<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Paradox from Randomization-Based Causal Inference</td>
<td>Peng Ding</td>
<td>331</td>
</tr>
<tr>
<td>Understanding Ding’s Apparent Paradox</td>
<td>Peter M. Aronow and Molly R. Offer-Westort</td>
<td>346</td>
</tr>
<tr>
<td>Randomization-Based Tests for “No Treatment Effects”</td>
<td>EunYi Chung</td>
<td>349</td>
</tr>
<tr>
<td>Inference from Randomized (Factorial) Experiments</td>
<td>R. A. Bailey</td>
<td>352</td>
</tr>
<tr>
<td>An Apparent Paradox Explained</td>
<td>Wen Wei Loh, Thomas S. Richardson and James M. Robins</td>
<td>356</td>
</tr>
<tr>
<td>Rejoinder: A Paradox from Randomization-Based Causal Inference</td>
<td>Peng Ding</td>
<td>362</td>
</tr>
<tr>
<td>Logistic Regression: From Art to Science</td>
<td>Dimitris Bertsimas and Angela King</td>
<td>367</td>
</tr>
<tr>
<td>Principles of Experimental Design for Big Data Analysis</td>
<td>Christopher C. Drovandi, Christopher C. Holmes, James M. McGree, Kerrie Mengersen, Sylvia Richardson and Elizabeth G. Ryan</td>
<td>385</td>
</tr>
<tr>
<td>Importance Sampling: Intrinsic Dimension and Computational Cost</td>
<td>S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A. M. Stuart</td>
<td>405</td>
</tr>
<tr>
<td>Estimation of Causal Effects with Multiple Treatments: A Review and New Ideas</td>
<td>Michael J. Lopez and Roee Gutman</td>
<td>432</td>
</tr>
<tr>
<td>Models for the Assessment of Treatment Improvement: The Ideal and the Feasible</td>
<td>P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos and C. Matrán</td>
<td>469</td>
</tr>
</tbody>
</table>
A Paradox from Randomization-Based Causal Inference

Peng Ding

Abstract. Under the potential outcomes framework, causal effects are defined as comparisons between potential outcomes under treatment and control. To infer causal effects from randomized experiments, Neyman proposed to test the null hypothesis of zero average causal effect (Neyman’s null), and Fisher proposed to test the null hypothesis of zero individual causal effect (Fisher’s null). Although the subtle difference between Neyman’s null and Fisher’s null has caused a lot of controversies and confusions for both theoretical and practical statisticians, a careful comparison between the two approaches has been lacking in the literature for more than eighty years. We fill this historical gap by making a theoretical comparison between them and highlighting an intriguing paradox that has not been recognized by previous researchers. Logically, Fisher’s null implies Neyman’s null. It is therefore surprising that, in actual completely randomized experiments, rejection of Neyman’s null does not imply rejection of Fisher’s null for many realistic situations, including the case with constant causal effect. Furthermore, we show that this paradox also exists in other commonly-used experiments, such as stratified experiments, matched-pair experiments and factorial experiments. Asymptotic analyses, numerical examples and real data examples all support this surprising phenomenon. Besides its historical and theoretical importance, this paradox also leads to useful practical implications for modern researchers.

Key words and phrases: Average null hypothesis, Fisher randomization test, potential outcome, randomized experiment, repeated sampling property, sharp null hypothesis.

REFERENCES


DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential out-


Understanding Ding’s Apparent Paradox

Peter M. Aronow and Molly R. Offer-Westort

REFERENCES


Peter M. Aronow is Associate Professor, Departments of Political Science and Biostatistics, Yale University, 77 Prospect Street, New Haven, Connecticut 06520, USA (e-mail: peter.aronow@yale.edu). Molly R. Offer-Westort is a graduate student, Departments of Political Science and Statistics and Data Science, Yale University, 77 Prospect Street, New Haven, Connecticut 06520, USA (e-mail: molly.offer-westort@yale.edu).
Randomization-Based Tests for “No Treatment Effects”

EunYi Chung

Abstract. Although both Fisher’s and Neyman’s tests are for testing “no treatment effects,” they both test fundamentally different null hypotheses. While Neyman’s null concerns the average casual effect, Fisher’s null focuses on the individual causal effect. When conducting a test, researchers need to understand what is really being tested and what underlying assumptions are being made. If these fundamental issues are not fully appreciated, dubious conclusions regarding causal effects can be made.

Key words and phrases: Fisher’s randomization test, Neyman’s randomization test, treatment effect, Wilcoxon–Mann–Whitney rank sum test.

REFERENCES


EunYi Chung is Assistant Professor of Economics, Department of Economics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA (e-mail: eunyi@illinois.edu).
Inference from Randomized (Factorial) Experiments

R. A. Bailey

Abstract. This is a contribution to the discussion of the interesting paper by Ding [Statist. Sci. 32 (2017) 331–345], which contrasts approaches attributed to Neyman and Fisher. I believe that Fisher’s usual assumption was unit-treatment additivity, rather than the “sharp null hypothesis” attributed to him. Fisher also developed the notion of interaction in factorial experiments. His explanation leads directly to the concept of marginality, which is essential for the interpretation of data from any factorial experiment.

Key words and phrases: Factorial design, marginality, randomisation, unit-treatment additivity.

REFERENCES

NELDER, J. A. (1965). The analysis of randomised experiments with orthogonal block structure. II. Treatment structure and the...


An Apparent Paradox Explained

Wen Wei Loh, Thomas S. Richardson and James M. Robins

REFERENCES


Wen Wei Loh is Postdoctoral Research Associate, Department of Biostatistics, University of North Carolina, CB #7420, Chapel Hill, North Carolina 27599, USA (e-mail: wloh@u.washington.edu). Thomas S. Richardson is Professor and Chair, Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195, USA (e-mail: thomasr@u.washington.edu). James M. Robins is Mitchell L. and Robin LaFoley Dong Professor of Epidemiology, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA (e-mail: robins@hsph.harvard.edu).
Rejoinder: A Paradox from Randomization-Based Causal Inference

Peng Ding

REFERENCES


Peng Ding is Assistant Professor, Department of Statistics, University of California, Berkeley, 425 Evans Hall, Berkeley, California 94720, USA (e-mail: pengdingpku@berkeley.edu).
Abstract. A high quality logistic regression model contains various desirable properties: predictive power, interpretability, significance, robustness to error in data and sparsity, among others. To achieve these competing goals, modelers incorporate these properties iteratively as they hone in on a final model. In the period 1991–2015, algorithmic advances in Mixed-Integer Linear Optimization (MILO) coupled with hardware improvements have resulted in an astonishing 450 billion factor speedup in solving MILO problems. Motivated by this speedup, we propose modeling logistic regression problems algorithmically with a mixed integer nonlinear optimization (MINLO) approach in order to explicitly incorporate these properties in a joint, rather than sequential, fashion. The resulting MINLO is flexible and can be adjusted based on the needs of the modeler. Using both real and synthetic data, we demonstrate that the overall approach is generally applicable and provides high quality solutions in realistic timelines as well as a guarantee of suboptimality. When the MINLO is infeasible, we obtain a guarantee that imposing distinct statistical properties is simply not feasible.

Key words and phrases: Logistic regression, computational statistics, mixed integer nonlinear optimization.

REFERENCES


Dimitris Bertsimas is Professor, Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (e-mail: dbertsim@mit.edu). Angela King is Data Scientist, End to End Analytics, San Francisco, California, USA (e-mail: aking10@mit.edu).
Principles of Experimental Design for Big Data Analysis

Christopher C. Drovandi, Christopher C. Holmes, James M. McGree, Kerrie Mengersen, Sylvia Richardson and Elizabeth G. Ryan

Abstract. Big Datasets are endemic, but are often notoriously difficult to analyse because of their size, heterogeneity and quality. The purpose of this paper is to open a discourse on the potential for modern decision theoretic optimal experimental design methods, which by their very nature have traditionally been applied prospectively, to improve the analysis of Big Data through retrospective designed sampling in order to answer particular questions of interest. By appealing to a range of examples, it is suggested that this perspective on Big Data modelling and analysis has the potential for wide generality and advantageous inferential and computational properties. We highlight current hurdles and open research questions surrounding efficient computational optimisation in using retrospective designs, and in part this paper is a call to the optimisation and experimental design communities to work together in the field of Big Data analysis.

Key words and phrases: Active learning, Big Data, dimension reduction, experimental design, sub-sampling.

REFERENCES


Christopher C. Drovandi is Senior Lecturer, James M. McGree is Associate Professor, Kerrie Mengersen is Professor, School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, 4000 (e-mail: c.drovandi@qut.edu.au, james.mcgree@qut.edu.au, k.mengersen@qut.edu.au). Christopher C. Holmes is Professor of Biostatistics, Department of Statistics, and Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom, OX1 3TG (e-mail: cholmes@stats.ox.ac.uk). Sylvia Richardson is Professor, MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, United Kingdom, CB2 0SR (e-mail: sylvia.richardson@mrc-bsu.cam.ac.uk). Elizabeth G. Ryan is Statistician, Biostatistics & Health Informatics Department, King’s College London, United Kingdom, SE5 8AF (e-mail: elizabeth.ryan@kcl.ac.uk).


Importance Sampling: Intrinsic Dimension and Computational Cost

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A. M. Stuart

Abstract. The basic idea of importance sampling is to use independent samples from a proposal measure in order to approximate expectations with respect to a target measure. It is key to understand how many samples are required in order to guarantee accurate approximations. Intuitively, some notion of distance between the target and the proposal should determine the computational cost of the method. A major challenge is to quantify this distance in terms of parameters or statistics that are pertinent for the practitioner. The subject has attracted substantial interest from within a variety of communities. The objective of this paper is to overview and unify the resulting literature by creating an overarching framework. A general theory is presented, with a focus on the use of importance sampling in Bayesian inverse problems and filtering.

Key words and phrases: Importance sampling, notions of dimension, small noise, absolute continuity, inverse problems, filtering.

REFERENCES


S. Agapiou is Lecturer, Department of Mathematics and Statistics, University of Cyprus, 1 University Avenue, 2109 Nicosia, Cyprus (e-mail: agapiou.sergios@ucy.ac.cy). O. Papaspiliopoulos is ICREA Research Professor, ICREA & Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain (e-mail: omiros.papaspiliopoulos@upf.edu). D. Sanz-Alonso is Postdoctoral Research Associate, Division of Applied Mathematics, Brown University, Providence, Rhode Island 02906, USA (e-mail: daniel_sanz-alonso1@brown.edu). A. M. Stuart is Bren Professor of Computing and Mathematical Sciences, Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California 91125, USA (e-mail: astuart@caltech.edu).


Estimation of Causal Effects with Multiple Treatments: A Review and New Ideas

Michael J. Lopez and Roee Gutman

Abstract. The propensity score is a common tool for estimating the causal effect of a binary treatment in observational data. In this setting, matching, subclassification, imputation or inverse probability weighting on the propensity score can reduce the initial covariate bias between the treatment and control groups. With more than two treatment options, however, estimation of causal effects requires additional assumptions and techniques, the implementations of which have varied across disciplines. This paper reviews current methods, and it identifies and contrasts the treatment effects that each one estimates. Additionally, we propose possible matching techniques for use with multiple, nominal categorical treatments, and use simulations to show how such algorithms can yield improved covariate similarity between those in the matched sets, relative the pre-matched cohort. To sum, this manuscript provides a synopsis of how to notate and use causal methods for categorical treatments.

Key words and phrases: Causal inference, propensity score, multiple treatments, matching, observational data.

REFERENCES


Michael J. Lopez is Assistant Professor of Statistics, Department of Mathematics, Skidmore College, 815 N. Broadway, Saratoga Springs, New York 12866, USA (e-mail: mlopez1@skidmore.edu). Roee Gutman is Assistant Professor of Biostatistics, Department of Biostatistics, Brown University, 121 S. Main Street, Providence, Rhode Island 02912, USA (e-mail: roee_gutman@brown.edu).
treatment effects. *Biometrika* **96** 187–199. MR2482144


On the Choice of Difference Sequence in a Unified Framework for Variance Estimation in Nonparametric Regression

Wenlin Dai, Tiejun Tong and Lixing Zhu

Abstract. Difference-based methods do not require estimating the mean function in nonparametric regression and are therefore popular in practice. In this paper, we propose a unified framework for variance estimation that combines the linear regression method with the higher-order difference estimators systematically. The unified framework has greatly enriched the existing literature on variance estimation that includes most existing estimators as special cases. More importantly, the unified framework has also provided a smart way to solve the challenging difference sequence selection problem that remains a long-standing controversial issue in nonparametric regression for several decades. Using both theory and simulations, we recommend to use the ordinary difference sequence in the unified framework, no matter if the sample size is small or if the signal-to-noise ratio is large. Finally, to cater for the demands of the application, we have developed a unified R package, named VarED, that integrates the existing difference-based estimators and the unified estimators in nonparametric regression and have made it freely available in the R statistical program http://cran.r-project.org/web/packages/.

Key words and phrases: Difference-based estimator, nonparametric regression, optimal difference sequence, ordinary difference sequence, residual variance.

REFERENCES


Wenlin Dai is Postdoctoral Fellow, CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia (e-mail: wenlin.dai@kaust.edu.sa). Tiejun Tong is Associate Professor and Corresponding Author, Department of Mathematics, Hong Kong Baptist University, Hong Kong (e-mail: tongt@hkbu.edu.hk). Lixing Zhu is Chair Professor, Department of Mathematics, Hong Kong Baptist University, Hong Kong (e-mail: lzhu@hkbu.edu.hk).
Models for the Assessment of Treatment Improvement: The Ideal and the Feasible

P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos and C. Matrán

Abstract. Comparisons of different treatments or production processes are the goals of a significant fraction of applied research. Unsurprisingly, two-sample problems play a main role in statistics through natural questions such as “Is the the new treatment significantly better than the old?” However, this is only partially answered by some of the usual statistical tools for this task. More importantly, often practitioners are not aware of the real meaning behind these statistical procedures. We analyze these troubles from the point of view of the order between distributions, the stochastic order, showing evidence of the limitations of the usual approaches, paying special attention to the classical comparison of means under the normal model. We discuss the unfeasibility of statistically proving stochastic dominance, but show that it is possible, instead, to gather statistical evidence to conclude that slightly relaxed versions of stochastic dominance hold.

Key words and phrases: Stochastic dominance, similarity, two-sample comparison, trimmed distributions, winsorized distributions, Behrens–Fisher problem, index of stochastic dominance.

REFERENCES


P. C. Álvarez-Esteban is Professor, Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid, Paseo de Belén, 7, 47011-Valladolid, Spain (e-mail: pedro@eio.uva.es). E. del Barrio is Professor, Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid, Paseo de Belén, 7, 47011-Valladolid, Spain (e-mail: tasio@eio.uva.es). J. A. Cuesta-Albertos is Professor, Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Avda. de los Castros, 48, 39005-Santander, Spain (e-mail: cuestaj@unican.es). C. Matrán is Professor, Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid, Paseo de Belén, 7, 47011-Valladolid, Spain (e-mail: carlos.matran@uva.es).


