Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss and Simpson’s Paradox
Ruobin Gong and Xiao-Li Meng

A Unified Primal Dual Active Set Algorithm for Nonconvex Sparse Recovery
Jian Huang, Yuling Jiao, Bangti Jin, Jin Liu, Xiliang Lu and Can Yang

The Box–Cox Transformation: Review and Extensions
Anthony C. Atkinson, Marco Riani and Aldo Corbellini

Noncommutative Probability and Multiplicative Cascades
Ian W. McKeague

A Selective Overview of Deep Learning
Jianqing Fan, Cong Ma and Yiqiao Zhong

Stochastic Approximation: From Statistical Origin to Big-Data, Multidisciplinary Applications
Tze Leung Lai and Hongsong Yuan

Robust High-Dimensional Factor Models with Applications to Statistical Machine Learning
Jianqing Fan, Kaizheng Wang, Yiqiao Zhong and Ziwei Zhu

A Conversation with Dennis Cook
Efstathia Bura, Bing Li, Lexin Li, Christopher Nachtsheim, Daniel Pena, Claude Setodji and Robert E. Weiss
EDITOR
Sonia Petrone
Università Bocconi

DEPUTY EDITOR
Peter Green
University of Bristol and University of Technology Sydney

ASSOCIATE EDITORS
Shankar Bhamidi
University of North Carolina
Jay Breidt
Colorado State University
Peter Bühlmann
ETH Zürich
John Chambers
Stanford University
Michael J. Daniels
University of Florida
Philip Dawid
University of Cambridge
Robin Evans
University of Oxford
Stefano Favaro
Università di Torino
Alan Gelfand
Duke University
Peter Green
University of Bristol and University of Technology Sydney

Chris Holmes
University of Oxford
Susan Holmes
Stanford University
Tailen Hsing
University of Michigan
Michael I. Jordan
University of California, Berkeley
Ian McKeague
Columbia University
Peter Müller
University of Texas
Luc Pronzato
Université Nice
Nancy Reid
University of Toronto
Jason Roy
Rutgers University
Chiara Sabatti
Stanford University

Bodhisattva Sen
Columbia University
Glenn Shafer
Rutgers Business
School–Newark and New Brunswick
Royal Holloway College, University of London
Michael Stein
University of Chicago
Jon Wellner
University of Washington
Bin Yu
University of California, Berkeley
Giacomo Zanella
Università Bocconi
Cun-Hui Zhang
Rutgers University

MANAGING EDITOR
Robert Keener
University of Michigan

PRODUCTION EDITOR
Patrick Kelly

EDITORIAL COORDINATOR
Kristina Mattson

PAST EXECUTIVE EDITORS
Morris Eaton, 2001
George Casella, 2002–2004
David Madigan, 2008–2010
Leon J. Gleser, 1998–2000
Jon A. Wellner, 2011–2013
Richard Tweedie, 2001
Peter Green, 2014–2016
Cun-Hui Zhang, 2017–2019
Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss and Simpson’s Paradox

Ruobin Gong and Xiao-Li Meng

Abstract. Imprecise probabilities alleviate the need for high-resolution and unwarranted assumptions in statistical modeling. They present an alternative strategy to reduce irreproducible findings. However, updating imprecise models requires the user to choose among alternative updating rules. Competing rules can result in incompatible inferences, and exhibit dilation, contraction and sure loss, unsettling phenomena that cannot occur with precise probabilities and the regular Bayes rule. We revisit some famous statistical paradoxes and show that the logical fallacy stems from a set of marginally plausible yet jointly incommensurable model assumptions akin to the trio of phenomena above. Discrepancies between the generalized Bayes (\(B\)) rule, Dempster’s (\(D\)) rule and the Geometric (\(G\)) rule as competing updating rules for Choquet capacities of order 2 are discussed. We note that (1) \(B\)-rule cannot contract nor induce sure loss, but is the most prone to dilation due to “overfitting” in a certain sense; (2) in absence of prior information, both \(B\)-and \(G\)-rules are incapable to learn from data however informative they may be; (3) \(D\)- and \(G\)-rules can mathematically contradict each other by contracting while the other dilating. These findings highlight the invaluable role of judicious judgment in handling low-resolution information, and the care that needs to be take when applying updating rules to imprecise probability models.

Key words and phrases: Imprecise probability, model uncertainty, Choquet capacity, belief function, coherence, Monty Hall problem.

REFERENCES


Comment: On the History and Limitations of Probability Updating

Glenn Shafer

Abstract. Gong and Meng show that we can gain insights into classical paradoxes about conditional probability by acknowledging that apparently precise probabilities live within a larger world of imprecise probability. They also show that the notion of updating becomes problematic in this larger world. A closer look at the historical development of the notion of updating can give us further insights into its limitations.

Key words and phrases: Bayes’s rule of conditioning, Dempster’s rule, conditional probability, conditionalization, imprecise probabilities, probability protocols, relative probability, updating.

REFERENCES


naines, Paris.


CZUBER, E. (1908). Wahrscheinlichkeitsrechnung und Ihre Anwen-


178.


Glenn Shafer is University Professor, Rutgers University, Newark, New Jersey (e-mail: gshafer@rutgers.edu).
Comment: Settle the Unsettling: An Inferential Models Perspective

Chuanhai Liu and Ryan Martin

Abstract. Here, we demonstrate that the inferential model (IM) framework, unlike the updating rules that Gong and Meng show to be unreliable, provides valid and efficient inferences/prediction while not being susceptible to sure loss. In this sense, the IM framework settles what Gong and Meng characterized as “unsettling.”

Key words and phrases: Belief function, efficiency, lower and upper probability, inferential models, validity.

REFERENCES


Chuanhai Liu is Professor, Department of Statistics, Purdue University, West Lafayette, Indiana, USA (e-mail: chuanhai@purdue.edu). Ryan Martin is Professor, Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA (e-mail: rgmarti3@ncsu.edu).
Comment: Moving Beyond Sets of Probabilities

Gregory Wheeler

Abstract. The theory of lower previsions is designed around the principles of coherence and sure-loss avoidance, thus steers clear of all the updating anomalies highlighted in Gong and Meng’s “Judicious Judgment Meets Unsettling Updating: Dilation, Sure Loss and Simpson’s Paradox” except dilation. In fact, the traditional problem with the theory of imprecise probability is that coherent inference is too complicated rather than unsettling. Progress has been made simplifying coherent inference by demoting sets of probabilities from fundamental building blocks to secondary representations that are derived or discarded as needed.

Key words and phrases: Desirable gambles, lower previsions, imprecise probabilities, dilation.

REFERENCES


Gregory Wheeler is Professor of Theoretical Philosophy and Computer Science, Frankfurt School of Finance & Management, 60322 Frankfurt am Main, Germany (e-mail: g.wheeler@fs.de).
Comment: On Focusing, Soft and Strong Revision of Choquet Capacities and Their Role in Statistics

Thomas Augustin and Georg Schollmeyer

Abstract. We congratulate Ruobin Gong and Xiao-Li Meng on their thought-provoking paper demonstrating the power of imprecise probabilities in statistics. In particular, Gong and Meng clarify important statistical paradoxes by discussing them in the framework of generalized uncertainty quantification and different conditioning rules used for updating. In this note, we characterize all three conditioning rules as envelopes of certain sets of conditional probabilities. This view also suggests some generalizations that can be seen as compromise rules. Similar to Gong and Meng, our derivations mainly focus on Choquet capacities of order 2, and so we also briefly discuss in general their role as statistical models. We conclude with some general remarks on the potential of imprecise probabilities to cope with the multidimensional nature of uncertainty.

Key words and phrases: Imprecise probabilities, Choquet capacities, updating, neighborhood models, generalized Bayes rule, Dempster’s rule of conditioning.

REFERENCES


Rejoinder: Let’s Be Imprecise in Order to Be Precise (About What We Don’t Know)

Ruobin Gong and Xiao-Li Meng

REFERENCES


Ruobin Gong is Assistant Professor of Statistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, USA (e-mail: rg915@stat.rutgers.edu). Xiao-Li Meng is Whipple V. N. Jones Professor of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA (e-mail: meng@stat.harvard.edu).
A Unified Primal Dual Active Set Algorithm for Nonconvex Sparse Recovery

Jian Huang, Yuling Jiao, Bangti Jin, Jin Liu, Xiliang Lu and Can Yang

Abstract. In this paper, we consider the problem of recovering a sparse signal based on penalized least squares formulations. We develop a novel algorithm of primal-dual active set type for a class of nonconvex sparsity-promoting penalties, including \( \ell_0 \), bridge, smoothly clipped absolute deviation, capped \( \ell_1 \) and minimax concavity penalty. First, we establish the existence of a global minimizer for the related optimization problems. Then we derive a novel necessary optimality condition for the global minimizer using the associated thresholding operator. The solutions to the optimality system are coordinatewise minimizers, and under minor conditions, they are also local minimizers. Upon introducing the dual variable, the active set can be determined using the primal and dual variables together. Further, this relation lends itself to an iterative algorithm of active set type which at each step involves first updating the primal variable only on the active set and then updating the dual variable explicitly. When combined with a continuation strategy on the regularization parameter, the primal dual active set method is shown to converge globally to the underlying regression target under certain regularity conditions. Extensive numerical experiments with both simulated and real data demonstrate its superior performance in terms of computational efficiency and recovery accuracy compared with the existing sparse recovery methods.

Key words and phrases: Nonconvex penalty, sparsity, primal-dual active set algorithm, continuation, consistency.

REFERENCES


Jian Huang is Professor, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242 (e-mail: jian-huang@uiowa.edu). Yuling Jiao is Associate Professor, School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, 430063, P.R. China (e-mail: yulingjiaomath@whu.edu.cn). Bangti Jin is Professor, Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK (e-mail: b.jin@ucl.ac.uk; bangti.jin@gmail.com). Jin Liu is Assistant Professor, Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore (e-mail: jin.liu@duke-nus.edu.sg). Xiliang Lu is Professor, School of Mathematics and Statistics, Wuhan University, Wuhan 430072, P.R. China, and Hubei Key Laboratory of Computational Science (Wuhan University), Wuhan, 430072, China (e-mail: xll.math@whu.edu.cn). Can Yang is Associate Professor, Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (e-mail: macyang@ust.hk).


The Box–Cox Transformation: Review and Extensions

Anthony C. Atkinson, Marco Riani and Aldo Corbellini

Abstract. The Box–Cox power transformation family for nonnegative responses in linear models has a long and interesting history in both statistical practice and theory, which we summarize. The relationship between generalized linear models and log transformed data is illustrated. Extensions investigated include the transform both sides model and the Yeo–Johnson transformation for observations that can be positive or negative. The paper also describes an extended Yeo–Johnson transformation that allows positive and negative responses to have different power transformations. Analyses of data show this to be necessary. Robustness enters in the fan plot for which the forward search provides an ordering of the data. Plausible transformations are checked with an extended fan plot. These procedures are used to compare parametric power transformations with nonparametric transformations produced by smoothing.

Key words and phrases: ACE, AVAS, constructed variable, extended Yeo–Johnson transformation, forward search, linked plots, robust methods.

REFERENCES


Anthony C. Atkinson is Emeritus Professor, Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom (e-mail: a.c.atkinson@lse.ac.uk). Marco Riani is Full Professor, Dipartimento di Scienze Economiche e Aziendale and Interdepartmental Centre for Robust Statistics, Università di Parma, 43100 Parma, Italy (e-mail: mriani@unipr.it). Aldo Corbellini is Senior Lecturer, Dipartimento di Scienze Economiche e Aziendale and Interdepartmental Centre for Robust Statistics, Università di Parma, 43100 Parma, Italy (e-mail: aldo.corbellini@unipr.it).


Noncommutative Probability and Multiplicative Cascades

Ian W. McKeague

Abstract. Various aspects of standard model particle physics might be explained by a suitably rich algebra acting on itself, as suggested by Furey (2015). The present paper develops the asymptotics of large causal tree diagrams that combine freely independent elements in such an algebra. The Marčenko–Pastur law and Wigner’s semicircle law are shown to emerge as limits of normalized sum-over-paths of nonnegative elements assigned to the edges of causal trees. These results are established in the setting of noncommutative probability. Trees with classically independent positive edge weights (random multiplicative cascades) were originally proposed by Mandelbrot as a model displaying the fractal features of turbulence. The novelty of the present work is the use of noncommutative (free) probability to allow the edge weights to take values in an algebra. An application to theoretical neuroscience is also discussed.

Key words and phrases: Mandelbrot cascade, Marčenko–Pastur law, martingale convergence, random matrices, Wigner’s semicircle law.

REFERENCES


A Selective Overview of Deep Learning

Jianqing Fan, Cong Ma and Yiqiao Zhong

Abstract. Deep learning has achieved tremendous success in recent years. In simple words, deep learning uses the composition of many nonlinear functions to model the complex dependency between input features and labels. While neural networks have a long history, recent advances have significantly improved their empirical performance in computer vision, natural language processing and other predictive tasks. From the statistical and scientific perspective, it is natural to ask: What is deep learning? What are the new characteristics of deep learning, compared with classical statistical methods? What are the theoretical foundations of deep learning?

To answer these questions, we introduce common neural network models (e.g., convolutional neural nets, recurrent neural nets, generative adversarial nets) and training techniques (e.g., stochastic gradient descent, dropout, batch normalization) from a statistical point of view. Along the way, we highlight new characteristics of deep learning (including depth and overparametrization) and explain their practical and theoretical benefits. We also sample recent results on theories of deep learning, many of which are only suggestive. While a complete understanding of deep learning remains elusive, we hope that our perspectives and discussions serve as a stimulus for new statistical research.

Key words and phrases: Neural networks, overparametrization, stochastic gradient descent, approximation theory, generalization error.

REFERENCES


Stochastic Approximation: From Statistical Origin to Big-Data, Multidisciplinary Applications

Tze Leung Lai and Hongsong Yuan

Abstract. Stochastic approximation was introduced in 1951 to provide a new theoretical framework for root finding and optimization of a regression function in the then-nascent field of statistics. This review shows how it has evolved in response to other developments in statistics, notably time series and sequential analysis, and to applications in artificial intelligence, economics and engineering. Its resurgence in the big data era has led to new advances in both theory and applications of this microcosm of statistics and data science.

Key words and phrases: Control, gradient boosting, optimization, recursive stochastic algorithms, regret, weak greedy variable selection.

REFERENCES


Tze Leung Lai is Professor, Department of Statistics, Stanford University, 390 Jane Stanford Way, Stanford, California 94305, USA (e-mail: lai@stanford.edu). Hongsong Yuan is Associate Professor, School of Information Management and Engineering, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China (e-mail: yuan.hongsong@shufe.edu.cn).


Robust High-Dimensional Factor Models with Applications to Statistical Machine Learning

Jianqing Fan, Kaizheng Wang, Yiqiao Zhong and Ziwei Zhu

Abstract. Factor models are a class of powerful statistical models that have been widely used to deal with dependent measurements that arise frequently from various applications from genomics and neuroscience to economics and finance. As data are collected at an ever-growing scale, statistical machine learning faces some new challenges: high dimensionality, strong dependence among observed variables, heavy-tailed variables and heterogeneity. High-dimensional robust factor analysis serves as a powerful toolkit to conquer these challenges.

This paper gives a selective overview on recent advance on high-dimensional factor models and their applications to statistics including Factor-Adjusted Robust Model selection (FarmSelect) and Factor-Adjusted Robust Multiple testing (FarmTest). We show that classical methods, especially principal component analysis (PCA), can be tailored to many new problems and provide powerful tools for statistical estimation and inference. We highlight PCA and its connections to matrix perturbation theory, robust statistics, random projection, false discovery rate, etc., and illustrate through several applications how insights from these fields yield solutions to modern challenges. We also present far-reaching connections between factor models and popular statistical learning problems, including network analysis and low-rank matrix recovery.

Key words and phrases: Factor model, PCA, covariance estimation, perturbation bounds, robustness, random sketch, FarmSelect, FarmTest.

REFERENCES


A Conversation with Dennis Cook

Efstathia Bura, Bing Li, Lexin Li, Christopher Nachtsheim, Daniel Pena, Claude Setodji and Robert E. Weiss

Abstract. Dennis Cook is a Full Professor, School of Statistics, at the University of Minnesota. He received his BS degree in Mathematics from Northern Montana College, and MS and PhD degrees in Statistics from Kansas State University. He has served as Chair of the Department of Applied Statistics, Director of the Statistical Center and Director of the School of Statistics, all at the University of Minnesota.

His research areas include dimension reduction, linear and nonlinear regression, experimental design, statistical diagnostics, statistical graphics and population genetics. He has authored over 200 research articles and is author or co-author of two textbooks—An Introduction to Regression Graphics and Applied Regression Including Computing and Graphics—and three research monographs, Influence and Residuals in Regression, Regression Graphics: Ideas for Studying Regressions through Graphics and An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics.

He has served as Associate Editor of the Journal of the American Statistical Association, The Journal of Quality Technology, Biometrika, Journal of the Royal Statistical Society and Statistica Sinica. He is a four-time recipient of the Jack Youden Prize for Best Expository Paper in Technometrics as well as the Frank Wilcoxon Award for Best Technical Paper. He received the 2005 COPSS Fisher Lecture and Award, and he is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics. The following conversation took place on March 22, 2019, following the banquet at the conference, “Cook’s Distance and Beyond: A Conference Celebrating the Contributions of R. Dennis Cook.” The interviewers were, Efstathia Bura (Effie), Bing Li, Lexin Li, Christopher Nachtsheim (Chris), Daniel Pena, Claude Messan Setodji and Robert Weiss (Rob).

Key words and phrases: Bayesian methods, central subspace, Cook’s distance, envelope methods, influence, neural networks, optimal experimental design, regression diagnostics, sufficient dimension reduction.
INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

______________________________

IMS OFFICERS

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

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

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

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

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

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

IMS EDITORS

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

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

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


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

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

COMPUTER AGE
STATISTICAL
INERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE