drug-resistance data.

REFERENCES


Key words and phrases. Adaptive data analysis, Bayesian inference, carving, conditional inference, convex constraints, selective inference.


Privacy-preserving data analysis is a rising challenge in contemporary statistics, as the privacy guarantees of statistical methods are often achieved at the expense of accuracy. In this paper, we investigate the tradeoff between statistical accuracy and privacy in mean estimation and linear regression, under both the classical low-dimensional and modern high-dimensional settings. A primary focus is to establish minimax optimality for statistical estimation with the \((\varepsilon, \delta)\)-differential privacy constraint. By refining the “tracing adversary” technique for lower bounds in the theoretical computer science literature, we improve existing minimax lower bound for low-dimensional mean estimation and establish new lower bounds for high-dimensional mean estimation and linear regression problems. We also design differentially private algorithms that attain the minimax lower bounds up to logarithmic factors. In particular, for high-dimensional linear regression, a novel private iterative hard thresholding algorithm is proposed. The numerical performance of differentially private algorithms is demonstrated by simulation studies and applications to real data sets.

REFERENCES


DISTRIBUTED STATISTICAL INFERENCE FOR MASSIVE DATA

BY SONG XI CHEN¹ AND LIUHUA PENG²

¹Guanghua School of Management and Center for Statistical Science, Peking University, songxichen@pku.edu.cn
²School of Mathematics and Statistics, University of Melbourne, liuhua.peng@unimelb.edu.au

This paper considers distributed statistical inference for general symmetric statistics in the context of massive data with efficient computation. Estimation efficiency and asymptotic distributions of the distributed statistics are provided, which reveal different results between the nondegenerate and degenerate cases, and show the number of the data subsets plays an important role. Two distributed bootstrap methods are proposed and analyzed to approximation the underlying distribution of the distributed statistics with improved computation efficiency over existing methods. The accuracy of the distributional approximation by the bootstrap are studied theoretically. One of the methods, the pseudo-distributed bootstrap, is particularly attractive if the number of datasets is large as it directly resamples the subset-based statistics, assumes less stringent conditions and its performance can be improved by studentization.

REFERENCES


Key words and phrases. Distributed bootstrap, distributed statistics, massive data, pseudo-distributed bootstrap.


CONSTRUCTION OF MIXED ORTHOGONAL ARRAYS WITH HIGH STRENGTH

BY SHANQI PANG\(^1\), JING WANG\(^2\), DENNIS K. J. LIN\(^3\) AND MIN-QIAN LIU\(^4\)

\(^1\)College of Mathematics and Information Science, Henan Normal University, pangshanqi@263.net
\(^2\)College of Mathematics and Science, Shanghai Normal University, wangjing19@126.com
\(^3\)Department of Statistics, Purdue University, dkjlin@purdue.edu
\(^4\)School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, mqliu@nankai.edu.cn

A considerable portion of the work on mixed orthogonal arrays applies specifically to arrays of strength 2. Although strength \(t = 2\) is arguably the most important case for statistical applications, there is an urgent need for better methods for \(t \geq 3\). However, the knowledge on the existence of arrays for \(t \geq 3\) is rather limited. In this paper, new construction methods for symmetric and asymmetric orthogonal arrays (OAs) with high strength are proposed by using lower strength orthogonal partitions of spaces and OAs. A positive answer is provided to the open problem in Hedayat, Sloane and Stufken (Orthogonal Arrays: Theory and Applications (1999) Springer) on developing better methods and tools for the construction of mixed orthogonal arrays with strength \(t \geq 3\). Not only are the methods straightforward, but also they are useful for constructing symmetric or asymmetric OAs of arbitrary strengths, numbers of levels and various sizes. The constructed OAs can be utilized to generate more OAs. The resulting OAs have a high degree of flexibility and many other desirable properties. Some selective OAs are tabulated for practical uses.

REFERENCES


MSC2020 subject classifications. Primary 62K15; secondary 05B15.

Key words and phrases. High strength orthogonal arrays, Kronecker product, orthogonal partitions.


FOUNDATIONS OF STRUCTURAL CAUSAL MODELS WITH CYCLES AND LATENT VARIABLES

BY STEPHAN BONGERS¹,*, PATRICK FORRÉ¹,†, JONAS PETERS² AND JORIS M. MOOIJ³

¹Informatics Institute, University of Amsterdam, *s.r.bongers@uva.nl; †p.d.forre@uva.nl
²Department of Mathematical Sciences, University of Copenhagen, jonas.peters@math.ku.dk
³Korteweg–De Vries Institute, University of Amsterdam, j.m.mooij@uva.nl

Structural causal models (SCMs), also known as (nonparametric) structural equation models (SEMs), are widely used for causal modeling purposes. In particular, acyclic SCMs, also known as recursive SEMs, form a well-studied subclass of SCMs that generalize causal Bayesian networks to allow for latent confounders. In this paper, we investigate SCMs in a more general setting, allowing for the presence of both latent confounders and cycles. We show that in the presence of cycles, many of the convenient properties of acyclic SCMs do not hold in general: they do not always have a solution; they do not always induce unique observational, interventional and counterfactual distributions; a marginalization does not always exist, and if it exists the marginal model does not always respect the latent projection; they do not always satisfy a Markov property; and their graphs are not always consistent with their causal semantics. We prove that for SCMs in general each of these properties does hold under certain solvability conditions. Our work generalizes results for SCMs with cycles that were only known for certain special cases so far. We introduce the class of simple SCMs that extends the class of acyclic SCMs to the cyclic setting, while preserving many of the convenient properties of acyclic SCMs. With this paper, we aim to provide the foundations for a general theory of statistical causal modeling with SCMs.

REFERENCES


MSC2020 subject classifications. Primary 62A09, 68T30; secondary 68T37.
Key words and phrases. Structural causal models, causal graph, cycles, interventions, counterfactuals, solvability, Markov properties, marginalization.


MEASURING DEPENDENCE IN THE WASSERSTEIN DISTANCE FOR
BAYESIAN NONPARAMETRIC MODELS

BY MARTA CATALANO¹, ANTONIO LIJOI²,* AND IGOR PRÜNSTER²,†

¹Department of Statistics, University of Warwick, marta.catalano@warwick.ac.uk
²Department of Decision Sciences, Bocconi University, *antonio.lioi@unibocconi.it; †igor.pruenster@unibocconi.it

The proposal and study of dependent Bayesian nonparametric models has been one of the most active research lines in the last two decades, with random vectors of measures representing a natural and popular tool to define them. Nonetheless, a principled approach to understand and quantify the associated dependence structure is still missing. We devise a general, and not model-specific, framework to achieve this task for random measure based models, which consists in: (a) quantify dependence of a random vector of probabilities in terms of closeness to exchangeability, which corresponds to the maximally dependent coupling with the same marginal distributions, that is, the comonotonic vector; (b) recast the problem in terms of the underlying random measures (in the same Fréchet class) and quantify the closeness to comonotonicity; (c) define a distance based on the Wasserstein metric, which is ideally suited for spaces of measures, to measure the dependence in a principled way. Several results, which represent the very first in the area, are obtained. In particular, useful bounds in terms of the underlying Lévy intensities are derived relying on compound Poisson approximations. These are then specialized to popular models in the Bayesian literature leading to interesting insights.

REFERENCES


MSC2020 subject classifications. 62C10, 60G57, 60G09.
Key words and phrases. Bayesian nonparametrics, completely random measures, completely random vectors, compound Poisson, dependence, independent increments, Lévy copula, Wasserstein distance.


BRIDGING CONVEX AND NONCONVEX OPTIMIZATION IN ROBUST PCA:
NOISE, OUTLIERS AND MISSING DATA

BY YUXIN CHEN1, JIANQING FAN2,*, CONG MA3 AND YULING YAN2,†

1Department of Electrical and Computer Engineering, Princeton University, yuxin.chen@princeton.edu
2Department of Operations Research and Financial Engineering, Princeton University, *jqfan@princeton.edu;
†yulingy@princeton.edu
3Department of Statistics, University of Chicago, congma@uchicago.edu

This paper delivers improved theoretical guarantees for the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis in the presence of random noise) remains highly suboptimal, which we strengthen in this paper. When the unknown matrix is well conditioned, incoherent and of constant rank, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the $\ell_\infty$ loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. The key analysis idea lies in bridging the convex program in use and an auxiliary nonconvex optimization algorithm, and hence the title of this paper.

REFERENCES


MSC2020 subject classifications. Primary 62F10; secondary 62B10.

Key words and phrases. Robust principal component analysis, convex relaxation, $\ell_\infty$ guarantees, leave-one-out analysis.


EFFICIENCY OF DELAYED-ACCEPTANCE RANDOM WALK METROPOLIS ALGORITHMS

BY CHRIS SHERLOCK\(^1\), ALEXANDRE H. THIERY\(^2\) AND ANDREW GOLIGHTLY\(^3\)

\(^1\)Department of Mathematics and Statistics, Lancaster University, c.sherlock@lancaster.ac.uk
\(^2\)Department of Statistics and Applied Probability, National University of Singapore, a.h.thiery@gmail.com
\(^3\)School of Mathematics, Statistics and Physics, Newcastle University, andrew.golightly@ncl.ac.uk

Delayed-acceptance Metropolis–Hastings and delayed-acceptance pseudo-marginal Metropolis–Hastings algorithms can be applied when it is computationally expensive to calculate the true posterior or an unbiased stochastic approximation thereof, but a computationally cheap deterministic approximation is available. An initial accept–reject stage uses the cheap approximation for computing the Metropolis–Hastings ratio; proposals which are accepted at this stage are subjected to a further accept–reject step, which corrects for the error in the approximation. Since the expensive posterior, or the approximation thereof, is only evaluated for proposals which are accepted at the first stage, the cost of the algorithm is reduced and larger scalings may be used.

We focus on the random walk Metropolis (RWM) and consider the delayed-acceptance RWM and the delayed-acceptance pseudo-marginal RWM. We provide a framework for incorporating relatively general deterministic approximations into the theoretical analysis of high-dimensional targets. Justified by diffusion-approximation arguments, we derive expressions for the limiting efficiency and acceptance rates in high-dimensional settings. Finally, these theoretical insights are leveraged to formulate practical guidelines for the efficient tuning of the algorithms. The robustness of these guidelines and predicted properties are verified against simulation studies, all of which are strictly outside of the domain of validity of our limit results.

REFERENCES


MSC2020 subject classifications. 65C05, 65C40, 60F05.

Key words and phrases. Markov chain Monte Carlo, delayed-acceptance, diffusion limit, Pseudo-marginal MCMC, random-walk Metropolis.


SEMIPARAMETRIC OPTIMAL ESTIMATION WITH NONIGNORABLE NONRESPONSE DATA

BY KOSUKE MORIKAWA\textsuperscript{1} AND JAE KWANG KIM\textsuperscript{2}

\textsuperscript{1}Graduate School of Engineering Science, Osaka University, morikawa@sigmath.es.osaka-u.ac.jp
\textsuperscript{2}Department of Statistics, Iowa State University, jkim@iastate.edu

When the response mechanism is believed to be not missing at random (NMAR), a valid analysis requires stronger assumptions on the response mechanism than standard statistical methods would otherwise require. Semiparametric estimators have been developed under the parametric model assumptions on the response mechanism. In this paper, a new statistical test is proposed to guarantee model identifiability without using instrumental variable assumption. Furthermore, we develop optimal semiparametric estimation for parameters such as the population mean. Specifically, we propose two semiparametric optimal estimators that do not require any model assumptions other than the response mechanism. Asymptotic properties of the proposed estimators are discussed. An extensive simulation study is presented to compare with some existing methods. We present an application of our method using Korean labor and income panel survey data.

REFERENCES


Key words and phrases. Estimating functions, identification, incomplete data, not missing at random (NMAR), semiparametric efficient estimation.


TWO-LEVEL PARALLEL FLATS DESIGNS

BY CHUNYAN WANG\textsuperscript{1} AND ROBERT W. MEE\textsuperscript{2}

\textsuperscript{1}School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, 2120150092@mail.nankai.edu.cn
\textsuperscript{2}Department of Business Analytics and Statistics, University of Tennessee, rmee@utk.edu

Regular $2^n - p$ designs are also known as single flat designs. Parallel flats designs (PFDs) consisting of three parallel flats (3-PFDs) are the most frequently utilized PFDs, due to their simple structure. Generalizing to $f$-PFD with $f > 3$ is more challenging. This paper aims to study the general theory for the $f$-PFD for any $f \geq 3$. We propose a method for obtaining the confounding frequency vectors for all nonequivalent $f$-PFDs, and to find the least $G$-aberration (or highest D-efficiency) $f$-PFD constructed from any single flat. PFDs are particularly useful for constructing nonregular fraction, split-plot or randomized block designs. We also characterize the quaternary code design series as PFDs. Finally, we show how designs constructed by concatenating regular fractions from different families may also have a parallel flats structure. Examples are given throughout to illustrate the results.

REFERENCES


MSC2020 subject classifications. Primary 62K15; secondary 62K05.

Key words and phrases. Confounding frequency vector, equivalent design, $G$-aberration, nonregular fractional factorial design, quaternary code, Sylvester Hadamard matrix, two-factor interaction model.
BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE