

STATISTICAL SCIENCE

Volume 38, Number 2

May 2023

Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations <i>David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofia S. Villar</i>	185
Comment: A Quarter Century of Methodological Research in Response-Adaptive Randomization <i>Anastasia Ivanova and William F. Rosenberger</i>	209
Comment: Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations <i>Yunshan Duan, Peter Müller and Yuan Ji</i>	212
Comment: Advancing Clinical Trials with Novel Designs and Implementations <i>Lorenzo Trippa and Yanxun Xu</i>	216
Comment: Group Sequential Designs with Response-Adaptive Randomisation <i>Christopher Jennison</i>	219
Comment: Is Response-Adaptive Randomization a “Good Thing” or Not in Clinical Trials? Why We Cannot Take Sides <i>Alessandra Giovagnoli</i>	224
Comment: Response Adaptive Randomization in Practice <i>Scott M. Berry and Kert Viele</i>	229
Rejoinder: Response-Adaptive Randomization in Clinical Trials <i>David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofia S. Villar</i>	233
Efficient Data Augmentation Techniques for Some Classes of State Space Models <i>Linda S. L. Tan</i>	240
Distributed Bayesian Inference in Massive Spatial Data <i>Rajarshi Guhaniyogi, Cheng Li, Terrance Savitsky and Sanvesh Srivastava</i>	262
The Secret Life of I. J. Good <i>Sandy Zabell</i>	285
Cross-Study Replicability in Cluster Analysis <i>Lorenzo Masoero, Emma Thomas, Giovanni Parmigiani, Svitlana Tyekucheva and Lorenzo Trippa</i>	303
Principal Fairness for Human and Algorithmic Decision-Making <i>Kosuke Imai and Zhichao Jiang</i>	317
Confidence and Discoveries with E -values <i>Vladimir Vovk and Ruodu Wang</i>	329
Conversations with Gábor J. Székely <i>Yulia R. Gel, Edsel A. Peña and Huixia Judy Wang</i>	355

Statistical Science [ISSN 0883-4237 (print); ISSN 2168-8745 (online)], Volume 38, Number 2, May 2023. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices. **POSTMASTER:** Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Copyright © 2023 by the Institute of Mathematical Statistics
Printed in the United States of America

Statistical Science

Volume 38, Number 2 (185–367) May 2023

Volume 38

Number 2

May 2023

Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations (with Discussion)

David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofía S. Villar

Efficient Data Augmentation Techniques for Some Classes of State Space Models

Linda S. L. Tan

Distributed Bayesian Inference in Massive Spatial Data

Rajarshi Guhaniyogi, Cheng Li, Terrance Savitsky and Sanvesh Srivastava

The Secret Life of I. J. Good

Sandy Zabell

Cross-Study Replicability in Cluster Analysis

Lorenzo Masoero, Emma Thomas, Giovanni Parmigiani, Svitlana Tyekucheva and Lorenzo Trippa

Principal Fairness for Human and Algorithmic Decision-Making

Kosuke Imai and Zhichao Jiang

Confidence and Discoveries with E-values

Vladimir Vovk and Ruodu Wang

Conversations with Gábor J. Székely

Yulia R. Gel, Edsel A. Peña and Huixia Judy Wang

EDITOR

Moulinath Banerjee
University of Michigan

ASSOCIATE EDITORS

Shankar Bhamidi
University of North Carolina
Jay Breidt
University of Chicago
Alicia Carriquiry
Iowa State University
Nilanjan Chatterjee
Johns Hopkins University
Yang Chen
University of Michigan
Bertrand Clarke
University of Nebraska-Lincoln
Michael J. Daniels
University of Florida
Philip Dawid
University of Cambridge
Holger Dette
Ruhr-Universität Bochum
Robin Evans
University of Oxford
Stefano Favaro
Università di Torino
Subhashis Ghoshal
North Carolina State University

Peter Green
*University of Bristol and
University of Technology
Sydney*
Chris Holmes
University of Oxford
Tailen Hsing
University of Michigan
Po-Ling Loh
University of Cambridge
Ian McKeague
Columbia University
George Michailidis
*University of California, Los
Angeles*
Peter Müller
University of Texas
Axel Munk
*Georg-August-University of
Goettingen*
Jean Opsomer
Westat
Sonia Petrone
Università Bocconi, Milan

Nancy Reid
University of Toronto
Thomas Richardson
University of Washington
Pietro Rigo
Università di Bologna
Jason Roy
Rutgers University
Purnamrita Sarkar
University of Texas at Austin
Richard Samworth
University of Cambridge
Bodhisattva Sen
Columbia University
Yuekai Sun
University of Michigan
Bin Yu
*University of California,
Berkeley*
Giacomo Zanella
Università Bocconi, Milan

MANAGING EDITOR

Dan Nordman
Iowa State University

PRODUCTION EDITOR

Patrick Kelly

EDITORIAL COORDINATOR

Kristina Mattson

PAST EXECUTIVE EDITORS

Morris H. DeGroot, 1986–1988
Carl N. Morris, 1989–1991
Robert E. Kass, 1992–1994
Paul Switzer, 1995–1997
Leon J. Gleser, 1998–2000
Richard Tweedie, 2001
Morris Eaton, 2001
George Casella, 2002–2004
Edward I. George, 2005–2007
David Madigan, 2008–2010
Jon A. Wellner, 2011–2013
Peter Green, 2014–2016
Cun-Hui Zhang, 2017–2019
Sonia Petrone, 2020–2022

Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations

David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofía S. Villar

Abstract. Response-Adaptive Randomization (RAR) is part of a wider class of data-dependent sampling algorithms, for which clinical trials are typically used as a motivating application. In that context, patient allocation to treatments is determined by randomization probabilities that change based on the accrued response data in order to achieve experimental goals. RAR has received abundant theoretical attention from the biostatistical literature since the 1930s and has been the subject of numerous debates. In the last decade, it has received renewed consideration from the applied and methodological communities, driven by well-known practical examples and its widespread use in machine learning. Papers on the subject present different views on its usefulness, and these are not easy to reconcile. This work aims to address this gap by providing a broad, balanced and fresh review of methodological and practical issues to consider when debating the use of RAR in clinical trials.

Key words and phrases: Ethics, patient allocation, power, sample size imbalance, time trends, type I error control.

REFERENCES

- ANGUS, D. C., BERRY, S., LEWIS, R. J., AL-BEIDH, F., ARABI, Y., VAN BENTUM-PUIJK, W., BHIMANI, Z., BON-TEN, M., BROGLIO, K. et al. (2020). The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. Rationale and design. *Ann. Amer. Thorac. Soc.* **17** 879–891. <https://doi.org/10.1513/AnnalsATS.202003-192SD>
- ANSCOMBE, F. J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58** 365–383. [MR0152096](https://doi.org/10.2307/1402871)
- ARMITAGE, P. (1985). The search for optimality in clinical trials. *Int. Stat. Rev.* **53** 15–24. [MR0959040](https://doi.org/10.2307/1402871)
- ATKINSON, A. C. and BISWAS, A. (2014). *Randomized Response-Adaptive Designs in Clinical Trials*. CRC Press, Boca Raton, FL.
- ATKINSON, A. C., BISWAS, A. and PRONZATO, L. (2011). Covariate-balanced response-adaptive designs for clinical trials with continuous responses that target allocation probabilities. Technical Report NII1042-DAE, Isaac Newton Institute for Mathematical Sciences, Cambridge.
- BAI, Z. D., HU, F. and ROSENBERGER, W. F. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *Ann. Statist.* **30** 122–139. [MR1892658](https://doi.org/10.1214/aos/1015362187)
- BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2010). Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika* **97** 935–946. [MR2746162](https://doi.org/10.1093/biomet/asq055)
- BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2015). *Adaptive Designs for Sequential Treatment Allocation*. Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL. [MR3616325](https://doi.org/10.1093/biomet/asq055)
- BALDI ANTOGNINI, A., NOVELLI, M. and ZAGORAIOU, M. (2022a). A simple solution to the inadequacy of asymptotic likelihood-based inference for response-adaptive clinical trials: Likelihood-based inference for RAR trials. *Statist. Papers* **63** 157–180. [MR4372934](https://doi.org/10.1007/s00362-021-01234-3)
- BALDI ANTOGNINI, A., NOVELLI, M. and ZAGORAIOU, M. (2022b). A new inferential approach for response-adaptive clinical trials: The variance-stabilized bootstrap. The variance-stabilized bootstrap for RA designs. *TEST* **31** 235–254. [MR4386433](https://doi.org/10.1007/s11749-021-00777-9)
- BALDI ANTOGNINI, A., VAGHEGGINI, A. and ZAGORAIOU, M. (2018). Is the classical Wald test always suitable under response-adaptive randomization? *Stat. Methods Med. Res.* **27** 2294–2311. [MR3825908](https://doi.org/10.1177/0962280216680241)
- BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2011). The covariate-adaptive biased coin design for balancing clinical trials in the pres-

David S. Robertson is a Senior Research Associate, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (e-mail: david.robertson@mrc-bsu.cam.ac.uk). Kim May Lee is a Research Fellow, King's College London, London, UK (e-mail: kim.lee@kcl.ac.uk). Boryana C. López-Kolkovska is a Statistical Science Associate Director, AstraZeneca, Cambridge, UK (e-mail: boryana.kolkovska@astrazeneca.com). Sofía S. Villar is a MRC Investigator, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (e-mail: sofia.villar@mrc-bsu.cam.ac.uk).

- ence of prognostic factors. *Biometrika* **98** 519–535. MR2836404 <https://doi.org/10.1093/biomet/asr021>
- BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2012). Multi-objective optimal designs in comparative clinical trials with covariates: The reinforced doubly adaptive biased coin design. *Ann. Statist.* **40** 1315–1345. MR3015027 <https://doi.org/10.1214/12-AOS1007>
- BALDI ANTOGNINI, A., VAGHEGGINI, A., ZAGORAIOU, M. and NOVELLI, M. (2018). A new design strategy for hypothesis testing under response adaptive randomization. *Electron. J. Stat.* **12** 2454–2481. MR3832098 <https://doi.org/10.1214/18-EJS1458>
- BARKER, A. D., SIGMAN, C. C., KELLOFF, G. J., HYLTON, N. M., BERRY, D. A. and ESSERMAN, L. J. (2009). I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin. Pharmacol. Ther.* **86** 97–100.
- BARNETT, H. Y., VILLAR, S. S., GEYS, H. and JAKI, T. (2021). A novel statistical test for treatment differences in clinical trials using a response-adaptive forward-looking Gittins index rule. *Biometrics* **79** 86–97. <https://doi.org/10.1111/biom.13581>
- BAROHN, R. J., GAJEWSKI, B., PASNOOR, M., BROWN, A., HERBELIN, L. L., KIMMINAU, K. S., MUDARANTHAKAM, D. P., JAWDAT, O., DIMACHKIE, M. M. and PATIENT ASSISTED INTERVENTION FOR NEUROPATHY: COMPARISON OF TREATMENT IN REAL LIFE SITUATIONS (PAIN-CONTROLS) STUDY TEAM (2021). Patient Assisted Intervention for Neuropathy: Comparison of Treatment in Real Life Situations (PAIN-CONTROLS): Bayesian adaptive comparative effectiveness randomized trial. *JAMA Neurology* **78** 68–76.
- BARTLETT, R., ROLOFF, D., CORNELL, R., ANDREWS, A., DILLON, P. and ZWISCHENBERGER, J. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* **76** 479–487.
- BEAUCHAMP, T. L. Informed consent. In *Medical Ethics*, 2nd ed. (R. M. Veatch, ed.) 185–508. Jones and Bartlett, Boston, MA.
- BELLO, G. A. and SABO, R. T. (2016). Outcome-adaptive allocation with natural lead-in for three-group trials with binary outcomes. *J. Stat. Comput. Simul.* **86** 2441–2449. MR3502173 <https://doi.org/10.1080/00949655.2015.1114116>
- BERRY, D. A. (2004). Bayesian statistics and the efficiency and ethics of clinical trials. *Statist. Sci.* **19** 175–187. MR2086326 <https://doi.org/10.1214/0883423040000000044>
- BERRY, D. A. and EICK, S. G. (1995). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Stat. Med.* **14** 231–246.
- BERRY, S. M., PETZOLD, E. A., DULL, P., THIELMAN, N. M., CUNNINGHAM, C. K., COREY, G. R., MCCLAIN, M. T., HOOVER, D. L., RUSSELL, J. et al. (2016). A response adaptive randomization platform trial for efficient evaluation of Ebola virus treatments: A model for pandemic response. *Clin. Trials* **13** 22–30.
- BLACKWELL, M., HONAKER, J. and KING, G. (2017). A unified approach to measurement error and missing data: Overview and applications. *Sociol. Methods Res.* **46** 303–341. MR3671517 <https://doi.org/10.1177/0049124115585360>
- BOWDEN, J. and TRIPPA, L. (2017). Unbiased estimation for response adaptive clinical trials. *Stat. Methods Med. Res.* **26** 2376–2388. MR3712238 <https://doi.org/10.1177/0962280215597716>
- BRITTAİN, E. H. and PROSCHAN, M. A. (2016). Comments on Berry et al.'s response-adaptive randomization platform trial for Ebola. *Clin. Trials* **13** 566–567.
- BURTON, P. R., GURRINA, L. C. and HUSSEY, M. H. (1997). Interpreting the clinical trials of extracorporeal membrane oxygenation in the treatment of persistent pulmonary hypertension of the newborn. *Semin. Neonatol.* **2** 69–79.
- CHEN, X., LEE, K. M., VILLAR, S. S. and ROBERTSON, D. S. (2022). Some performance considerations when using multi-armed bandit algorithms in the presence of missing data. *PLoS ONE.* **17** e0274272.
- CHENG, Y., SU, F. and BERRY, D. A. (2003). Choosing sample size for a clinical trial using decision analysis. *Biometrika* **90** 923–936. MR2024767 <https://doi.org/10.1093/biomet/90.4.923>
- CHEVRET, S. (2012). Bayesian adaptive clinical trials: A dream for statisticians only? *Stat. Med.* **31** 1002–1013. MR2925674 <https://doi.org/10.1002/sim.4363>
- CHOW, S.-C. and CHANG, M. (2007). *Adaptive Design Methods in Clinical Trials*. CRC Press, Boca Raton, FL. MR2857824
- COAD, D. S. (1991). Sequential tests for an unstable response variable. *Biometrika* **78** 113–121. MR1118236 <https://doi.org/10.1093/biomet/78.1.113>
- COAD, D. S. and GOVINDARAJULU, Z. (2000). Corrected confidence intervals following a sequential adaptive clinical trial with binary responses. *J. Statist. Plann. Inference* **91** 53–64. MR1792363 [https://doi.org/10.1016/S0378-3758\(00\)00129-4](https://doi.org/10.1016/S0378-3758(00)00129-4)
- COAD, D. S. and IVANOVA, A. (2001). Bias calculations for adaptive urn designs. *Sequential Anal.* **20** 91–116. MR1855413 <https://doi.org/10.1081/SQA-100106051>
- COLTON, T. (1963). A model for selecting one of two medical treatments. *J. Amer. Statist. Assoc.* **58** 388–400. MR0149637
- DAS, S. and LO, A. W. (2017). Re-inventing drug development: A case study of the I-SPY 2 breast cancer clinical trials program. *Contemp. Clin. Trials* **62** 168–174. <https://doi.org/10.1016/j.cct.2017.09.002>
- DAWSON, A. (2009). The normative status of the requirement to gain an informed consent in clinical trials: Comprehension, obligations and empirical evidence. In *The Limits of Consent: A Sociolegal Approach to Human Subject Research in Medicine* (O. Corrigan, J. McMillan, K. Liddell, M. Richards and C. Weijer, eds.) 99–113. Oxford Univ. Press, Oxford.
- DELIU, N., WILLIAMS, J. J. and VILLAR, S. S. (2021). Efficient inference without trading-off regret in bandits: An allocation probability test for Thompson sampling. arXiv preprint, arXiv:2111.00137.
- EISELE, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *J. Statist. Plann. Inference* **38** 249–261. MR1256599 [https://doi.org/10.1016/0378-3758\(94\)90038-8](https://doi.org/10.1016/0378-3758(94)90038-8)
- FASERU, B., ELLERBECK, E. F., CATLEY, D., GAJEWSKI, B. J., SCHEUERMANN, T. S., SHIREMAN, T. I., MUSSULMAN, L. M., NAZIR, N., BUSH, T. et al. (2017). Changing the default for tobacco-cessation treatment in an inpatient setting: Study protocol of a randomized controlled trial. *Trials* **18** 379.
- FLOURNOY, N., HAINES, L. M. and ROSENBERGER, W. F. (2013). A graphical comparison of response-adaptive randomization procedures. *Stat. Biopharm. Res.* **5** 126–141.
- FREEDMAN, B. (1987). Equipoise and the ethics of clinical research. *N. Engl. J. Med.* **317** 141–145.
- GALBETE, A., MOLER, J. A. and PLO, F. (2016). Randomization tests in recursive response-adaptive randomization procedures. *Statistics* **50** 418–434. MR3452994 <https://doi.org/10.1080/02331888.2015.1050020>
- GALBETE, A. and ROSENBERGER, W. F. (2016). On the use of randomization tests following adaptive designs. *J. Biopharm. Statist.* **26** 466–474.
- GLIMM, E. and ROBERTSON, D. S. (2022). Familywise error rate control for block response-adaptive randomization. *Stat. Methods Med. Res.* <https://doi.org/10.1177/09622802231167437>
- GRIEVE, A. P. (2017). Response-adaptive clinical trials: Case studies in the medical literature. *Pharm. Stat.* **16** 64–86. <https://doi.org/10.1002/pst.1778>
- GU, X. and LEE, J. J. (2010). A simulation study for comparing testing statistics in response-adaptive randomization. *BMC Med. Res. Methodol.* **10** 48.

- GUOLO, A. (2008). Robust techniques for measurement error correction: A review. *Stat. Methods Med. Res.* **17** 555–580. MR2654662 <https://doi.org/10.1177/0962280207081318>
- HADAD, V., HIRSHBERG, D. A., ZHAN, R., WAGER, S. and ATHEY, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proc. Natl. Acad. Sci. USA* **118** Paper No. e2014602118, 10. MR4294058 <https://doi.org/10.1073/pnas.2014602118>
- HU, F. and ROSENBERGER, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons. *J. Amer. Statist. Assoc.* **98** 671–678. MR2011680 <https://doi.org/10.1198/016214503000000576>
- HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials. Wiley Series in Probability and Statistics.* Wiley Interscience, Hoboken, NJ. MR2245329 <https://doi.org/10.1002/047005588X>
- HU, F. and ZHANG, L.-X. (2004a). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann. Statist.* **32** 268–301. MR2051008 <https://doi.org/10.1214/aos/1079120137>
- HU, F. and ZHANG, L.-X. (2004b). Asymptotic normality of urn models for clinical trials with delayed response. *Bernoulli* **10** 447–463. MR2061440 <https://doi.org/10.3150/bj/1089206406>
- HU, F., ZHANG, L.-X. and HE, X. (2009). Efficient randomized-adaptive designs. *Ann. Statist.* **37** 2543–2560. MR2543702 <https://doi.org/10.1214/08-AOS655>
- HU, J., ZHU, H. and HU, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *J. Amer. Statist. Assoc.* **110** 357–367. MR3338508 <https://doi.org/10.1080/01621459.2014.903846>
- HU, F., ZHANG, L.-X., CHEUNG, S. H. and CHAN, W. S. (2008). Doubly adaptive biased coin designs with delayed responses. *Canad. J. Statist.* **36** 541–559. MR2532251 <https://doi.org/10.1002/cjs.5550360404>
- IVANOVA, A. (2003). A play-the-winner-type urn design with reduced variability. *Metrika* **58** 1–13. MR1999248 <https://doi.org/10.1007/s001840200220>
- JACKO, P. (2019). The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths. arXiv preprint, arXiv:1906.10173.
- JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* CRC Press/CRC, Boca Raton, FL. MR1710781
- JENNISON, C. and TURNBULL, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Anal.* **20** 209–234. MR1868774 <https://doi.org/10.1081/SQA-100107646>
- JEON, Y. and HU, F. (2010). Optimal adaptive designs for binary response trials with three treatments. *Stat. Biopharm. Res.* **2** 310–318.
- JIANG, Y., ZHAO, W. and DURKALSKI-MAULDIN, V. (2020). Time-trend impact on treatment estimation in two-arm clinical trials with a binary outcome and Bayesian response adaptive randomization. *J. Biopharm. Statist.* **30** 69–88.
- JOHNSON, R., JACKSON, C., PRESANIS, A., VILLAR, S. S. and ANGELIS, D. D. (2022). Quantifying efficiency gains of innovative designs of two-arm vaccine trials for Covid-19 using an epidemic simulation model. *Stat. Biopharm. Res.* **14** 33–41. <https://doi.org/10.1080/19466315.2021.1939774>
- KAIBEL, C. and BIEMANN, T. (2021). Rethinking the gold standard with multi-armed bandits: Machine learning allocation algorithms for experiments. *Organ. Res. Methods* **24** 78–103.
- KARRISON, T. G., HUO, D. and CHAPPELL, R. (2003). A group sequential, response-adaptive design for randomized clinical trials. *Control. Clin. Trials* **24** 506–522.
- KAUFMANN, E. and GARIVIER, A. (2017). Learning the distribution with largest mean: Two bandit frameworks. In *Journées MAS 2016 de la SMAI—Phénomènes Complexes et Hétérogènes. ESAIM Proc. Surveys* **60** 114–131. EDP Sci., Les Ulis. MR3772476 <https://doi.org/10.1051/proc/201760114>
- KAUFMANN, E., KORDA, N. and MUNOS, R. (2012). Thompson sampling: An asymptotically optimal finite-time analysis. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **7568** 199–213. Springer, Heidelberg. MR3042891 https://doi.org/10.1007/978-3-642-34106-9_18
- KIM, E. S., HERBST, R. S., WISTUBA, I. I., LEE, J. J., BLUMENSCHNEIN, G. R., TSAO, A., STEWART, D. J., HICKS, M. E., ERASMUS, J. JR et al. (2011). The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discov.* **1** 44–53.
- KORN, E. L. and FREIDLIN, B. (2011a). Outcome-adaptive randomization: Is it useful? *J. Clin. Oncol.* **29** 771–776.
- KORN, E. L. and FREIDLIN, B. (2011b). Reply to Y. Yuan et al. *J. Clin. Oncol.* **29** e393.
- KORN, E. L. and FREIDLIN, B. (2017). Adaptive clinical trials: Advantages and disadvantages of various adaptive design elements. *J. Natl. Cancer Inst.* **109** dxj013.
- KORN, E. L. and FREIDLIN, B. (2022). Time trends with response-adaptive randomization: The inevitability of inefficiency. *Clin. Trials* **19** 158–161.
- LAAGE, T., LOEWY, J. W., MENON, S., MILLER, E. R., PULKSTENIS, E., KAN-DOBROSKY, N. and COFFEY, C. (2017). Ethical considerations in adaptive design clinical trials. *Ther. Innov. Regul. Sci.* **51** 190–199.
- LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms.* Cambridge Univ. Press, Cambridge, UK.
- LEE, J. J., CHEN, N. and YIN, G. (2012). Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clin. Cancer Res.* **18** 4498–4507.
- LEE, K. M. and LEE, J. J. (2021). Evaluating Bayesian adaptive randomization procedures with adaptive clip methods for multi-arm trials. *Stat. Methods Med. Res.* **30** 1273–1287. MR4259898 <https://doi.org/10.1177/0962280221995961>
- LEE, K. M., MITRA, R. and BIEDERMANN, S. (2018). Optimal design when outcome values are not missing at random. *Statist. Sinica* **28** 1821–1838. MR3838993
- LI, X. and WANG, X. (2012). Variance-penalized response-adaptive randomization with mismeasurement. *J. Statist. Plann. Inference* **142** 2128–2135. MR2903417 <https://doi.org/10.1016/j.jspi.2012.02.016>
- LI, X. and WANG, X. (2013). Response adaptive designs with misclassified responses. *Comm. Statist. Theory Methods* **42** 2071–2083. MR3053882 <https://doi.org/10.1080/03610926.2011.602488>
- LIN, J. and BUNN, V. (2017). Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp. Clin. Trials* **54** 48–59. <https://doi.org/10.1016/j.cct.2017.01.003>
- LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics.* Wiley Interscience, Hoboken, NJ. MR1925014 <https://doi.org/10.1002/9781119013563>
- LONDON, A. J. (2018). Learning health systems, clinical equipoise and the ethics of response adaptive randomization. *J. Med. Ethics* **44** 409–415.
- MAGARET, A. S., JACOB, S. T., HALLORAN, M. E., GUTHRIE, K. A., MAGARET, C. A., JOHNSTON, C., SIMON, N. R. and WALD, A. (2020). Multigroup, adaptively randomized trials are advantageous for comparing coronavirus disease 2019 (Covid-19) interventions. *Ann. Intern. Med.* **173** 576–577. <https://doi.org/10.7326/M20-2933>

- MARSCHNER, I. C. (2021). A general framework for the analysis of adaptive experiments. *Statist. Sci.* **36** 465–492. MR4293100 <https://doi.org/10.1214/20-STSS803>
- MCGREE, J. M., DROVANDI, C. C., THOMPSON, M. H., ECCLESTON, J. A., DUFFULL, S. B., MENGERSEN, K., PETTITT, A. N. and GOGGIN, T. (2012). Adaptive Bayesian compound designs for dose finding studies. *J. Statist. Plann. Inference* **142** 1480–1492. MR2891500 <https://doi.org/10.1016/j.jspi.2011.12.029>
- MELFI, V. F. and PAGE, C. (2000). Estimation after adaptive allocation. *J. Statist. Plann. Inference* **87** 353–363. MR1771124 [https://doi.org/10.1016/S0378-3758\(99\)00198-6](https://doi.org/10.1016/S0378-3758(99)00198-6)
- METELKINA, A. and PRONZATO, L. (2017). Information-regret compromise in covariate-adaptive treatment allocation. *Ann. Statist.* **45** 2046–2073. MR3718161 <https://doi.org/10.1214/16-AOS1518>
- MORRIS, T. P., WHITE, I. R. and CROWTHER, M. J. (2019). Using simulation studies to evaluate statistical methods. *Stat. Med.* **38** 2074–2102. MR3937487 <https://doi.org/10.1002/sim.8086>
- O'BRIEN, B., GREEN, C. E., AL-JURDI, R., CHANG, L., LIJFFIJT, M., IQBAL, S., IQBAL, T., SWANN, A. C. and MATHEW, S. J. (2019). Bayesian adaptive randomization trial of intravenous ketamine for veterans with late-life, treatment-resistant depression. *Contemp. Clin. Trials Commun.* **16** 100432. <https://doi.org/10.1016/j.conctc.2019.100432>
- PAPADIMITRAKOPOULOU, V., LEE, J. J., WISTUBA, I., TSAO, A., FOSSELLA, F., KALHOR, N., GUPTA, S., AVERETT BYERS, L., IZZO, J. et al. (2016). The BATTLE-2 study: A biomarker-integrated targeted therapy study in previously treated patients with advanced non-small-cell lung cancer. *J. Clin. Oncol.* **34** 3638–3647.
- PITT, E. R. (2021). Optimising first in human trials. Ph.D. thesis, Univ. Bath, Bath. Available at https://purehost.bath.ac.uk/ws/portalfiles/portal/226805176/LizziPitt_final_thesis.pdf.
- PROSCHAN, M. A. and DODD, L. E. (2019). Re-randomization tests in clinical trials. *Stat. Med.* **38** 2292–2302. MR3939818 <https://doi.org/10.1002/sim.8093>
- PROSCHAN, M. and EVANS, S. (2020). Resist the temptation of response-adaptive randomization. *Clin. Infect. Dis.* **71** 3002–3004.
- REMAP-CAP INVESTIGATORS (2021). Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N. Engl. J. Med.* **385** 1491–1502.
- ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246 <https://doi.org/10.1090/S0002-9904-1952-09620-8>
- ROBERTSON, D. S. and WASON, J. M. S. (2019). Familywise error control in multi-armed response-adaptive trials. *Biometrics* **75** 885–894. MR4012094 <https://doi.org/10.1111/biom.13042>
- ROBERTSON, D. S., CHOODARI-OSKOOEI, B., DIMAIRO, M., FLIGHT, L. and JAKI, T. (2023). Point estimation for adaptive trial designs I: A methodological review. *Stat. Med.* **42** 122–145. <https://doi.org/10.1002/sim.9605>
- ROBERTSON, D. S., CHOODARI-OSKOOEI, B., DIMAIRO, M., FLIGHT, L. and JAKI, T. (2023). Point estimation for adaptive trial designs II: Practical considerations and guidance. *Stat. Med.* <https://doi.org/10.1002/sim.9734>
- ROSENBERGER, W. F. (2015). A conversation with Nancy Flournoy. *Statist. Sci.* **30** 133–146. MR3317758 <https://doi.org/10.1214/14-STSS495>
- ROSENBERGER, W. F. and HU, F. (1999). Bootstrap methods for adaptive designs. *Stat. Med.* **18** 1757–1767.
- ROSENBERGER, W. F. and HU, F. (2004). Maximising power and minimizing treatment failures in clinical trials. *Clin. Trials* **1** 141–147.
- ROSENBERGER, W. F. and LACHIN, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice*. Wiley Series in Probability and Statistics. Wiley Interscience, New York. MR1914364 <https://doi.org/10.1002/0471722103>
- ROSENBERGER, W. F. and LACHIN, J. M. (2016). *Randomization in Clinical Trials: Theory and Practice*, 2nd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. MR3443072 <https://doi.org/10.1002/9781118742112>
- ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. MR2483911 <https://doi.org/10.1214/08-STSS269>
- ROSENBERGER, W. F., SVERDLOV, O. and HU, F. (2012). Adaptive randomization for clinical trials. *J. Biopharm. Statist.* **22** 719–736. MR2931067 <https://doi.org/10.1080/10543406.2012.676535>
- ROSENBERGER, W. F., VIDYASHANKAR, A. N. and AGARWAL, D. K. (2001). Covariate-adjusted response-adaptive designs for binary response. *J. Biopharm. Statist.* **11** 227–236.
- ROSENBERGER, W. F., STALLARD, N., IVANOVA, A., HARPER, C. N. and RICKS, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57** 909–913. MR1863454 <https://doi.org/10.1111/j.0006-341X.2001.00909.x>
- ROSNER, G. L. (2020). Bayesian adaptive design in drug development. In *Bayesian Methods in Pharmaceutical Research* (E. Lesaffre, G. Baio and B. Boulangier, eds.) CRC Press/CRC Press, Boca Raton, FL.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. With comments by R. J. A. Little and a reply by the author. MR0455196 <https://doi.org/10.1093/biomet/63.3.581>
- RYAN, E. G., DROVANDI, C. C., MCGREE, J. M. and PETTITT, A. N. (2016). A review of modern computational algorithms for Bayesian optimal design. *Int. Stat. Rev.* **84** 128–154. MR3491282 <https://doi.org/10.1111/insr.12107>
- SABO, R. T. (2014). Adaptive allocation for binary outcomes using decreasingly informative priors. *J. Biopharm. Statist.* **24** 569–578. MR3196161 <https://doi.org/10.1080/10543406.2014.888441>
- SAMANIEGO, F. J. (2010). *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. Springer Series in Statistics. Springer, New York. MR2664350 <https://doi.org/10.1007/978-1-4419-5941-6>
- SIMON, R. and SIMON, N. R. (2011). Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization. *Statist. Probab. Lett.* **81** 767–772. MR2793742 <https://doi.org/10.1016/j.spl.2010.12.018>
- SIU, L. L., IVY, S. P., DIXON, E. L., GRAVELL, A. E., REEVES, S. A. and ROSNER, G. L. (2017). Challenges and opportunities in adapting clinical trial design for immunotherapies. *Clin. Cancer Res.* **23** 4950–4958. <https://doi.org/10.1158/1078-0432.CCR-16-3079>
- STALLARD, N. and ROSENBERGER, W. F. (2020). Comparison of Bayesian and frequentist group-sequential clinical trial designs. *BMC Med. Res. Methodol.* **20**.
- SUGARMAN, J., DOUGLAS, C. MCCRORY, D. C., POWELL, D., KRASNY, A., ADAMS, B., BALL, E. and CASSELL, C. (1999). Empirical research on informed consent. *Hastings Cent. Rep.* **29**(suppl) S1–S42.
- SVERDLOV, O., ed. (2016). *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*. Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL. MR3642968
- SVERDLOV, O. and ROSENBERGER, W. F. (2013a). On recent advances in optimal allocation designs in clinical trials. *J. Stat. Theory Pract.* **7** 753–773. MR3196633 <https://doi.org/10.1080/15598608.2013.783726>
- SVERDLOV, O. and ROSENBERGER, W. F. (2013b). Randomization in clinical trials: Can we eliminate bias? *Clin. Invest.* **3** 37–47.

- TAMURA, R. N., FARIES, D. E., ANDERSEN, J. S. and HEILIGENSTEIN, J. H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J. Amer. Statist. Assoc.* **89** 768–776.
- THALL, P. F. (2020). *Statistical Remedies for Medical Researchers. Springer Series in Pharmaceutical Statistics.*
- THALL, P. F., FOX, P. and WATHEN, J. (2015). Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann. Oncol.* **26** 1621–1628.
- THALL, P. F., FOX, P. S. and WATHEN, J. K. (2016). Some caveats for outcome adaptive randomization in clinical trials. In *Modern Adaptive Randomized Clinical Trials* (O. Sverdlov, ed.). *Chapman & Hall/CRC Biostat. Ser.* 287–305. CRC Press, Boca Raton, FL. MR3676361
- THALL, P. F. and WATHEN, J. K. (2007). Practical Bayesian adaptive randomization in clinical trials. *Eur. J. Cancer* **43** 859–866.
- THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.
- TORGERSON, D. J. and CAMPBELL, M. K. (2000). Use of unequal randomization to aid the economic efficiency of clinical trials. *BMJ* **321** 759.
- TRIPPA, L., LEE, E. Q., WEN, P. Y., BATCHELOR, T. T., CLOUGHESY, T., PARMIGIANI, G. and ALEXANDER, B. M. (2012). Bayesian adaptive trial design for patients with recurrent glioblastoma. *J. Clin. Oncol.* **30** 3258–3263.
- TYMOFYEV, Y., ROSENBERGER, W. F. and HU, F. (2007). Implementing optimal allocation in sequential binary response experiments. *J. Amer. Statist. Assoc.* **102** 224–234. MR2345540 <https://doi.org/10.1198/016214506000000906>
- U.S. FOOD AND DRUG ADMINISTRATION (2019). Adaptive designs for clinical trials of drugs and biologics. Available at <https://www.fda.gov/media/78495/download>. Accessed 8 March 2022.
- VENTZ, S., PARMIGIANI, G. and TRIPPA, L. (2017). Combining Bayesian experimental designs and frequentist data analyses: Motivations and examples. *Appl. Stoch. Models Bus. Ind.* **33** 302–313. MR3666933 <https://doi.org/10.1002/asmb.2249>
- VICKERSTAFF, V., OMAR, R. and AMBLER, G. (2019). Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Med. Res. Methodol.* **19** 129.
- VIELE, K., BROGLIO, K., MCGLOTHLIN, A. and SAVILLE, B. R. (2020a). Comparison of methods for control allocation in multiple arm studies using response adaptive randomization. *Clin. Trials* **17** 52–20.
- VIELE, K., SAVILLE, B. R., MCGLOTHLIN, A. and BROGLIO, K. (2020b). Comparison of response adaptive randomization features in multiarm clinical trials with control. *Pharm. Stat.* **19** 602–612.
- VILLAR, S. S., BOWDEN, J. and WASON, J. (2015). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statist. Sci.* **30** 199–215. MR3353103 <https://doi.org/10.1214/14-STS504>
- VILLAR, S. S., BOWDEN, J. and WASON, J. (2018). Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? *Pharm. Stat.* **17** 182–197. <https://doi.org/10.1002/pst.1845>
- VILLAR, S. S., ROBERTSON, D. S. and ROSENBERGER, W. F. (2021). The temptation of overgeneralizing response-adaptive randomization. *Clin. Infect. Dis.* **73** e842.
- VILLAR, S. S., WASON, J. and BOWDEN, J. (2015). Response-adaptive randomization for multi-arm clinical trials using the forward looking Gittins index rule. *Biometrics* **71** 969–978. MR3436722 <https://doi.org/10.1111/biom.12337>
- WAGENMAKERS, E. J., LEE, M., LODEWYCKX, T. and IVERSON, G. (2008). Bayesian versus frequentist inference. In *Bayesian Evaluation of Informative Hypotheses* 181–207. Springer, New York, NY.
- WANG, Y. and ROSENBERGER, W. F. (2020). Randomization-based interval estimation in randomized clinical trials. *Stat. Med.* **39** 2843–2854. MR4133143 <https://doi.org/10.1002/sim.8577>
- WANG, Y., ZHU, H. and LEE, J. J. (2020). Evaluation of bias for outcome adaptive randomization designs with binary endpoints. *Stat. Interface* **13** 287–315. MR4091798 <https://doi.org/10.4310/SII.2020.v13.n3.a2>
- WARE, J. H. (1989). Investigating therapies of potentially great benefit: ECMO. *Statist. Sci.* **4** 298–340. With comments and a rejoinder by the author. MR1041761
- WASON, J. M. S., BROCKLEHURST, P. and YAP, C. (2019). When to keep it simple—adaptive designs are not always useful. *BMC Med.* **17** 152.
- WASON, J. M. S. and TRIPPA, L. (2014). A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Stat. Med.* **33** 2206–2221. MR3256687 <https://doi.org/10.1002/sim.6086>
- WATHEN, J. K. and THALL, P. F. (2017). A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clin. Trials* **14** 432–440. <https://doi.org/10.1177/1740774517692302>
- WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.
- WILLIAMSON, S. F. and VILLAR, S. S. (2020). A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. *Biometrics* **76** 197–209. MR4098555 <https://doi.org/10.1111/biom.13119>
- WILLIAMSON, S. F., JACKO, P., VILLAR, S. S. and JAKI, T. (2017). A Bayesian adaptive design for clinical trials in rare diseases. *Comput. Statist. Data Anal.* **113** 136–153. MR3662397 <https://doi.org/10.1016/j.csda.2016.09.006>
- WOODCOCK, J. and LAVANGE, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *N. Engl. J. Med.* **377** 62–70. <https://doi.org/10.1056/NEJMr1510062>
- YUAN, Y. and YIN, G. (2011). On the usefulness of outcome adaptive randomization. *J. Clin. Oncol.* **29** 771–776.
- ZAGORAIOU, M. (2017). Choosing a covariate-adaptive randomization procedure in practice. *J. Biopharm. Statist.* **27** 845–857. <https://doi.org/10.1080/10543406.2017.1289944>
- ZELLEN, M. (1969). Play the winner rule and the controlled clinical trial. *J. Amer. Statist. Assoc.* **64** 131–146. MR0240938
- ZHANG, L. and ROSENBERGER, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* **62** 562–569. MR2236838 <https://doi.org/10.1111/j.1541-0420.2005.00496.x>
- ZHANG, L. and ROSENBERGER, W. F. (2007). Response-adaptive randomization for survival trials: The parametric approach. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **56** 153–165. MR2359239 <https://doi.org/10.1111/j.1467-9876.2007.00571.x>
- ZHANG, L.-X., CHAN, W. S., CHEUNG, S. H. and HU, F. (2007). A generalized drop-the-loser urn for clinical trials with delayed responses. *Statist. Sinica* **17** 387–409. MR2352516

Comment: A Quarter Century of Methodological Research in Response-Adaptive Randomization

Anastasia Ivanova and William F. Rosenberger

REFERENCES

- HU, F., ROSENBERGER, W. F. and ZHANG, L.-X. (2006). Asymptotically best response-adaptive randomization procedures. *J. Statist. Plann. Inference* **136** 1911–1922. MR2255603 <https://doi.org/10.1016/j.jspi.2005.08.011>
- IVANOVA, A. and HOBERMAN, S. (2015). Higher order response-adaptive urn designs for clinical trials with highly successful treatments. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 175–189. MR3293924 <https://doi.org/10.1111/rssc.12066>
- KORN, E. L. and FREIDLIN, B. (2011). Outcome-adaptive randomization: Is it useful? *J. Clin. Oncol.* **29** 771–776. <https://doi.org/10.1200/JCO.2010.31.1423>
- MILES, L. F., BURT, C., ARROWSMITH, J., MCKIE, M. A., VILLAR, S. S., GOVENDER, P. and FALTER, F. (2021). Optimal pro-tamine dosing after cardiopulmonary bypass: The PRODOSE adaptive randomised controlled trial. *PLoS Med.* **18** e1003658.
- PROPER, J. and MURRAY, T. A. (2022). An alternative metric for evaluating the potential patient benefit of response-adaptive randomization procedures. *Biometrics*. To appear.
- PROSCHAN, M. and EVANS, S. (2020). Resist the temptation of response-adaptive randomization. *Clin. Infect. Dis.* **71** 3002–3004.
- ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci.* **11** 137–149.
- VILLAR, S. S. and ROSENBERGER, W. F. (2018). Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward looking Gittins index rule. *Biometrics* **74** 49–57. MR3777925 <https://doi.org/10.1111/biom.12738>
- XIAO, Y., LIU, Z. and HU, F. (2017). Bayesian doubly adaptive randomization in clinical trials. *Sci. China Math.* **60** 2503–2514. MR3736374 <https://doi.org/10.1007/s11425-016-0056-1>
- ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Ann. Statist.* **35** 1166–1182. MR2341702 <https://doi.org/10.1214/009053606000001424>

Comment: Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations

Yunshan Duan, Peter Müller and Yuan Ji

REFERENCES

- [1] BARTLETT, R. H., GAZZANIGA, A., TOOMASIAN, J., CORAN, A., ROLOFF, D., RUCKER, R. and CORWIN, A. (1986). Extracorporeal membrane oxygenation (ECMO) in neonatal respiratory failure. 100 cases. *Ann. Surg.* **204** 236.
- [2] BERNARDO, J.-M. and SMITH, A. F. M. (2009). *Bayesian Theory*. John Wiley & Sons. MR1274699 <https://doi.org/10.1002/9780470316870>
- [3] KORN, E. L. and FREIDLIN, B. (2011). Outcome-adaptive randomization: Is it useful? *J. Clin. Oncol.* **29** 771–776. <https://doi.org/10.1200/JCO.2010.31.1423>
- [4] MÜLLER, P., BERRY, D. A., GRIEVE, A. P., SMITH, M. and KRAMS, M. (2007). Simulation-based sequential Bayesian design. *J. Statist. Plann. Inference* **137** 3140–3150. MR2364157 <https://doi.org/10.1016/j.jspi.2006.05.021>
- [5] MÜLLER, P., DUAN, Y. and GARCIA TEC, M. (2022). Simulation-based sequential design. *Pharm. Stat.* **21** 729–739.
- [6] ROBERTSON, D. S., LEE, K. M., LOPEZ-KOLKOVSKA, B. C. and VILLAR, S. S. (2023). Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statist. Sci.* **38** 185–208.
- [7] THALL, P., FOX, P. and WATHEN, J. (2015). Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann. Oncol.* **26** 1621–1628. <https://doi.org/10.1093/annonc/mdv238>
- [8] VENTZ, S., CELLAMARE, M., BACALLADO, S. and TRIPPA, L. (2019). Bayesian uncertainty directed trial designs. *J. Amer. Statist. Assoc.* **114** 962–974. MR4011751 <https://doi.org/10.1080/01621459.2018.1497497>
- [9] XU, Y., TRIPPA, L., MÜLLER, P. and JI, Y. (2016). Subgroup-based adaptive (SUBA) designs for multi-arm biomarker trials. *Stat. Biosci.* **8** 159–180. <https://doi.org/10.1007/s12561-014-9117-1>

Comment: Advancing Clinical Trials with Novel Designs and Implementations

Lorenzo Trippa and Yanxun Xu

REFERENCES

- [1] ALEXANDER, B. M., TRIPPA, L., GAFFEY, S., ARRILLAGA-ROMANY, I. C., LEE, E. Q., RINNE, M. L., AHLUWALIA, M. S., COLMAN, H., FELL, G. et al. (2019). Individualized screening trial of innovative glioblastoma therapy (INSIGHT): A Bayesian adaptive platform trial to develop precision medicines for patients with glioblastoma. *JCO Precis. Oncol.* **3** 1–13.
- [2] BERRY, D. A. (2011). Adaptive clinical trials: The promise and the caution. *J. Clin. Oncol.* **29** 606–609.
- [3] BROGLIO, K., MEURER, W. J., DURKALSKI, V., PAULS, Q., CONNOR, J., BERRY, D., LEWIS, R. J., JOHNSTON, K. C. and BARSAN, W. G. (2022). Comparison of Bayesian vs frequentist adaptive trial design in the stroke hyperglycemia insulin network effort trial. *JAMA Netw. Open* **5** e2211616. <https://doi.org/10.1001/jamanetworkopen.2022.11616>
- [4] BROGLIO, K. R. and BERRY, D. A. (2009). Detecting an overall survival benefit that is derived from progression-free survival. *J. Natl. Cancer Inst.* **101** 1642–1649. <https://doi.org/10.1093/jnci/djp369>
- [5] CELLAMARE, M., VENTZ, S., BAUDIN, E., MITNICK, C. D. and TRIPPA, L. (2017). A Bayesian response-adaptive trial in tuberculosis: The endTB trial. *Clin. Trials* **14** 17–28. <https://doi.org/10.1177/1740774516665090>
- [6] CHENG, P. M., MONTAGNON, E., YAMASHITA, R., PAN, I., CADRIN-CHENEVERT, A., PERDIGÓN ROMERO, F., CHARTRAND, G., KADOURY, S. and TANG, A. (2021). Deep learning: An update for radiologists. *Radiographics* **41** 1427–1445.
- [7] FLIGHT, L., BRENNAN, A., CHICK, S. E., FORSTER, M., JULIOUS, S. and THARMANATHAN, P. (2022). Value-adaptive clinical trial designs for efficient delivery of research—actions, opportunities and challenges for publicly funded trials. Discussion paper, School of Health and Related Research, University of Sheffield.
- [8] GOLCHI, S. (2022). Estimating design operating characteristics in Bayesian adaptive clinical trials. *Canad. J. Statist.* **50** 417–436. [MR4429845 https://doi.org/10.1002/cjs.11699](https://doi.org/10.1002/cjs.11699)
- [9] GREEN, A. K., REEDER-HAYES, K. E., CORTY, R. W., BASCH, E., MILOWSKY, M. I., DUSETZINA, S. B., BENNETT, A. V. and WOOD, W. A. (2015). The project data sphere initiative: Accelerating cancer research by sharing data. *Oncol.* **20** 464–e20.
- [10] KOTECHA, G., VENTZ, S. and TRIPPA, L. (2023). Prospectively shared control data across concurrent randomised clinical trials. *Eur. J. Cancer* **181** 18–20. <https://doi.org/10.1016/j.ejca.2022.11.038>
- [11] LO, A. W. and THAKOR, R. T. (2022). Financing biomedical innovation. *Annu. Rev. Financ. Econ.* **14** 231–270.
- [12] ROBERTSON, D. S., LEE, K. M., LOPEZ-KOLKOVSKA, B. C. and VILLAR, S. S. (2023). Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statist. Sci.*
- [13] ROSS, J. S., WALDSTREICHER, J., BAMFORD, S., BERLIN, J. A., CHILDERS, K., DESAI, N. R., GAMBLE, G., GROSS, C. P., KUNTZ, R. et al. (2018). Overview and experience of the YODA project with clinical trial data sharing after 5 years. *Sci. Data* **5** 1–14.
- [14] VENTZ, S., ALEXANDER, B. M., PARMIGIANI, G., GELBER, R. D. and TRIPPA, L. (2017). Designing clinical trials that accept new arms: An example in metastatic breast cancer. *J. Clin. Oncol.* **35** 3160–3168. <https://doi.org/10.1200/JCO.2016.70.1169>
- [15] VENTZ, S., COMMENT, L., LOUV, B., RAHMAN, R., WEN, P. Y., ALEXANDER, B. M. and TRIPPA, L. (2022). The use of external control data for predictions and futility interim analyses in clinical trials. *Neuro-Oncol.* **24** 247–256. <https://doi.org/10.1093/neuonc/noab141>
- [16] VENTZ, S., LAI, A., CLOUGHESY, T. F., WEN, P. Y., TRIPPA, L. and ALEXANDER, B. M. (2019). Design and evaluation of an external control arm using prior clinical trials and real-world data. *Clin. Cancer Res.* **25** 4993–5001.
- [17] WANG, H. (2016). Response adaptive randomization using surrogate and primary endpoints. Ph.D. thesis, Virginia Commonwealth University, Richmond, VA.
- [18] XU, Y., CONSTANTINE, F., YUAN, Y. and PRITCHETT, Y. L. (2020). ASIED: A Bayesian adaptive subgroup-identification enrichment design. *J. Biopharm. Statist.* **30** 623–638.
- [19] XU, Y., MÜLLER, P., TSIMBERIDOU, A. M. and BERRY, D. (2019). A nonparametric Bayesian basket trial design. *Biom. J.* **61** 1160–1174. [MR4013340 https://doi.org/10.1002/bimj.201700162](https://doi.org/10.1002/bimj.201700162)

Comment: Group Sequential Designs with Response-Adaptive Randomisation

Christopher Jennison

Abstract. Group sequential Phase III trial designs enable early stopping for positive or negative study outcomes. Response-adaptive randomisation can be included in such designs with the sampling ratio in each group of subjects determined by the current treatment effect estimate. We demonstrate the potential of adaptive randomisation to reduce the number of patients receiving the inferior treatment, even when there is a delay in observing each patient's response. We also observe that using a fixed but unequal sampling ratio may offer a simpler way to achieve the same objectives.

Key words and phrases: Response-adaptive randomisation, group sequential test, inferior treatment number.

REFERENCES

- [1] HAMPSON, L. V. and JENNISON, C. (2013). Group sequential tests for delayed responses (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 3–54. [MR3008270](#) <https://doi.org/10.1111/j.1467-9868.2012.01030.x>
- [2] JENNISON, C., JOHNSTONE, I. M. and TURNBULL, B. W. (1982). Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *Statistical Decision Theory and Related Topics, III, Vol. 2* (West Lafayette, Ind., 1981) 55–86. Academic Press, New York. [MR0705308](#)
- [3] JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press/CRC, Boca Raton, FL. [MR1710781](#)
- [4] JENNISON, C. and TURNBULL, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Anal.* **20** 209–234. [MR1868774](#) <https://doi.org/10.1081/SQA-100107646>
- [5] PAMPALLONA, S. and TSIATIS, A. A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *J. Statist. Plann. Inference* **42** 19–35. [MR1309622](#) [https://doi.org/10.1016/0378-3758\(94\)90187-2](https://doi.org/10.1016/0378-3758(94)90187-2)

Comment: Is Response-Adaptive Randomization a “Good Thing” or Not in Clinical Trials? Why We Cannot Take Sides

Alessandra Giovagnoli

Abstract. The paper by Robertson et al. intends to provide “a unified, broad and fresh review of methodological and practical issues to consider” as a contribution to the ongoing debate concerning RAR in clinical trials. Simulations carried out by different authors seem to disprove its usefulness both for statistical inference and as a safeguard for the care of patients in the trial. I argue that the arguments brought forward so far are inconclusive, since the inferential considerations are sometimes incomplete or incorrect, and some simulation studies unconvincing. A Bayesian stand is very common, but often not fully understood.

Key words and phrases: Bayesian response-adaptive randomization, doubly adaptive, modelling and simulation, simply adaptive, Thompson’s sampling.

REFERENCES

- [1] ATKINSON, A. C. and BISWAS, A. (2005). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics* **61** 118–125. MR2135851 <https://doi.org/10.1111/j.0006-341X.2005.031002.x>
- [2] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2015). *Adaptive Designs for Sequential Treatment Allocation*. Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL. MR3616325
- [3] BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2015). Adaptive biased coins: Achieving better balance without compromising randomness. In *Modern Adaptive Randomized Clinical Trials* (O. Sverdlov, ed.). *Statistical and Practical Aspects Chapman and Hall/CRC Biostatistics Series* 55–82.
- [4] FLOURNOY, N., MAY, C. and SECCHI, P. (2012). Asymptotically optimal response-adaptive designs for allocating the best treatment: An overview. *Int. Stat. Rev.* **80** 293–305. MR2990356 <https://doi.org/10.1111/j.1751-5823.2011.00173.x>
- [5] GERALDES, M., MELFI, V., PAGE, C. and ZHANG, H. (2006). The doubly adaptive weighted differences design. *J. Statist. Plann. Inference* **136** 1923–1939. MR2255604 <https://doi.org/10.1016/j.jspi.2005.08.012>
- [6] GIOVAGNOLI, A. and VERDINELLI, I. (2023). Bayesian adaptive randomization with compound utility functions. *Statist. Sci.* **38** 52–67. MR4535394 <https://doi.org/10.1214/21-sts848>
- [7] GIOVAGNOLI, A. and ZAGORAIOU, M. (2012). Simulation of clinical trials: A review with emphasis on the design issues. *Statistica* **LXXII** 63–80.
- [8] PADILLA, J. J., DIALLO, S. Y. and TOLK, A. (2011). Do we need M&S science?. *SCS M&S Mag.* **4** 161–166.
- [9] ROYALL, R. M. (1991). Ethics and statistics in randomized clinical trials. *Statist. Sci.* **6** 52–88. MR1108816
- [10] XIAO, Y., LIU, Z. and HU, F. (2017). Bayesian doubly adaptive randomization in clinical trials. *Sci. China Math.* **60** 2503–2514. MR3736374 <https://doi.org/10.1007/s11425-016-0056-1>

Comment: Response Adaptive Randomization in Practice

Scott M. Berry and Kert Viele

Abstract. Robertson et al. provide an elegant, thorough and accessible overview of the rich theoretical background of response adaptive randomization. Many completed and ongoing clinical trials utilize these methods with demonstrated success. We provide a summary of multiple real world examples of response adaptive randomization, and a discussion of themes that arise in planning and executing response adaptive trials.

Key words and phrases: Response adaptive randomization, adaptive clinical trials, seamless clinical trials.

REFERENCES

- [1] ALLAREDDY, V. and CHEIFETZ, I. M. (2019). Clinical trials and future directions in pediatric acute respiratory distress syndrome. *Ann. Transl. Med.* **7** 514. <https://doi.org/10.21037/atm.2019.09.14>
- [2] JONES, A. E., PUSKARICH, M. A., SHAPIRO, N. I. et al. (2018). Effect of levocarnitine vs placebo as an adjunctive treatment for septic shock: The rapid administration of carnitine in sepsis (RACE) randomized clinical trial. *JAMA Netw. Open* **1** e186076. <https://doi.org/10.1001/jamanetworkopen.2018.6076>
- [3] LEWIS, R. L., ANGUS, D. A., LATERRE, P. F., KJOLBYE, A. L., VAN DER MEULEN, E., BLEMINGS, A., GRAVES, T., RUSSELL, J. A., CARLSON, J. E. et al. (2018). Rationale and design of an adaptive phase 2b/3 clinical trial of selepressin for adults in septic shock. Selepressin evaluation programme for sepsis-induced shock-adaptive clinical trial. *Ann. Amer. Thorac. Soc.* **15** 250–257.
- [4] SKRIVANEK, Z., BERRY, S., BERRY, D., CHIEN, J., GEIGER, M. J., ANDERSON, J. H. and GAYDOS, B. (2012). Application of adaptive design methodology in development of a long-acting glucagon-like peptide-1 analog (dulaglutide): Statistical design and simulations. *J. Diabetes Sci. Technol.* **6** 1305–1318. <https://doi.org/10.1177/193229681200600609>
- [5] SWANSON, C. J., ZHANG, Y., DHADDA, S. et al. (2021). A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-A β protofibril antibody. *Alzheimer's Res. Ther.* **13** 80. <https://doi.org/10.1186/s13195-021-00813-8>
- [6] <https://investor.lilly.com/news-releases/news-release-details/lilly-reports-solid-fourth-quarter-and-full-year-2021-financial>.
- [7] <https://siren.network/clinical-trials/icecap>.

Rejoinder: Response-Adaptive Randomization in Clinical Trials

David S. Robertson, Kim May Lee, Boryana C. López-Kolkovska and Sofía S. Villar

REFERENCES

- BERGER, V. W., BOUR, L. J., CARTER, K., CHIPMAN, J. J., EVERETT, C. C., HEUSSEN, N., HEWITT, C., HILGERS, R., LUO, Y. A. et al. (2021). A roadmap to using randomization in clinical trials. *BMC Med. Res. Methodol.* **21** 1–24.
- BERRY, D. A. (1978). Modified two-armed bandit strategies for certain clinical trials. *J. Amer. Statist. Assoc.* **73** 339–345.
- CHENG, Y. and BERRY, D. A. (2007). Optimal adaptive randomized designs for clinical trials. *Biometrika* **94** 673–689. MR2410016 <https://doi.org/10.1093/biomet/asm049>
- HEINZE, G., BOULESTEIX, A., KAMMER, M., MORRIS, T. P. and WHITE, I. R. (2022). Phases of methodological research in biostatistics-building the evidence base for new methods. *Biom. J.* <https://doi.org/10.1002/bimj.202200222>
- JACKO, P. (2019). The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths. arXiv preprint. arXiv:1906.10173.
- JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press/CRC, Boca Raton, FL. MR1710781
- LAW, M., COUTURIER, D., CHOODARI-OSKOEI, B., CROUT, P., GAMBLE, C., PALLMANN, P., PILLING, M., ROBERTSON, D. S., ROBLING, M., SYDES, M. R., VILLAR, S. S., WASON, J. M. S., WHEELER, G., WILLIAMSON, S. F., YAP, C. and JAKI, T. (2022). Medicines and healthcare products regulatory agency's "Consultation on proposals for legislative changes for clinical trials". Preprint. <https://doi.org/10.21203/rs.3.rs-2210654/v1>
- MAYER, C., PEREVOZSKAYA, I., LEONOV, S., DRAGALIN, V., PRITCHETT, Y., BEDDING, A., HARTFORD, A., FARDIPOUR, P. and CICONETTI, G. (2019). Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Stat. Biopharm. Res.* **11** 325–335.
- MORRIS, T. P., WHITE, I. R. and CROWTHER, M. J. (2019). Using simulation studies to evaluate statistical methods. *Stat. Med.* **38** 2074–2102. MR3937487 <https://doi.org/10.1002/sim.8086>
- MOZGUNOV, P., JAKI, T. and PAOLETTI, X. (2020). A benchmark for dose finding studies with continuous outcomes. *Biostatistics* **21** 189–201. MR4132543 <https://doi.org/10.1093/biostatistics/kxy045>
- MOZGUNOV, P., PAOLETTI, X. and JAKI, T. (2022). A benchmark for dose-finding studies with unknown ordering. *Biostatistics* **23** 721–737. MR4454951 <https://doi.org/10.1093/biostatistics/kxaa054>
- UMA, L. O., GRAYLING, M. J., WASON, J. M. S. and ZHENG, H. (2022). Bayesian modelling strategies for borrowing of information in randomised basket trials. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **71** 2014–2037. MR4511139 <https://doi.org/10.1111/rssc.12602>
- PAWEL, S., KOOK, L. and REEVE, K. (2023). Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biom. J.* <https://doi.org/10.1002/bimj.202200091>
- PITT, L. (2021). Optimising first in human trials. PhD Thesis, Univ. Bath. Available at <https://researchportal.bath.ac.uk/en/studentTheses/optimising-first-in-human-trials>.
- PRESS, W. H. (2009). Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proc. Natl. Acad. Sci. USA* **106** 22387–22392.
- ROBERTSON, D. S., CHOODARI-OSKOEI, B., DIMAIRO, M., FLIGHT, L., PALLMANN, P. and JAKI, T. (2023a). Point estimation for adaptive trial designs I: A methodological review. *Stat. Med.* **42** 122–145. MR4527724 <https://doi.org/10.1002/sim.9605>
- ROBERTSON, D. S., CHOODARI-OSKOEI, B., DIMAIRO, M., FLIGHT, L., PALLMANN, P. and JAKI, T. (2023b). Point estimation after adaptive trials II: Practical considerations and guidance. *Stat. Med.* <https://doi.org/10.1002/sim.9734>
- ROBERTSON, D. S., LÓPEZ-KOLKOVSKA, B. C., LEE, K. M. and VILLAR, S. S. (2023c). Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statist. Sci.* **38** 185–208.
- ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci.* **11** 137–149.
- SHIN, J., RAMDAS, A. and RINALDO, A. (2019). Are sample means in multi-armed bandits positively or negatively biased? *Adv. Neural Inf. Process. Syst.* **32**.
- SKLAR, M. J. (2022). Adaptive experiments and a rigorous framework for type I error verification and computational experiment design. PhD Thesis, Stanford Univ. arXiv:2205.09369.
- SMITH, A. L. and VILLAR, S. S. (2018). Bayesian adaptive bandit-based designs using the Gittins index for multi-armed trials with normally distributed endpoints. *J. Appl. Stat.* **45** 1052–1076. MR3774531 <https://doi.org/10.1080/02664763.2017.1342780>
- VENTZ, S., BARRY, W. T., PARMIGIANI, G. and TRIPPA, L. (2017). Bayesian response-adaptive designs for basket trials. *Biometrics* **73** 905–915. MR3713124 <https://doi.org/10.1111/biom.12668>
- VILLAR, S. S., BOWDEN, J. and WASON, J. (2015). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statist. Sci.* **30** 199–215. MR3353103 <https://doi.org/10.1214/14-STS504>
- VILLAR, S. S., WASON, J. and BOWDEN, J. (2015). Response-adaptive randomization for multi-arm clinical trials using the forward looking Gittins index rule. *Biometrics* **71** 969–978. MR3436722 <https://doi.org/10.1111/biom.12337>
- WILLIAMSON, S. F., JACKO, P., VILLAR, S. S. and JAKI, T. (2017). A Bayesian adaptive design for clinical trials in rare dis-

David S. Robertson is a Senior Research Associate, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (e-mail: david.robertson@mrc-bsu.cam.ac.uk). Kim May Lee is a Research Fellow, King's College London, London, UK (e-mail: kim.lee@kcl.ac.uk). Boryana C. López-Kolkovska is a Statistical Science Associate Director at AstraZeneca, Cambridge, UK (e-mail: boryana.kolkovska@astrazeneca.com). Sofía S. Villar is an MRC Investigator, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (e-mail: sofia.villar@mrc-bsu.cam.ac.uk).

eases. *Comput. Statist. Data Anal.* **113** 136–153. MR3662397
<https://doi.org/10.1016/j.csda.2016.09.006>

XIAO, Y., LIU, Z. and HU, F. (2017). Bayesian doubly adaptive randomization in clinical trials. *Sci. China Math.* **60** 2503–2514. MR3736374 <https://doi.org/10.1007/s11425-016-0056-1>

ZHENG, H. and WASON, J. M. S. (2022). Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics* **23** 120–135. MR4366039 <https://doi.org/10.1093/biostatistics/kxaa019>

Efficient Data Augmentation Techniques for Some Classes of State Space Models

Linda S. L. Tan

Abstract. Data augmentation improves the convergence of iterative algorithms, such as the EM algorithm and Gibbs sampler by introducing carefully designed latent variables. In this article, we first propose a data augmentation scheme for the first-order autoregression plus noise model, where optimal values of working parameters introduced for recentering and rescaling of the latent states, can be derived analytically by minimizing the fraction of missing information in the EM algorithm. The proposed data augmentation scheme is then utilized to design efficient Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference of some non-Gaussian and non-linear state space models, via a mixture of normals approximation coupled with a block-specific reparametrization strategy. Applications on simulated and benchmark real data sets indicate that the proposed MCMC sampler can yield improvements in simulation efficiency compared with centering, non-centering and even the ancillarity-sufficiency interweaving strategy.

Key words and phrases: Data augmentation, state space model, stochastic volatility model, EM algorithm, reparametrization, Markov chain Monte Carlo, ancillarity-sufficiency interweaving strategy.

REFERENCES

- ABANTO-VALLE, C. A. and DEY, D. K. (2014). State space mixed models for binary responses with scale mixture of normal distributions links. *Comput. Statist. Data Anal.* **71** 274–287. MR3131970 <https://doi.org/10.1016/j.csda.2013.01.009>
- ALMEIDA, C. and CZADO, C. (2012). Efficient Bayesian inference for stochastic time-varying copula models. *Comput. Statist. Data Anal.* **56** 1511–1527. MR2892357 <https://doi.org/10.1016/j.csda.2011.08.015>
- ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 <https://doi.org/10.1111/j.1467-9868.2009.00736.x>
- BASS, M. R. and SAHU, S. K. (2017). A comparison of centering parameterisations of Gaussian process-based models for Bayesian computation using MCMC. *Stat. Comput.* **27** 1491–1512. MR3687322 <https://doi.org/10.1007/s11222-016-9700-z>
- BAUWENS, L. and VEREDAS, D. (2004). The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *J. Econometrics* **119** 381–412. MR2057105 [https://doi.org/10.1016/S0304-4076\(03\)00201-X](https://doi.org/10.1016/S0304-4076(03)00201-X)
- BERLINET, A. and ROLAND, C. (2009). Parabolic acceleration of the EM algorithm. *Stat. Comput.* **19** 35–47. MR2470556 <https://doi.org/10.1007/s11222-008-9067-x>
- BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Rev.* **59** 65–98. MR3605826 <https://doi.org/10.1137/141000671>
- BITTO, A. and FRÜHWIRTH-SCHNATTER, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *J. Econometrics* **210** 75–97. MR3944764 <https://doi.org/10.1016/j.jeconom.2018.11.006>
- BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 <https://doi.org/10.1080/01621459.2017.1285773>
- CARTER, C. K. and KOHN, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *J. Roy. Statist. Soc. Ser. B* **59** 255–268. MR1436567 <https://doi.org/10.1111/1467-9868.00067>
- CARVALHO, C. M., JOHANNES, M. S., LOPES, H. F. and POLSON, N. G. (2010). Particle learning and smoothing. *Statist. Sci.* **25** 88–106. MR2741816 <https://doi.org/10.1214/10-STS325>
- CHRISTENSEN, O. F., ROBERTS, G. O. and SKÖLD, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *J. Comput. Graph. Statist.* **15** 1–17. MR2269360 <https://doi.org/10.1198/106186006X100470>
- CRYER, J. D. and CHAN, K. S. (2008). Time series analysis with applications in R. Springer, New York.
- DE VALPINE, P., TUREK, D., PACIOREK, C. J., ANDERSON-BERGMAN, C., TEMPLE LANG, D. and BODIK, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Statist.* **26** 403–413. MR3640196 <https://doi.org/10.1080/10618600.2016.1172487>
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501577

- DOUCET, A., DE FREITAS, N. and GORDON, N. (2012). Sequential Monte Carlo methods in practice. Springer, New York.
- ENGLER, R. F. and RUSSELL, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66** 1127–1162. MR1639411 <https://doi.org/10.2307/2999632>
- FEARNHEAD, P. and MELIGKOTSIDOU, L. (2016). Augmentation schemes for particle MCMC. *Stat. Comput.* **26** 1293–1306. MR3538638 <https://doi.org/10.1007/s11222-015-9603-4>
- FENG, D., JIANG, G. J. and SONG, P. X. K. (2004). Stochastic conditional duration models with “leverage effect” for financial transaction data. *J. Financ. Econ.* **2** 390–421. <https://doi.org/10.1093/jffinec/nbh016>
- FRÜHWIRTH-SCHNATTER, S. (2004). Efficient Bayesian parameter estimation. In *State Space and Unobserved Component Models* 123–151. Cambridge Univ. Press, Cambridge. MR2077636 <https://doi.org/10.1017/CBO9780511617010.008>
- FRÜHWIRTH-SCHNATTER, S. and FRÜHWIRTH, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Comput. Statist. Data Anal.* **51** 3509–3528. MR2367817 <https://doi.org/10.1016/j.csda.2006.10.006>
- FRÜHWIRTH-SCHNATTER, S. and WAGNER, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *J. Econometrics* **154** 85–100. MR2558953 <https://doi.org/10.1016/j.jeconom.2009.07.003>
- GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika* **82** 479–488. MR1366275 <https://doi.org/10.1093/biomet/82.3.479>
- GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1996). Efficient parametrizations for generalized linear mixed models. In *Bayesian Statistics, 5 (Alicante, 1994)* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). *Oxford Sci. Publ.* 165–180. Oxford Univ. Press, New York. MR1425405
- GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6** 721–741.
- GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 <https://doi.org/10.1111/j.1467-9868.2010.00765.x>
- GOPLERUD, M. (2022). Fast and accurate estimation of non-nested binomial hierarchical models using variational inference. *Bayesian Anal.* **17** 623–650. <https://doi.org/10.1214/21-BA1266>
- HARVEY, A., RUIZ, E. and SHEPHARD, N. (1994). Multivariate stochastic variance models. *Rev. Econ. Stud.* **61** 247–264.
- HENDERSON, N. C. and VARADHAN, R. (2019). Damped Anderson acceleration with restarts and monotonicity control for accelerating EM and EM-like algorithms. *J. Comput. Graph. Statist.* **28** 834–846. MR4045852 <https://doi.org/10.1080/10618600.2019.1594835>
- HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779
- HOSSZEJNI, D. and KASTNER, G. (2021). Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *J. Stat. Softw.* **100** 1–34.
- JAMSHIDIAN, M. and JENNRICH, R. I. (1997). Acceleration of the EM algorithm by using quasi-Newton methods. *J. Roy. Statist. Soc. Ser. B* **59** 569–587. MR1452026 <https://doi.org/10.1111/1467-9868.00083>
- KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82** 35–45. MR3931993
- KANTAS, N., DOUCET, A., SINGH, S. S., MACIEJOWSKI, J. and CHOPIN, N. (2015). On particle methods for parameter estimation in state-space models. *Statist. Sci.* **30** 328–351. MR3383884 <https://doi.org/10.1214/14-STSS11>
- KASTNER, G. and FRÜHWIRTH-SCHNATTER, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Comput. Statist. Data Anal.* **76** 408–423. MR3209449 <https://doi.org/10.1016/j.csda.2013.01.002>
- KASTNER, G., FRÜHWIRTH-SCHNATTER, S. and LOPES, H. F. (2017). Efficient Bayesian inference for multivariate factor stochastic volatility models. *J. Comput. Graph. Statist.* **26** 905–917. MR3765354 <https://doi.org/10.1080/10618600.2017.1322091>
- KIM, S., SHEPHARD, N. and CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65** 361–393.
- KLEPPE, T. S. (2019). Dynamically rescaled Hamiltonian Monte Carlo for Bayesian hierarchical models. *J. Comput. Graph. Statist.* **28** 493–507. MR4007736 <https://doi.org/10.1080/10618600.2019.1584901>
- KREUZER, A. and CZADO, C. (2020). Efficient Bayesian inference for nonlinear state space models with univariate autoregressive state equation. *J. Comput. Graph. Statist.* **29** 523–534. MR4153179 <https://doi.org/10.1080/10618600.2020.1725523>
- KROESE, D. P. and CHAN, J. C. C. (2014). *Statistical Modeling and Computation*. Springer, New York. MR3155490 <https://doi.org/10.1007/978-1-4614-8775-3>
- LI, M. and SCHARTH, M. (2022). Leverage, asymmetry, and heavy tails in the high-dimensional factor stochastic volatility model. *J. Bus. Econom. Statist.* **40** 285–301. MR4356573 <https://doi.org/10.1080/07350015.2020.1806853>
- LIU, C., RUBIN, D. B. and WU, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85** 755–770. MR1666758 <https://doi.org/10.1093/biomet/85.4.755>
- MENG, X.-L. and RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278. MR1243503 <https://doi.org/10.1093/biomet/80.2.267>
- MENG, X.-L. and VAN DYK, D. (1997). The EM algorithm—An old folk-song sung to a fast new tune. *J. Roy. Statist. Soc. Ser. B* **59** 511–567. MR1452025 <https://doi.org/10.1111/1467-9868.00082>
- MENG, X.-L. and VAN DYK, D. (1998). Fast EM-type implementations for mixed effects models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 559–578. MR1625942 <https://doi.org/10.1111/1467-9868.00140>
- MICHAUD, N., DE VALPINE, P., TUREK, D., PACIOREK, C. J. and NGUYEN, D. (2020). Sequential Monte Carlo methods in the nimble R package.
- MOGENSEN, P. K. and RISETH, A. N. (2018). Optim: A mathematical optimization package for Julia. *J. Open Sour. Softw.* **3** 615. <https://doi.org/10.21105/joss.00615>
- NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447
- OLSSON, R. K. and HANSEN, L. K. (2006). Linear state-space models for blind source separation. *J. Mach. Learn. Res.* **7** 2585–2602. MR2274451
- OMORI, Y., CHIB, S., SHEPHARD, N. and NAKAJIMA, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *J. Econometrics* **140** 425–449. MR2408914 <https://doi.org/10.1016/j.jeconom.2006.07.008>
- ORMEROD, J. T. and WAND, M. P. (2010). Explaining variational approximations. *Amer. Statist.* **64** 140–153. MR2757005 <https://doi.org/10.1198/tast.2010.09058>
- OSMUNDSEN, K. K., KLEPPE, T. S. and LIESENFELD, R. (2021). Importance sampling-based transport map Hamiltonian Monte Carlo for Bayesian hierarchical models. *J. Comput. Graph. Statist.*

- 30 906–919. MR4356594 <https://doi.org/10.1080/10618600.2021.1923519>
- PAL, A. and PRAKASH, P. (2017). Practical time series analysis. Packt Publishing, Birmingham, Mumbai.
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2003). Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian Statistics, 7 (Tenerife, 2002)* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 307–326. Oxford Univ. Press, New York. MR2003180
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Statist. Sci.* **22** 59–73. MR2408661 <https://doi.org/10.1214/088342307000000014>
- PITT, M. K. and SHEPHARD, N. (1999a). Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models. *J. Time Series Anal.* **20** 63–85. MR1678569 <https://doi.org/10.1111/1467-9892.00126>
- PITT, M. K. and SHEPHARD, N. (1999b). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328 <https://doi.org/10.2307/2670179>
- PITT, M. K. and SHEPHARD, N. (1999c). Time-varying covariances: A factor stochastic volatility approach. In *Bayesian Statistics, 6 (Alcoceber, 1998)* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 547–570. Oxford Univ. Press, New York. MR1724873
- PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.
- ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 <https://doi.org/10.1111/1467-9868.00070>
- SAÂDAOUI, F. (2010). Acceleration of the EM algorithm via extrapolation methods: Review, comparison and new methods. *Comput. Statist. Data Anal.* **54** 750–766. MR2744430 <https://doi.org/10.1016/j.csda.2008.11.011>
- SAHU, S. K. and ROBERTS, G. O. (1999). On convergence of the EM algorithm and the Gibbs sampler. *Stat. Comput.* **9** 55–64.
- SHUMWAY, R. H. and STOFFER, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Anal.* **3** 253–264.
- SHUMWAY, R. H. and STOFFER, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*, 4th ed. Springer Texts in Statistics. Springer, Cham. MR3642322 <https://doi.org/10.1007/978-3-319-52452-8>
- SIMPSON, M., NIEMI, J. and ROY, V. (2017). Interweaving Markov chain Monte Carlo strategies for efficient estimation of dynamic linear models. *J. Comput. Graph. Statist.* **26** 152–159. MR3610416 <https://doi.org/10.1080/10618600.2015.1105748>
- STRICKLAND, C. M., FORBES, C. S. and MARTIN, G. M. (2006). Bayesian analysis of the stochastic conditional duration model. *Comput. Statist. Data Anal.* **50** 2247–2267. MR2225569 <https://doi.org/10.1016/j.csda.2005.07.005>
- TAK, H., YOU, K., GHOSH, S. K., SU, B. and KELLY, J. (2020). Data transforming augmentation for heteroscedastic models. *J. Comput. Graph. Statist.* **29** 659–667. MR4153189 <https://doi.org/10.1080/10618600.2019.1704295>
- TAN, L. S. L. (2019). Explicit inverse of tridiagonal matrix with applications in autoregressive modelling. *IMA J. Appl. Math.* **84** 679–695. MR4011557 <https://doi.org/10.1093/imamat/hxz010>
- TAN, L. S. L. (2021). Use of model reparametrization to improve variational Bayes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 30–57. MR4220983 <https://doi.org/10.1111/rssb.12399>
- TAN, L. S. L. (2023). Supplement to “Efficient data augmentation techniques for some classes of state space models.” <https://doi.org/10.1214/22-ST867SUPP>
- TAN, L. S. L. and NOTT, D. J. (2013). Variational inference for generalized linear mixed models using partially noncentered parametrizations. *Statist. Sci.* **28** 168–188. MR3112404 <https://doi.org/10.1214/13-sts418>
- TAN, S. L. and NOTT, D. J. (2014). Variational approximation for mixtures of linear mixed models. *J. Comput. Graph. Statist.* **23** 564–585. MR3215825 <https://doi.org/10.1080/10618600.2012.761138>
- TAN, M., TIAN, G.-L., FANG, H.-B. and NG, K. W. (2007). A fast EM algorithm for quadratic optimization subject to convex constraints. *Statist. Sinica* **17** 945–964. MR2408642
- TAYLOR, S. J. (1982). Financial returns modelled by the product of two stochastic processes—A study of daily sugar prices, 1961–79. In *Time Series Analysis: Theory and Practice, Vol. 1* (O. D. Anderson, ed.) 203–226. Elsevier, North-Holland, Amsterdam.
- STAN DEVELOPMENT TEAM (2019). Stan modeling language users guide and reference manual. Version 2.28.
- R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- TSAY, R. S. (2010). *Analysis of Financial Time Series*, 3rd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. MR2778591 <https://doi.org/10.1002/9780470644560>
- VAN DYK, D. A. and MENG, X.-L. (2001). The art of data augmentation. *J. Comput. Graph. Statist.* **10** 1–50. MR1936358 <https://doi.org/10.1198/10618600152418584>
- WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103. MR0684867 <https://doi.org/10.1214/aos/1176346060>
- YANG, B., STROUD, J. R. and HUERTA, G. (2018). Sequential Monte Carlo smoothing with parameter estimation. *Bayesian Anal.* **13** 1133–1157. MR3855366 <https://doi.org/10.1214/17-BA1088>
- YU, Y. and MENG, X.-L. (2011). To center or not to center: That is not the question—An ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *J. Comput. Graph. Statist.* **20** 531–570. MR2878987 <https://doi.org/10.1198/jcgs.2011.203main>
- ZANELLA, G. and ROBERTS, G. (2021). Multilevel linear models, Gibbs samplers and multigrid decompositions. *Bayesian Anal.* **16** 1308–1390. MR4381136 <https://doi.org/10.1214/20-BA1242>
- ZHOU, H., ALEXANDER, D. and LANGE, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Stat. Comput.* **21** 261–273. MR2774856 <https://doi.org/10.1007/s11222-009-9166-3>
- ZHOU, L. and TANG, Y. (2021). Linearly preconditioned nonlinear conjugate gradient acceleration of the PX-EM algorithm. *Comput. Statist. Data Anal.* **155** Paper No. 107056, 13 pp. MR4161784 <https://doi.org/10.1016/j.csda.2020.107056>

Distributed Bayesian Inference in Massive Spatial Data

Rajarshi Guhaniyogi, Cheng Li, Terrance Savitsky and Sanvesh Srivastava

Abstract. Gaussian process (GP) regression is computationally expensive in spatial applications involving massive data. Various methods address this limitation, including a small number of Bayesian methods based on distributed computations (or the divide-and-conquer strategy). Focusing on the latter literature, we achieve three main goals. First, we develop an extensible Bayesian framework for distributed spatial GP regression that embeds many popular methods. The proposed framework has three steps that partition the entire data into many subsets, apply a readily available Bayesian spatial process model in parallel on all the subsets, and combine the posterior distributions estimated on all the subsets into a pseudo posterior distribution that conditions on the entire data. The combined pseudo posterior distribution replaces the full data posterior distribution in prediction and inference problems. Demonstrating our framework's generality, we extend posterior computations for (nondistributed) spatial process models with a stationary full-rank and a nonstationary low-rank GP priors to the distributed setting. Second, we contrast the empirical performance of popular distributed approaches with some widely-used, nondistributed alternatives and highlight their relative advantages and shortcomings. Third, we provide theoretical support for our numerical observations and show that the Bayes L_2 -risks of the combined posterior distributions obtained from a subclass of the divide-and-conquer methods achieves the near-optimal convergence rate in estimating the true spatial surface with various types of covariance functions. Additionally, we provide upper bounds on the number of subsets to achieve these near-optimal rates.

Key words and phrases: Distributed Bayesian inference, Gaussian process, low-rank Gaussian process, massive spatial data, Wasserstein barycenter.

REFERENCES

- AGUEH, M. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182 <https://doi.org/10.1137/100805741>
- ANDERES, E., HUSER, R., NYCHKA, D. and CORAM, M. (2013). Nonstationary positive definite tapering on the plane. *J. Comput. Graph. Statist.* **22** 848–865. MR3173746 <https://doi.org/10.1080/10618600.2012.729982>
- ANDERSON, C., LEE, D. and DEAN, N. (2014). Identifying clusters in Bayesian disease mapping. *Biostatistics* **15** 457–469.
- BAI, Y., SONG, P. X.-K. and RAGHUNATHAN, T. E. (2012). Joint composite estimating functions in spatiotemporal models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 799–824. MR2988907 <https://doi.org/10.1111/j.1467-9868.2012.01035.x>
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 <https://doi.org/10.1111/j.1467-9868.2008.00663.x>
- BARBIAN, M. H. and ASSUNÇÃO, R. M. (2017). Spatial subensemble estimator for large geostatistical data. *Spat. Stat.* **22** 68–88. MR3726176 <https://doi.org/10.1016/j.spasta.2017.08.004>

Rajarshi Guhaniyogi is Associate Professor, Department of Statistics, Texas A & M University, College Station, Texas 77843, USA (e-mail: rajguhaniyogi@tamu.edu). Cheng Li is Assistant Professor, Department of Statistics and Data Science, National University of Singapore, Singapore (e-mail: stalic@nus.edu.sg). Terrance Savitsky is Senior Research Mathematical Statistician, U.S. Bureau of Labor Statistics, Washington D.C. 20212, USA (e-mail: savitsky.terrance@bls.gov). Sanvesh Srivastava is Associate Professor, Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, Iowa 52240, USA (e-mail: sanvesh-srivastava@uiowa.edu).

- BERLINER, L. M., WIKLE, C. K. and CRESSIE, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate* **13** 3953–3968.
- BEVILACQUA, M., CAAMAÑO-CARRILLO, C. and PORCU, E. (2022). Unifying compactly supported and Matérn covariance functions in spatial statistics. *J. Multivariate Anal.* **189** Paper No. 104949, 17. MR4375672 <https://doi.org/10.1016/j.jmva.2022.104949>
- BEVILACQUA, M. and GAETAN, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Stat. Comput.* **25** 877–892. MR3375623 <https://doi.org/10.1007/s11222-014-9460-6>
- BOLIN, D. and LINDGREN, F. (2013). A comparison between Markov approximations and other methods for large spatial data sets. *Comput. Statist. Data Anal.* **61** 7–21. MR3062997 <https://doi.org/10.1016/j.csda.2012.11.011>
- BOLIN, D. and WALLIN, J. (2020). Multivariate type G Matérn stochastic partial differential equation random fields. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 215–239. MR4060983
- CHANDLER, R. E. and BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94** 167–183. MR2367830 <https://doi.org/10.1093/biomet/asm015>
- CHENG, G. and SHANG, Z. (2017). Computational limits of divide-and-conquer method. *J. Mach. Learn. Res.* **18** 1–37.
- CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank Kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639 <https://doi.org/10.1111/j.1467-9868.2007.00633.x>
- CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics.* Wiley, Hoboken, NJ. MR2848400
- CUTURI, M. and DOUCET, A. (2014). Fast computation of Wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning, JMLR W&CP* **32**.
- DALEY, D. J., PORCU, E. and BEVILACQUA, M. (2015). Classes of compactly supported covariance functions for multivariate random fields. *Stoch. Environ. Res. Risk Assess.* **29** 1249–1263.
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706 <https://doi.org/10.1080/01621459.2015.1044091>
- DI LORENZO, E., SCHNEIDER, N., COBB, K., FRANKS, P., CHHAK, K., MILLER, A., MCWILLIAMS, J., BOGRAD, S., ARANGO, H. et al. (2008). North Pacific gyre oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.* **35**.
- EIDSVIK, J., SHABY, B. A., REICH, B. J., WHEELER, M. and NIEMI, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.* **23** 295–315. MR3215812 <https://doi.org/10.1080/10618600.2012.760460>
- FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. MR2667597 <https://doi.org/10.1016/j.csda.2008.09.008>
- FINLEY, A. O., DATTA, A., COOK, B. D., MORTON, D. C., ANDERSEN, H. E. and BANERJEE, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *J. Comput. Graph. Statist.* **28** 401–414. MR3974889 <https://doi.org/10.1080/10618600.2018.1537924>
- FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261 <https://doi.org/10.1198/106186006X132178>
- GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* **24** 561–578. MR3357395 <https://doi.org/10.1080/10618600.2014.914442>
- GRAMACY, R. B. and HAALAND, B. (2016). Speeding up neighborhood search in local Gaussian process prediction. *Technometrics* **58** 294–303. MR3520659 <https://doi.org/10.1080/00401706.2015.1027067>
- GUHANIYOGI, R. and BANERJEE, S. (2018). Meta-Kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics* **60** 430–444. MR3878099 <https://doi.org/10.1080/00401706.2018.1437474>
- GUHANIYOGI, R. and BANERJEE, S. (2019). Multivariate spatial meta Kriging. *Statist. Probab. Lett.* **144** 3–8. MR3906771 <https://doi.org/10.1016/j.spl.2018.04.017>
- GUHANIYOGI, R. and SANJO, B. (2020). Large multi-scale spatial modeling using tree shrinkage priors. *Statist. Sinica* **30** 2023–2050. MR4260754 <https://doi.org/10.5705/ss.20>
- GUHANIYOGI, R., FINLEY, A. O., BANERJEE, S. and GELFAND, A. E. (2011). Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics* **22** 997–1007. MR2861573 <https://doi.org/10.1002/env.1131>
- GUHANIYOGI, R., LI, C., SAVITSKY, T. and SRIVASTAVA, S. (2023). Supplement to “Distributed Bayesian Inference in Massive Spatial Data.” <https://doi.org/10.1214/22-ST868SUPP>
- GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60** 415–429. MR3878098 <https://doi.org/10.1080/00401706.2018.1437476>
- GUINNESS, J. (2021). Gaussian process learning via Fisher scoring of Vecchia’s approximation. *Stat. Comput.* **31** Paper No. 25, 8. MR4224950 <https://doi.org/10.1007/s11222-021-09999-1>
- HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician’s Perspective.* Springer, New York. MR1467237 <https://doi.org/10.1007/b98818>
- HAZRA, A. and HUSER, R. (2021). Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. *Ann. Appl. Stat.* **15** 572–596. MR4298957 <https://doi.org/10.1214/20-aos1418>
- HEATON, M. J., CHRISTENSEN, W. F. and TERRES, M. A. (2017). Nonstationary Gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics* **59** 93–101. MR3604192 <https://doi.org/10.1080/00401706.2015.1102763>
- HEATON, M. J., DATTA, A., FINLEY, A., FURRER, R., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D., KATZFUSS, M. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.
- ILLIAN, J. B., SØRBYE, S. H. and RUE, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **6** 1499–1530. MR3058673 <https://doi.org/10.1214/11-AOAS530>
- KATZFUSS, M. (2017). A multi-resolution approximation for massive spatial datasets. *J. Amer. Statist. Assoc.* **112** 201–214. MR3646566 <https://doi.org/10.1080/01621459.2015.1123632>
- KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. MR4194207 <https://doi.org/10.1214/19-STS755>
- KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 <https://doi.org/10.1198/016214508000000959>
- KNORR-HELD, L. and RASSER, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56** 13–21.
- LEMONS, R. T. and SANSÓ, B. (2009). A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature. *J. Amer. Statist. Assoc.* **104** 5–18. MR2662306 <https://doi.org/10.1198/jasa.2009.0018>

- LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. MR3694589 <https://doi.org/10.1093/biomet/asx033>
- LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 <https://doi.org/10.1111/j.1467-9868.2011.00777.x>
- LINDSTEN, F., JOHANSEN, A. M., NAESSETH, C. A., KIRKPATRICK, B., SCHÖN, T. B., ASTON, J. A. D. and BOUCHARDCÔTÉ, A. (2017). Divide-and-conquer with sequential Monte Carlo. *J. Comput. Graph. Statist.* **26** 445–458. MR3640200 <https://doi.org/10.1080/10618600.2016.1237363>
- MEHROTRA, S., BRANTLEY, H., WESTMAN, J., BANGERTER, L. and MAITY, A. (2021). Divide-and-Conquer MCMC for Multivariate Binary Data. Preprint. Available at [arXiv:2102.09008](https://arxiv.org/abs/2102.09008).
- MINSKER, S. (2019). Distributed statistical estimation and rates of convergence in normal approximation. *Electron. J. Stat.* **13** 5213–5252. MR4043072 <https://doi.org/10.1214/19-EJS1647>
- MINSKER, S., SRIVASTAVA, S., LIN, L. and DUNSON, D. (2014). Scalable and robust Bayesian inference via the median posterior. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)* 1656–1664.
- MINSKER, S., SRIVASTAVA, S., LIN, L. and DUNSON, D. B. (2017). Robust and scalable Bayes via a median of subset posterior measures. *J. Mach. Learn. Res.* **18** Paper No. 124, 40. MR3763758
- NEISWANGER, W., WANG, C. and XING, E. (2014). Asymptotically exact, embarrassingly parallel MCMC. In *Proceedings of the 30th International Conference on Uncertainty in Artificial Intelligence* 623–632.
- NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. and SAIN, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *J. Comput. Graph. Statist.* **24** 579–599. MR3357396 <https://doi.org/10.1080/10618600.2014.914946>
- NYCHKA, D., HAMMERLING, D., SAIN, S. and LENNSEN, N. (2016). LatticeKrig: Multiresolution kriging based on Markov random fields. R package version 6.4.
- QUIÑONERO-CANDELA, J. and RASMUSSEN, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **6** 1939–1959. MR2249877
- RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704
- RIBATET, M., COOLEY, D. and DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist. Sinica* **22** 813–845. MR2954363
- RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- SANG, H. and HUANG, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 111–132. MR2885842 <https://doi.org/10.1111/j.1467-9868.2011.01007.x>
- SANTIN, G. and SCHABACK, R. (2016). Approximation of eigenfunctions in kernel-based spaces. *Adv. Comput. Math.* **42** 973–993. MR3533523 <https://doi.org/10.1007/s10444-015-9449-5>
- SAVITSKY, T. D. and SRIVASTAVA, S. (2018). Scalable Bayes under informative sampling. *Scand. J. Stat.* **45** 534–556. MR3858945 <https://doi.org/10.1111/sjos.12312>
- SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *Int. J. Manag. Sci. Eng. Manag.* **11** 78–88.
- SHANG, Z., HAO, B. and CHENG, G. (2019). Nonparametric Bayesian aggregation for massive data. *J. Mach. Learn. Res.* **20** Paper No. 140, 81. MR4030154
- SIMPSON, D., LINDGREN, F. and RUE, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* **23** 65–74. MR2873784 <https://doi.org/10.1002/env.1137>
- SRIVASTAVA, S., LI, C. and DUNSON, D. B. (2018). Scalable Bayes via barycenter in Wasserstein space. *J. Mach. Learn. Res.* **19** Paper No. 8, 35. MR3862415
- SRIVASTAVA, S. and XU, Y. (2021). Distributed Bayesian inference in linear mixed-effects models. *J. Comput. Graph. Statist.* **30** 594–611. MR4313463 <https://doi.org/10.1080/10618600.2020.1869025>
- SRIVASTAVA, S., CEVHER, V., DINH, Q. and DUNSON, D. (2015). WASP: Scalable Bayes via barycenters of subset posteriors. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics* 912–920.
- STEIN, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* **8** 1–19. MR3326818 <https://doi.org/10.1016/j.spasta.2013.06.003>
- STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376 <https://doi.org/10.1046/j.1369-7412.2003.05512.x>
- SU, Y. (2020). A divide and conquer algorithm of Bayesian density estimation. Preprint. Available at [arXiv:2002.07094](https://arxiv.org/abs/2002.07094).
- SZABÓ, B. and VAN ZANTEN, H. (2019). An asymptotic analysis of distributed nonparametric methods. *J. Mach. Learn. Res.* **20** Paper No. 87, 30. MR3960941
- VAN TREES, H. L. (2001). *Detection, Estimation, and Modulation Theory*. Wiley, New York.
- VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. (IMS) Collect.* **3** 200–222. IMS, Beachwood, OH. MR2459226 <https://doi.org/10.1214/074921708000000156>
- VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028
- VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852
- VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. MR0964183
- WANG, X. and DUNSON, D. B. (2013). Parallelizing MCMC via Weierstrass sampler. Preprint. Available at [arXiv:1312.4605](https://arxiv.org/abs/1312.4605).
- WANG, C. and SRIVASTAVA, S. (2021). Divide-and-Conquer Bayesian inference in hidden Markov models. Preprint. Available at [arXiv:2105.14395](https://arxiv.org/abs/2105.14395).
- WANG, X., GUO, F., HELLER, K. A. and DUNSON, D. B. (2015). Parallelizing MCMC with random partition trees. In *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, eds.) **28**.
- WENDLAND, H. (2005). *Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724
- WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 <https://doi.org/10.1201/9781420072884-c8>

- WIKLE, C. K. and HOLAN, S. H. (2011). Polynomial nonlinear spatio-temporal integro-difference equation models. *J. Time Series Anal.* **32** 339–350. MR2841788 <https://doi.org/10.1111/j.1467-9892.2011.00729.x>
- XUE, J. and LIANG, F. (2019). Double-parallel Monte Carlo for Bayesian analysis of big data. *Stat. Comput.* **29** 23–32. MR3905537 <https://doi.org/10.1007/s11222-017-9791-1>
- YANG, Y., BHATTACHARYA, A. and PATI, D. (2017). Frequentist coverage and sup-norm convergence rate in Gaussian process regression. Preprint. Available at arXiv:1708.04753.
- YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 <https://doi.org/10.1214/16-AOS1472>
- ZHANG, T. (2005). Learning bounds for kernel regression using effective data dimensionality. *Neural Comput.* **17** 2077–2098. MR2175849 <https://doi.org/10.1162/0899766054323008>
- ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. MR3450540
- ZHANG, M. M. and WILLIAMSON, S. A. (2019). Embarrassingly parallel inference for Gaussian processes. *J. Mach. Learn. Res.* **20** Paper No. 169, 26. MR4048980
- ZHU, H., WILLIAMS, C. K. I., ROHWER, R. J. and MORCINIEC, M. (1998). Gaussian regression and optimal finite dimensional linear models. In *Neural Networks and Machine Learning* (C. M. Bishop, ed.).

The Secret Life of I. J. Good

Sandy Zabell

Abstract. I. J. (“Jack”) Good was a leading Bayesian statistician for more than half a century after World War II, playing an important role in the post-war Bayesian revival. But his graduate training had been in pure mathematics rather than statistics (one of his doctoral advisors at Cambridge had been the famous G. H. Hardy). What was responsible for this metamorphosis from pure mathematician to applied and theoretical statistician? As Good himself only revealed in 1976, during the war he had initially served as an assistant to Alan Turing at Bletchley Park, working on the cryptanalysis of the German Naval Enigma, and it was from Turing that he acquired his life-long Bayesian philosophy. Declassified and other documents now permit us to understand in some detail how this came about, and indeed how many of the ideas Good explored and papers he wrote in the initial decades after the war, in fact, gave in sanitized form, results that had their origins in his wartime work. Drawing on these sources, this paper discusses the daily and very real use of Bayesian methods Turing and Good employed, and how this was gradually revealed by Good over the course of his life (including his return to classified work in the 1950s).

Key words and phrases: I. J. Good, Alan Turing, Bayesian statistics, Bletchley Park, cryptanalysis, Enigma machine, Banburismus, deciban, weight of evidence, Tunny.

REFERENCES

- ADAMS, J. F. (1985). Maxwell Herman Alexander Newman, 7 February 1897–22 February 1984. *Biogr. Mem. Fellows R. Soc.* **31** 437–452.
- ALEXANDER, C. H. O’D. (1945). Cryptographic History of Work on German Naval ENIGMA. HW 25/1, UK National Archives, Kew, London.
- BANKS, D. L. (1996). A conversation with I. J. Good. *Statist. Sci.* **11** 1–19. MR1437124 <https://doi.org/10.1214/ss/1032209661>
- BARNARD, G. A. (1946). Sequential tests in industrial statistics. *Suppl. J. R. Stat. Soc.* **8** 1–21. MR0019287
- BEESELY, P. (1977). *Very Special Intelligence: The Story of the Admiralty’s Operational Intelligence Centre, 1939–1945*. Hamish Hamilton, London. Reprinted in 2000 by Greenhill Books, London, with a new Introduction by W. J. R. Gardner, and a new Afterword by Ralph Erskine, and an updated Bibliography by Gardner and Erskine.
- BRITTON, J. L., ed. (1992) *Collected Works of A. M. Turing. Pure Mathematics* North-Holland, Amsterdam.
- BUDIANSKY, S. (2000). *Battle of Wits: The Complete Story of Codebreaking in World War II*. The Free Press, New York.
- BUDIANSKY, S. (2006). Colossus, codebreaking, and the digital age. In *Colossus: The Secrets of Bletchley Park’s Codebreaking Computers* (B. J. Copeland, ed.) 52–63. Oxford Univ. Press, Oxford.
- COPELAND, B. J., ed. (2006). *Colossus: The Secrets of Bletchley Park’s Codebreaking Computers* Oxford Univ. Press, Oxford.
- DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11** 427–444. MR0003527 <https://doi.org/10.1214/aoms/1177731829>
- FLOOD, J. L. (2011). Arthur Thomas Hatto: 1910–2010. In *Biographical Memoirs of Fellows of the British Academy* (R. Johnson, ed.) **10** 172–198. Oxford Univ. Press, Oxford.
- FRIEDMAN, W. F. (1922). *The Index of Coincidence and Its Applications in Cryptography*. Riverbank Laboratories, Geneva, IL. Department of Ciphers, Publication 22. Second ed., 1935, War Department, Office of the Chief Signal Officer, Government Printing Office, Washington, DC. Declassified and approved for release by NSA on 01-06-2014.
- GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin & Company, Limited, London.
- GOOD, I. J. (1951). Random motion on a finite Abelian group. *Proc. Camb. Philos. Soc.* **47** 756–762. MR0044061 <https://doi.org/10.1017/s0305004100027201>
- GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264. MR0061330 <https://doi.org/10.1093/biomet/40.3-4.237>
- GOOD, I. J. (1956). Some terminology and notation in information theory. *Proc. Inst. Elec. Engrs. C.* **103** 200–204. MR0076232
- GOOD, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. Ser. B* **20** 361–372. MR0102888
- GOOD, I. J. (1960). Weight of evidence, corroboration, explanatory power, information and the utility of experiments. *J. Roy. Statist.*

- Soc. Ser. B* **22** 319–331. MR0116394
- GOOD, I. J. (1961). Weight of evidence, causality and false-alarm probabilities. In *Information Theory (Symposium, London, 1960)* 125–136. Butterworths, Washington, DC. MR0131907
- GOOD, I. J. (1962). Analogues of Poisson’s summation formula. *Amer. Math. Monthly* **69** 259–266. MR0184006 <https://doi.org/10.2307/2312938>
- GOOD, I. J. (1965a). A list of properties of Bayes-Turing factors. *NSA Tech. J.* **10** 1–6. Issue 2 for 1965. Approved for Release by NSA on Appeal on 03-9-2011, FOIA Case #58820.
- GOOD, I. J. (1965b). *The Estimation of Probabilities. An Essay on Modern Bayesian Methods*. MIT Press, Cambridge, MA. MR0185724
- GOOD, I. J. (1967). On the principle of total evidence. *British J. Philos. Sci.* **17** 319–321.
- GOOD, I. J. (1968a). Corroboration, explanation, evolving probability, simplicity and a sharpened razor. *British J. Philos. Sci.* **19** 123–143. MR0242205 <https://doi.org/10.1093/bjps/19.2.123>
- GOOD, I. J. (1969). Statistics of language. In *Encyclopedia of Linguistics, Information and Control* (A. R. Meetham, ed.) 567–581. Pergamon, London.
- GOOD, I. J. (1971). 46656 varieties of Bayesians. *Amer. Statist.* **25** 62–63. Letter in the December 1971 issue.
- GOOD, I. J. (1973). The joint probability generating function for run-lengths in regenerative binary Markov chains, with applications. *Ann. Statist.* **1** 933–939. MR0341612
- GOOD, I. J. (1975). Explicativity, corroboration, and the relative odds of hypotheses. *Synthese* **30** 39–73.
- GOOD, I. J. (1976). Early work on computers at Bletchley. *Natl. Phys. Lab. Rep. Com. Sci.* **82**. Updated and reprinted in *Annals of the History of Computing* **1** (1979), 38–48.
- GOOD, I. J. (1979). Studies in the history of probability and statistics. XXXVII. A. M. Turing’s statistical work in World War II. *Biometrika* **66** 393–396. MR0548210 <https://doi.org/10.1093/biomet/66.2.393>
- GOOD, I. J. (1980). The contributions of Jeffreys to Bayesian statistics. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (A. Zellner, ed.) 21–34. North-Holland, Amsterdam.
- GOOD, I. J. (1982). A report on a lecture by Tom Flowers on the design of Colossus. *Ann. Hist. Comput.* **4** 53–59.
- GOOD, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Univ. Minnesota Press, Minneapolis, MN. MR0723501
- GOOD, I. J. (1988). The interface between statistics and philosophy of science. *Statist. Sci.* **3** 386–412. MR0996411
- GOOD, I. J. (1992a). Introductory remarks for the article in *Biometrika* **66** (1979). In *The Collected Works of A. M. Turing. Pure Mathematics* (J. L. Britton, ed.) 211–223. North-Holland, Amsterdam.
- GOOD, I. J. (1992b). The Bayes/non-Bayes compromise: A brief review. *J. Amer. Statist. Assoc.* **87** 597–606. MR1185188
- GOOD, I. J. (1993). Enigma and Fish. In *Codebreakers: The Inside Story of Bletchley Park* (F. H. Hinsley and A. Stripp, eds.) 149–166. Oxford Univ. Press, Oxford. Corrected paperback edition, 1994.
- GOOD, I. J. (2000). Turing’s anticipation of empirical Bayes in connection with the cryptanalysis of the Naval Enigma. *J. Stat. Comput. Simul.* **66** 101–111.
- GOOD, I. J. (2006). From Hut 8 to the Newmanry. In *Colossus: The Secrets of Bletchley Park’s Codebreaking Computers* (B. J. Copeland, ed.) 204–222. Oxford Univ. Press, Oxford.
- GOOD, I. J. and TOULMIN, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63. MR0077039 <https://doi.org/10.1093/biomet/43.1-2.45>
- GOOD, I. J. and TOULMIN, G. H. (1968). Coding theorems and weight of evidence. *IMA J. Appl. Math.* **4** 94–105.
- GREENBERG, J. (2014). *Gordon Welchman: Bletchley Park’s Architect of Ultra Intelligence*. Frontline Books, London.
- HALL, H. S. and KNIGHT, S. R. (1891). *Higher Algebra: A Sequel to Elementary Algebra for Schools*, 4th ed. Macmillan, London.
- HATTO, A. T. (1965). *The Nibelungenlied*. Penguin Books Inc, New York.
- HILTON, P. J. and WYLIE, S. (1967). *Homology Theory: An Introduction to Algebraic Topology*. Cambridge Univ. Press, Cambridge.
- HINSLEY, F. H. (1979–1990). *British Intelligence in the Second World War*. Her Majesty’s Stationery’s Office, London. Four volumes, Volume 3 issued in two parts.
- HINSLEY, F. H. and STRIPP, A. (1993). *Codebreakers: The Inside Story of Bletchley Park*. Oxford Univ. Press, Oxford. Corrected paperback reprinting, 1994.
- KAHN, D. (1967). *The Codebreakers: The Story of Secret Writing*. Macmillan, New York. Second edition, 1996, Macmillan, New York.
- KAHN, D. (1991). *Seizing the Enigma: The Race to Break the German U-Boat Codes 1939–1943*. Houghton Mifflin, Boston, MA. Second edition, 1998, Barnes and Noble, New York; 2012 edition, Naval Institute Press, Annapolis.
- KAHN, D. (2010). How I discovered World War II’s greatest spy. *Cryptologia* **34** 12–21. Reprinted in *How I Discovered World War II’s Greatest Spy, and Other Stories of Intelligence and Code*, CRC Press, Boca Raton, FL, 2014, Chapter 1.
- MAHON, A. P. (1945). The History of Hut 8. HW 25/2, UK National Archives, Kew, London.
- PLUMB, J. H. (1950). *England in the Eighteenth Century*. Pelican Books, London.
- REEDS, J. A. et al. (2015). *Breaking Teleprinter Ciphers at Bletchley Park*. Wiley–IEEE Press, Piscataway, NJ.
- REJEWSKI, M. (1981). How Polish mathematicians deciphered the Enigma. *Ann. Hist. Comput.* **3** 213–234. Includes discussion at the end by Cipher A. Deavours and I. J. Good.
- SEBAG-MONTEFIORE, H. (2000). *Enigma: The Battle for the Code*. Wiley, Hoboken, NJ.
- SINGH, S. (1999). *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*. Doubleday, New York. Very readable, but to be consulted with caution; see the review by Jim Reeds in the *Notices of the American Mathematical Society* **47** (2000) 369–372.
- TURING, S. (1959). *Alan M. Turing*. Heffer, Cambridge. MR0106802
- TURING, A. M. (2012a). The applications of probability to cryptography. Unpublished paper, c. 1941, UK National Archives, HW 25/37, released 2012 on the centenary of Turing’s birth.
- TURING, A. M. (2012b). Paper on the statistics of repetitions. Unpublished paper, c. 1941, UK National Archives, HW 25/38, released 2012 on the centenary of Turing’s birth.
- TURING, D. (2021). *X, Y & Z*, 2nd ed. The History Press, Cheltenham, UK. First edition, 2018. The author, Sir John Dermot Turing, is a nephew of Alan Turing, and has—in addition to a very successful career in his own right—recently written several books on the history of cryptography in WWII.
- WALD, A. (1945a). Sequential method of sampling for deciding between two courses of action. *J. Amer. Statist. Assoc.* **40** 277–306. MR0013276
- WALD, A. (1945b). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 <https://doi.org/10.1214/aoms/1177731118>
- WALD, A. (1947). *Sequential Analysis*. Wiley, New York. MR0020764
- WELCHMAN, G. (1982). *The Hut Six Story: Breaking the Enigma Codes*. McGraw-Hill, New York.

WINTERBOTHAM, F. W. (1974). *The Ultra Secret*. Weidenfeld and Nicolson, London.

ZABELL, S. (2012). Commentary on Alan M. Turing: The applications of probability to cryptography. *Cryptologia* **36** 191–214.

ZABELL, S. (2015). Statistics at Bletchley Park. In *Breaking Teleprinter Ciphers at Bletchley Park* (J. Reeds et al., eds.) lxxv–ci. Wiley–IEEE Press, Piscataway, NJ.

Cross-Study Replicability in Cluster Analysis

Lorenzo Masoero, Emma Thomas, Giovanni Parmigiani, Svitlana Tyekucheva and Lorenzo Trippa

Abstract. In cancer research, clustering techniques are widely used for exploratory analyses, playing a critical role in the identification of novel cancer subtypes and patient management. As data collected by multiple research groups grows, it is increasingly feasible to investigate the replicability of clustering procedures, that is, their ability to consistently recover biologically meaningful clusters across several data sets. In this paper, we review methods for replicability of clustering analyses, and discuss a novel framework for evaluating cross-study clustering replicability, useful when two or more studies are available. Our approach can be applied to any clustering algorithm and can employ different measures of similarity between partitions to quantify replicability, *globally* (i.e., for the whole sample) as well as *locally* (i.e., for individual clusters). Using experiments on synthetic and real gene expression data, we illustrate the usefulness of our procedure to evaluate if the same clusters are identified consistently across a collection of data sets.

Key words and phrases: Clustering, replicability, multiple studies.

REFERENCES

- ALBATINEH, A. N., NIEWIADOMSKA-BUGAJ, M. and MIHALKO, D. (2006). On similarity indices and correction for chance agreement. *J. Classification* **23** 301–313. MR2295924 <https://doi.org/10.1007/s00357-006-0017-z>
- ALEXE, G., DALGIN, G. S., RAMASWAMY, R., DELISI, C. and BHANOT, G. (2006). Data perturbation independent diagnosis and validation of breast cancer subtypes using clustering and patterns. *Cancer Inform.* **2**.
- ARRIETA, A. B., DÍAZ-RODRÍGUEZ, N., DEL SER, J., BENNETOT, A., TABIK, S., BARBADO, A., GARCÍA, S., GIL-LÓPEZ, S., MOLINA, D. et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf. Fusion* **58** 82–115.
- BEN-DAVID, S., PÁL, D. and SIMON, H. U. (2007). Stability of k -means clustering. In *Learning Theory. Lecture Notes in Computer Science* **4539** 20–34. Springer, Berlin. MR2397576 https://doi.org/10.1007/978-3-540-72927-3_4
- BERNAU, C., RIESTER, M., BOULESTEIX, A.-L., PARMIGIANI, G., HUTTENHOWER, C., WALDRON, L. and TRIPPA, L. (2014). Cross-study validation for the assessment of prediction algorithms. *Bioinformatics* **30**.
- BERTONI, A. and VALENTINI, G. (2007). Model order selection for bio-molecular data clustering. *BMC Bioinform.* **8**.
- BROCK, G., PIHUR, V., DATTA, S. and DATTA, S. (2008). clvalid: An R package for cluster validation. *J. Stat. Softw.* **25**.
- BRYAN, J. (2004). Problems in gene clustering based on gene expression data. *J. Multivariate Anal.* **90** 44–66. MR2086340 <https://doi.org/10.1016/j.jmva.2004.02.011>
- ESTER, M., KRIEGEL, H.-P., SANDER, J. and XU, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*.
- FANG, Y. and WANG, J. (2012). Selection of the number of clusters via the bootstrap method. *Comput. Statist. Data Anal.* **56** 468–477. MR2853747 <https://doi.org/10.1016/j.csda.2011.09.003>
- FRÄNTI, P. and SIERANOJA, S. (2018). K-means properties on six clustering benchmark datasets. *Appl. Intell.* **48**.
- FRÄNTI, P., VIRMAJOKI, O. and HAUTAMÄKI, V. (2006). Fast agglomerative clustering using a k -nearest neighbor graph. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**.
- FREY, B. J. and DUECK, D. (2007). Clustering by passing messages between data points. *Science* **315** 972–976. MR2292174 <https://doi.org/10.1126/science.1136800>
- HAIBE-KAINS, B., DESMEDT, C., LOI, S., CULHANE, A. C., BONTEMPI, G., QUACKENBUSH, J. and SOTIRIOU, C. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. *J. Natl. Cancer Inst.* **104**.
- HAYES, D. N., MONTI, S., PARMIGIANI, G., GILKS, C. B., NAOIKI, K., BHATTACHARJEE, A., SOCINSKI, M. A., PEROU, C. and MEYERSON, M. (2006). Gene expression profiling reveals re-

Lorenzo Masoero is Applied Scientist at Amazon. This research work was conducted prior to this position (e-mail: masoerl@amazon.com). Emma Thomas is Data Scientist at RAND. Giovanni Parmigiani is Professor in the Departments of Data Science, Dana-Farber Cancer Institute and Biostatistics, Harvard T.H. Chan School of Public Health. Svitlana Tyekucheva is Senior Research Scientist, Department of Data Science, Dana-Farber Cancer Institute and Department of Biostatistics, Harvard T.H. Chan School of Public Health. Lorenzo Trippa is Associate Professor, Department of Data Science, Dana-Farber Cancer Institute and Department of Biostatistics, Harvard T.H. Chan School of Public Health.

- producible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **24**.
- HENNIG, C. (2007). Cluster-wise assessment of cluster stability. *Comput. Statist. Data Anal.* **52** 258–271. [MR2409980 https://doi.org/10.1016/j.csda.2006.11.025](https://doi.org/10.1016/j.csda.2006.11.025)
- HENNIG, C. (2015). Package ‘fpc’. R-project, 91.
- HUBERT, L. and ARABIE, P. (1985). Comparing partitions. *J. Classification* **2**.
- JASKOWIAK, P. A., CAMPELLO, R. J. and COSTA, I. G. (2014). On the selection of appropriate distances for gene expression data clustering. *BMC Bioinform.* **15** S2.
- KAPP, A. V. and TIBSHIRANI, R. (2006). Are clusters found in one dataset present in another dataset? *Biostatistics* **8**.
- LANCASTER, H. O. and SENETA, E. (1969). Chi-square distribution. In *Encyclopedia of Biostatistics* **2**.
- LANGE, T., ROTH, V., BRAUN, M. L. and BUHMANN, J. M. (2004). Stability-based validation of clustering solutions. *Neural Comput.* **16**.
- LEVENSTIEN, M. A., YANG, Y. and OTT, J. (2003). Statistical significance for hierarchical clustering in genetic association and microarray expression studies. *BMC Bioinform.* **4**.
- LEVINE, E. and DOMANY, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Comput.* **13**.
- LIM, C. and YU, B. (2016). Estimation stability with cross-validation (ESCV). *J. Comput. Graph. Statist.* **25** 464–492. [MR3499690 https://doi.org/10.1080/10618600.2015.1020159](https://doi.org/10.1080/10618600.2015.1020159)
- LIU, Y., HAYES, D. N., NOBEL, A. and MARRON, J. S. (2008). Statistical significance of clustering for high-dimension, low-sample size data. *J. Amer. Statist. Assoc.* **103** 1281–1293. [MR2528840 https://doi.org/10.1198/016214508000000454](https://doi.org/10.1198/016214508000000454)
- LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. [MR0651807 https://doi.org/10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489)
- MAATEN, L. v. D. and HINTON, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**.
- MASOERO, L., THOMAS, E., PARMIGIANI, G., TYEKUCHEVA, S. and TRIPPA, L. (2023). Supplement to “Cross-study replicability in cluster analysis.” <https://doi.org/10.1214/22-STS871SUPP>
- MCSHANE, L. M., RADMACHER, M. D., FREIDLIN, B., YU, R., LI, M.-C. and SIMON, R. (2002). Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics* **18**.
- MÜLLER, P. and QUINTANA, F. (2010). Random partition models with regression on covariates. *J. Statist. Plann. Inference* **140** 2801–2808. [MR2651966 https://doi.org/10.1016/j.jspi.2010.03.002](https://doi.org/10.1016/j.jspi.2010.03.002)
- MURDOCH, W. J., SINGH, C., KUMBIER, K., ABBASI-ASL, R. and YU, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116** 22071–22080. [MR4030584 https://doi.org/10.1073/pnas.1900654116](https://doi.org/10.1073/pnas.1900654116)
- NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE (2019). *Reproducibility and Replicability in Science*. The National Academies Press, Washington, DC.
- PARKER, J. S., MULLINS, M., CHEANG, M. C., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X. et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**.
- PEROU, C. M., SØRLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H. et al. (2000). Molecular portraits of human breast tumours. *Nature* **406**.
- RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66**.
- SCHROEDER, M., HAIBE-KAINS, B., CULHANE, A., SOTIRIOU, C., BONTEMPI, G. and QUACKENBUSH, J. (2011a). breastCancerMAINZ: Gene expression dataset published by Schmidt et al. (2008) (MAINZ). R package version 1.16.0.
- SCHROEDER, M., HAIBE-KAINS, B., CULHANE, A., SOTIRIOU, C., BONTEMPI, G. and QUACKENBUSH, J. (2011b). breastCancerTRANSBIG: Gene expression dataset published by Desmedt et al. (2007) (TRANSBIG). R package version 1.16.0.
- SCHROEDER, M., HAIBE-KAINS, B., CULHANE, A., SOTIRIOU, C., BONTEMPI, G. and QUACKENBUSH, J. (2011c). breastCancerVDX: Gene expression datasets published by Wang et al. (2005) and Minn et al. (2007) (VDX). R package version 1.16.0.
- SMOLKIN, M. and GHOSH, D. (2003). Cluster stability scores for microarray data in cancer studies. *BMC Bioinform.* **4**.
- TIBSHIRANI, R. and WALTHER, G. (2005). Cluster validation by prediction strength. *J. Comput. Graph. Statist.* **14** 511–528. [MR2170199 https://doi.org/10.1198/106186005X59243](https://doi.org/10.1198/106186005X59243)
- TRIPPA, L., WALDRON, L., HUTTENHOWER, C. and PARMIGIANI, G. (2015). Bayesian nonparametric cross-study validation of prediction methods. *Ann. Appl. Stat.* **9** 402–428. [MR3341121 https://doi.org/10.1214/14-AOAS798](https://doi.org/10.1214/14-AOAS798)
- ULLMANN, U., HENNIG, C. and BOULESTEIX, A. (2022). Validation of cluster analysis results on validation data: A systematic framework. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **3** e1444. <https://doi.org/10.1002/widm.1444>
- VINH, N. X., EPPS, J. and BAILEY, J. (2009). Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, New York.
- VINH, N. X., EPPS, J. and BAILEY, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11** 2837–2854. [MR2738784](https://doi.org/10.1214/10-AOAS798)
- VON LUXBURG, U. (2010). Clustering stability: An overview. *Found. Trends Mach. Learn.* **2**.
- WADE, S. and GHAHRAMANI, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Anal.* **13** 559–626. With discussion and a reply by the authors. [MR3807860 https://doi.org/10.1214/17-BA1073](https://doi.org/10.1214/17-BA1073)
- WAKS, A. G. and WINER, E. P. (2019). Breast cancer treatment: A review. *JAMA* **321**.
- WARD, J. H. JR. (1963). Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* **58** 236–244. [MR0148188](https://doi.org/10.1080/01621459.1963.10501622)
- YU, B. (2013). Stability. *Bernoulli* **19** 1484–1500. [MR3102560 https://doi.org/10.3150/13-BEJSP14](https://doi.org/10.3150/13-BEJSP14)
- ZHANG, T., RAMAKRISHNAN, R. and LIVNY, M. (1996). Birch: An efficient data clustering method for very large databases. In *ACM Sigmod Record* ACM, New York.

Principal Fairness for Human and Algorithmic Decision-Making

Kosuke Imai and Zhichao Jiang

Abstract. Using the concept of principal stratification from the causal inference literature, we introduce a new notion of fairness, called principal fairness, for human and algorithmic decision-making. Principal fairness states that one should not discriminate among individuals who would be similarly affected by the decision. Unlike the existing statistical definitions of fairness, principal fairness explicitly accounts for the fact that individuals can be impacted by the decision. This causal fairness formulation also enables on-line or post-hoc fairness evaluation and policy learning. We also explain how principal fairness relates to the existing causality-based fairness criteria. In contrast to the counterfactual fairness criteria, for example, principal fairness considers the effects of decision in question rather than those of protected attributes of interest. Finally, we discuss how to conduct empirical evaluation and policy learning under the proposed principal fairness criterion.

Key words and phrases: Algorithmic fairness, causal inference, potential outcomes, principal stratification.

REFERENCES

- [1] AGARWAL, A., BEYGELZIMER, A., DUDÍK, M., LANGFORD, J. and WALLACH, H. (2018). A reductions approach to fair classification. In *International Conference on Machine Learning* 60–69. PMLR.
- [2] BAROCAS, S., HARDT, M. and NARAYANAN, A. (2019). *Fairness and Machine Learning*. fairmlbook.org. Available at <http://www.fairmlbook.org>.
- [3] BAROCAS, S. and SELBST, A. D. (2016). Big data’s disparate impact. *California Law Review* **104** 671–732.
- [4] BEN-MICHAEL, E., IMAI, K. and JIANG, Z. (2022). Policy learning with asymmetric utilities. Technical Report. ArXiv Preprint. Available at <https://arxiv.org/pdf/2206.10479.pdf>.
- [5] BERK, R., HEIDARI, H., JABBARI, S., KEARNS, M. and ROTH, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociol. Methods Res.* **50** 3–44. MR4198551 <https://doi.org/10.1177/0049124118782533>
- [6] BEUTEL, A., CHEN, J., DOSHI, T., QIAN, H., WOODRUFF, A., LUU, C., KREITMANN, P., BISCHOF, J. and CHI, E. H. (2019). Putting fairness principles into practice: Challenges, metrics, and improvements. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. AIES’19 453–459. Association for Computing Machinery, New York, NY, USA.
- [7] CELIS, L. E., HUANG, L., KESWANI, V. and VISHNOI, N. K. (2019). Classification with fairness constraints: A meta-algorithm with provable guarantees. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* 319–328.
- [8] CHIAPPA, S. (2019). Path-specific counterfactual fairness. In *Proceedings of the AAAI Conference on Artificial Intelligence* **33** 7801–7808.
- [9] CHOULDECHOVA, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* **5** 153–163. <https://doi.org/10.1089/big.2016.0047>
- [10] CHOULDECHOVA, A. and ROTH, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Commun. ACM* **63** 82–89.
- [11] CORBETT-DAVIES, S. and GOEL, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. Technical Report. Available at arXiv:1808.00023.
- [12] COSTON, A., MISHLER, A., KENNEDY, E. H. and CHOULDECHOVA, A. (2020). Counterfactual risk assessments, evaluation, and fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* 582–593.
- [13] D’AMOUR, A., SRINIVASAN, H., ATWOOD, J., BALJEKAR, P., SCULLEY, D. and HALPERN, Y. (2020). Fairness is not static: Deeper understanding of long term fairness via simulation studies. In *FAT*’20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* 525–534.
- [14] DWORK, C., HARDT, M., PITASSI, T., REINGOLD, O. and ZEMEL, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* 214–226. ACM, New York. MR3388391
- [15] FISHER, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, London.

Kosuke Imai is Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge, Massachusetts 02138, USA (e-mail: imai@harvard.edu). Zhichao Jiang (corresponding author) is Professor, School of Mathematics, Sun Yat-sen University, Guangzhou, Guangdong 510275, China (e-mail: jiangzhch7@mail.sysu.edu.cn).

- [16] FRANGAKIS, C. E. and RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58** 21–29. MR1891039 <https://doi.org/10.1111/j.0006-341X.2002.00021.x>
- [17] GREINER, D. J. and RUBIN, D. B. (2011). Causal effects of perceived immutable characteristics. *Rev. Econ. Stat.* **93** 775–785.
- [18] HARDT, M., PRICE, E. and SREBRO, N. (2016). Equality of opportunity in supervised learning. Technical Report. Available at [arXiv:1610.02413](https://arxiv.org/abs/1610.02413).
- [19] HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. MR0867618
- [20] IMAI, K., JIANG, Z., GREINER, D. J., HALEN, R. and SHIN, S. (2022). Experimental evaluation of computer-assisted human decision-making: Application to pretrial risk assessment instrument (with discussions). *J. Roy. Statist. Soc. Ser. A* To appear.
- [21] JACKSON, J. W. and VANDERWEELE, T. J. (2018). Decomposition analysis to identify intervention targets for reducing disparities. *Epidemiology* **29** 825–835. <https://doi.org/10.1097/EDE.0000000000000901>
- [22] JOHNDROW, J. E. and LUM, K. (2019). An algorithm for removing sensitive information: Application to race-independent recidivism prediction. *Ann. Appl. Stat.* **13** 189–220. MR3937426 <https://doi.org/10.1214/18-AOAS1201>
- [23] KALLUS, N. and ZHOU, A. (2019). Assessing disparate impact of personalized interventions: Identifiability and bounds. In *33rd Conference on Neural Information Processing Systems*.
- [24] KAMISHIMA, T., AKAHO, S. and SAKUMA, J. (2011). Fairness-aware learning through regularization approach. In *2011 IEEE 11th International Conference on Data Mining Workshops* 643–650.
- [25] KILBERTUS, N., CARULLA, M. R., PARASCANDOLO, G., HARDT, M., JANZING, D. and SCHÖLKOPF, B. (2017). Avoiding discrimination through causal reasoning. In *Advances in Neural Information Processing Systems* 656–666.
- [26] KLEINBERG, J., MULLAINATHAN, S. and RAGHAVAN, M. (2017). Inherent trade-offs in the fair determination of risk scores. In *8th Innovations in Theoretical Computer Science Conference* C. H. Papadimitrou, ed.). *LIPIcs. Leibniz Int. Proc. Inform.* **67** Art. No. 43. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3754967
- [27] KNOX, D., LOWE, W. and MUMMOLO, J. (2022). Administrative records mask racially biased policing. *Am. Polit. Sci. Rev.* **114** 619–637.
- [28] KUSNER, M., LOFTUS, J., RUSSELL, C. and SILVA, R. (2017). Counterfactual fairness. In *Proceedings of the 31st Conference on Neural Information Processing Systems*.
- [29] MITCHELL, S., POTASH, E., BAROCCAS, S., D’AMOUR, A. and LUM, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annu. Rev. Stat. Appl.* **8** 141–163. MR4243544 <https://doi.org/10.1146/annurev-statistics-042720-125902>
- [30] NABI, R. and SHPITSER, I. (2018). Fair inference on outcomes. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [31] NEYMAN, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, Section 9. (Translated in 1990). *Statist. Sci.* **5** 465–480. MR1092986
- [32] PLECKO, D. and BAREINBOIM, E. (2022). Causal Fairness Analysis. Technical Report. ArXiv Preprint. Available at <https://arxiv.org/abs/2207.11385>.
- [33] PLEČKO, D. and MEINSHAUSEN, N. (2020). Fair data adaptation with quantile preservation. *J. Mach. Learn. Res.* **21** Paper No. 242. MR4209528
- [34] RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *J. Educ. Psychol.* **66** 688–701.
- [35] VANDERWEELE, T. J. and SHPITSER, I. (2011). A new criterion for confounder selection. *Biometrics* **67** 1406–1413. MR2872391 <https://doi.org/10.1111/j.1541-0420.2011.01619.x>
- [36] ZAFAR, M. B., VALERA, I., GOMEZ RODRIGUEZ, M. and GUMMADI, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web. WWW’17* 1171–1180. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE.
- [37] ZHANG, J. and BAREINBOIM, E. (2018). Fairness in decision-making—the causal explanation formula. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI’18/IAAI’18/EAAI’18*. AAAI Press, Menlo Park, CA.

Confidence and Discoveries with E -values

Vladimir Vovk and Ruodu Wang

Abstract. We discuss systematically two versions of confidence regions: those based on p -values and those based on e -values, a recent alternative to p -values. Both versions can be applied to multiple hypothesis testing, and in this paper we are interested in procedures that control the number of false discoveries under arbitrary dependence between the base p - or e -values. We introduce a procedure that is based on e -values and show that it is efficient both computationally and statistically using simulated and real-world data sets. Comparison with the corresponding standard procedure based on p -values is not straightforward, but there are indications that the new one performs significantly better in some situations.

Key words and phrases: Hypothesis testing, multiple hypothesis testing, Bayes factor, discovery vector, discovery matrix.

REFERENCES

- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245](#) <https://doi.org/10.1214/aos/1013699998>
- BENJAMINI, Y., VEAUX, R. D. D., EFRON, B., EVANS, S., GLICKMAN, M., GRAUBARD, B. I., HE, X., MENG, X.-L., REID, N. et al. (2021). The ASA president’s task force statement on statistical significance and replicability. *Ann. Appl. Stat.* **15** 1084–1085.
- BERNARDO, J. M. and SMITH, A. F. M. (2000). *Bayesian Theory*. Wiley, Chichester.
- BERNOULLI, J. (1713). *Ars Conjectandi*. Thurnisius, Basel.
- CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, 2nd ed. Duxbury, Pacific Grove, CA.
- COURNOT, A.-A. (1843). *Exposition de la Théorie des Chances et des Probabilités*. Hachette, Paris.
- COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. [MR0370837](#)
- DE FINETTI, B. (2017). *Theory of Probability*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. [MR3643261](#) <https://doi.org/10.1002/9781119286387>
- DUBOIS, D. and PRADE, H. (1988). *Possibility Theory*. Plenum Press, New York.
- FISHER, R. A. (1973). *Statistical Methods and Scientific Inference*, 3rd ed. Hafner, New York.
- GÁCS, P. (2005). Uniform test of algorithmic randomness over a general space. *Theoret. Comput. Sci.* **341** 91–137. [MR2159646](#) <https://doi.org/10.1016/j.tcs.2005.03.054>
- GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p -value weighting. *Biometrika* **93** 509–524. [MR2261439](#) <https://doi.org/10.1093/biomet/93.3.509>
- GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. [MR2065197](#) <https://doi.org/10.1214/009053604000000283>
- GOEMAN, J. J., HEMERIK, J. and SOLARI, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *Ann. Statist.* **49** 1218–1238. [MR4255125](#) <https://doi.org/10.1214/20-aos1999>
- GOEMAN, J. J., MEIJER, R. and KREBS, T. (2019). *homme1*: Methods for closed testing with Simes inequality, in particular Hommel’s method. R package version 1.5, available on CRAN.
- GOEMAN, J. J., ROSENBLATT, J. D. and NICHOLS, T. E. (2019). The harmonic mean p -value: Strong versus weak control, and the assumption of independence. *Proc. Natl. Acad. Sci. USA* **116** 23382–23383.
- GOEMAN, J. J. and SOLARI, A. (2011a). Multiple testing for exploratory research. *Statist. Sci.* **26** 584–597. Correction: **28** 464. [MR2951390](#) <https://doi.org/10.1214/11-STS356>
- GOEMAN, J. J. and SOLARI, A. (2011b). Multiple testing for exploratory research: Rejoinder. *Statist. Sci.* **26** 608–612.
- GOEMAN, J. J., MEIJER, R. J., KREBS, T. J. P. and SOLARI, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika* **106** 841–856. [MR4046036](#) <https://doi.org/10.1093/biomet/asz041>
- GÖNEN, M., JOHNSON, W. O., LU, Y. and WESTFALL, P. H. (2005). The Bayesian two-sample t test. *Amer. Statist.* **59** 252–257. [MR2149766](#) <https://doi.org/10.1198/000313005X55233>
- GÖNEN, M., JOHNSON, W. O., LU, Y. and WESTFALL, P. H. (2019). Comparing objective and subjective Bayes factors for the two-sample comparison: The classification theorem in action. *Amer. Statist.* **73** 22–31. [MR3923052](#) <https://doi.org/10.1080/00031305.2017.1322142>
- GOOD, I. J. (1958). Significance tests in parallel and in series. *J. Amer. Statist. Assoc.* **53** 799–813. [MR0103560](#)
- GRÜNWARD, P., DE HEIDE, R. and KOOLEN, W. M. (2020). Safe testing Technical Report. Available at [arXiv:1906.07801](https://arxiv.org/abs/1906.07801) [math.ST].
- GRÜNWARD, P. and VAN OMMEN, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a pro-

Vladimir Vovk is Professor, Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, UK (e-mail: v.vovk@rhul.ac.uk). Ruodu Wang is Professor and University Research Chair, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada (e-mail: wang@uwaterloo.ca).

- posal for repairing it. *Bayesian Anal.* **12** 1069–1103. MR3724979 <https://doi.org/10.1214/17-BA1085>
- GUINDANI, M., MÜLLER, P. and ZHANG, S. (2009). A Bayesian discovery procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 905–925. MR2750250 <https://doi.org/10.1111/j.1467-9868.2009.00714.x>
- HEDENFALK, I., DUGGAN, D., CHEN, Y., RADMACHER, M., BITTNER, M., SIMON, R., MELTZER, P., GUSTERSON, B., ESTELLER, M. et al. (2001). Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344** 539–548.
- HELD, L. (2019). On the Bayesian interpretation of the harmonic mean p -value. *Proc. Natl. Acad. Sci. USA* **116** 5855–5856. MR3939777 <https://doi.org/10.1073/pnas.1900671116>
- HEMERIK, J. and GOEMAN, J. (2018). Exact testing with random permutations. *TEST* **27** 811–825. MR3878362 <https://doi.org/10.1007/s11749-017-0571-1>
- HEMERIK, J., SOLARI, A. and GOEMAN, J. J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika* **106** 635–649. MR3992394 <https://doi.org/10.1093/biomet/asz021>
- HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597
- HOMMEL, G. (1986). Multiple test procedures for arbitrary dependence structures. *Metrika* **33** 321–336. MR0868042 <https://doi.org/10.1007/BF01894765>
- JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford. MR0187257
- KLEENE, S. C. (1967). *Mathematical Logic*. Wiley, New York. MR0216930
- LEHMANN, E. L. (2011). *Fisher, Neyman, and the Creation of Classical Statistics*. Springer, New York. MR2798202 <https://doi.org/10.1007/978-1-4419-9500-1>
- LEHMANN, E. L. and ROMANO, J. P. (2022). *Testing Statistical Hypotheses*, 4th ed. Springer, Cham.
- LEVIN, L. A. (1976). Uniform tests of randomness. *Sov. Math., Dokl.* **17** 337–340.
- LY, A., VERHAGEN, J. and WAGENMAKERS, E.-J. (2016). Harold Jeffreys’s default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *J. Math. Psych.* **72** 19–32. MR3506022 <https://doi.org/10.1016/j.jmp.2015.06.004>
- NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection (with discussion). *J. R. Stat. Soc.* **97** 558–625.
- NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond. A* **236** 333–380.
- NEYMAN, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32** 128–150. MR0005582 <https://doi.org/10.1093/biomet/32.2.128>
- ROMANO, J. P. and WOLF, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. MR2351090 <https://doi.org/10.1214/009053606000001622>
- ROUDER, J. N., SPECKMAN, P. L., SUN, D. and MOREY, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16** 225–237.
- SARKAR, S. K. (2011). Simes’ test in multiple testing. In *International Encyclopedia of Statistical Science* (M. Lovric, ed.) 1325–1327. Springer, Berlin.
- SCHERVISH, M. J. (1995). *Theory of Statistics. Springer Series in Statistics*. Springer, New York. MR1354146 <https://doi.org/10.1007/978-1-4612-4250-5>
- SELLKE, T., BAYARRI, M. J. and BERGER, J. O. (2001). Calibration of p values for testing precise null hypotheses. *Amer. Statist.* **55** 62–71. MR1818723 <https://doi.org/10.1198/000313001300339950>
- SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ. MR0464340
- SHAFER, G. (2007). From Cournot’s principle to market efficiency. In *Augustin Cournot: Modelling Economics* (J.-P. Touffut, ed.) 55–95. Edward Elgar, Cheltenham.
- SHAFER, G. (2021). The language of betting as a strategy for statistical and scientific communication (with discussion). *J. R. Stat. Soc., A* **184** 407–478.
- SHAFER, G. (2022). Bayesian, fiducial, frequentist. In *Handbook on Bayesian, Fiducial and Frequentist (BFF) Inferences* (J. Berger, X.-L. Meng, N. Reid and M. Xie, eds.) CRC Press, Boca Raton. (to appear).
- SHAFER, G. and VOVK, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley, Hoboken, NJ.
- SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 <https://doi.org/10.1093/biomet/73.3.751>
- STOREY, J. D., DAI, J. Y. and LEEK, J. T. (2007). The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics* **8** 414–432.
- STOREY, J. D. and TIBSHIRANI, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100** 9440–9445. MR1994856 <https://doi.org/10.1073/pnas.1530509100>
- STOREY, J. D., BASS, A. J., DABNEY, A. and ROBINSON, D. (2019). q value: Q-value estimation for false discovery rate control. R package version 2.18.0, available on Bioconductor.
- STUART, A., ORD, J. K. and ARNOLD, S. (1999). *Kendall’s Advanced Theory of Statistics. Vol. 2A: Classical Inference and the Linear Model*, 6th ed. Arnold, London. MR1687411
- TIAN, J., CHEN, X., KATSEVICH, E., GOEMAN, J. and RAMDAS, A. (2021). Large-scale simultaneous inference under dependence. Technical Report. Scandinavian Journal of Statistics. Available at [arXiv:2102.11253](https://arxiv.org/abs/2102.11253).
- VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220
- VOVK, V. G. and V’YUGIN, V. V. (1993). On the empirical validity of the Bayesian method. *J. Roy. Statist. Soc. Ser. B* **55** 253–266. MR1210435
- VOVK, V. and WANG, R. (2020a). Combining p -values via averaging. *Biometrika* **107** 791–808. MR4186488 <https://doi.org/10.1093/biomet/asaa027>
- VOVK, V. and WANG, R. (2020b). True and false discoveries with independent e -values. Technical Report. Available at [arXiv:2003.00593](https://arxiv.org/abs/2003.00593) [stat.ME].
- VOVK, V. and WANG, R. (2021). E -values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 <https://doi.org/10.1214/20-aos2020>
- VOVK, V., WANG, B. and WANG, R. (2022). Admissible ways of merging p -values under arbitrary dependence. *Ann. Statist.* **50** 351–375. MR4382020 <https://doi.org/10.1214/21-aos2109>
- WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.
- WANG, M. and LIU, G. (2016). A simple two-sample Bayesian t -test for hypothesis testing. *Amer. Statist.* **70** 195–201. MR3511050 <https://doi.org/10.1080/00031305.2015.1093027>
- WANG, R. and RAMDAS, A. (2022). False discovery rate control with e -values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 822–852. MR4460577
- WIGGINS, G. A. R., WALKER, L. C. and PEARSON, J. F. (2020). Genome-wide gene expression analyses of BRCA1- and BRCA2-associated breast and ovarian tumours. *Cancers* **12** 3015.
- WILSON, D. J. (2019). The harmonic mean p -value for combining dependent tests. *Proc. Natl. Acad. Sci. USA* **116** 1195–1200. MR3904688 <https://doi.org/10.1073/pnas.1814092116>

Conversations with Gábor J. Székely

Yulia R. Gel, Edsel A. Peña and Huixia Judy Wang

Abstract. Gábor J. Székely was born in Budapest, Hungary on February 4, 1947. He graduated from Eötvös Loránd University (ELTE) with an M.S. degree in 1970, and a Ph.D. degree in 1971. He received his Candidate Degree from the Hungarian Academy of Sciences in 1976, and the Doctor of Science Degree (D. Sc.) from the Hungarian Academy of Sciences in 1986. Székely joined the Department of Probability Theory of ELTE in 1970. In 1989, he became the founding chair of the Department of Stochastics of the Budapest Institute of Technology (Technical University of Budapest). In 1995, he moved to the United States as a tenured full professor at Bowling Green State University (BGSU) in Bowling Green, Ohio. Before that, in 1990–1991, he was the first Lukacs Distinguished Professor at BGSU. Székely had several visiting positions, for example, at the University of Amsterdam in 1977 and at Yale University in 1989. Between 2006 and 2022, he served as a Program Director in the Statistics Program of the Division of Mathematical Sciences at the U.S. National Science Foundation. Székely has about 250 publications, including 6 books in several languages. In 1988, he received the Rollo Davidson Prize from Cambridge University, jointly with Imre Z. Ruzsa for their work on algebraic probability theory. In 2010, Székely became an Elected Fellow of the Institute of Mathematical Statistics mostly for his works dealing with physics concepts in statistics like energy statistics and distance correlation. He had the fortune to know and work with world-class mathematicians and statisticians like (in chronological order of their first meetings): P. Erdős, A. Rényi, Y. Linnik, B. de Finetti, A. N. Kolmogorov, H. Robbins, G. Pólya, L. Shepp, G. Wahba, C. R. Rao, B. Efron, P. Bickel and E. Seneta.

Key words and phrases: Distance correlation, energy of data, Lukacs Professorship, negative probability, uncertainty principle of games.

REFERENCES

- [1] BENNETT, C. D., GLASS, A. M. W. and SZÉKELY, G. J. (2004). Fermat's last theorem for rational exponents. *Amer. Math. Monthly* **111** 322–329. MR2057186 <https://doi.org/10.2307/4145241>
- [2] BUCZOLICH, Z. and SZÉKELY, G. J. (1989). When is a weighted average of ordered sample elements a maximum likelihood estimator of the location parameter? *Adv. in Appl. Math.* **10** 439–456. MR1023943 [https://doi.org/10.1016/0196-8858\(89\)90024-9](https://doi.org/10.1016/0196-8858(89)90024-9)
- [3] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics. Probability and Mathematical Statistics*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York. MR0666546
- [4] DYSON, F. (2009). Birds and frogs. *Notices Amer. Math. Soc.* **56** 212–223. MR2483565
- [5] EDELMANN, D., MÓRI, T. F. and SZÉKELY, G. J. (2021). On relationships between the Pearson and the distance correlation coefficients. *Statist. Probab. Lett.* **169** Paper No. 108960, 6. MR4163211 <https://doi.org/10.1016/j.spl.2020.108960>
- [6] FÜREDI, Z., SZÉKELY, G. J. and ZUBOR, Z. (1996). On the lottery problem. *J. Combin. Des.* **4** 5–10. MR1364095 [https://doi.org/10.1002/\(SICI\)1520-6610\(1996\)4:1<5::AID-JCD2>3.3.CO;2-W](https://doi.org/10.1002/(SICI)1520-6610(1996)4:1<5::AID-JCD2>3.3.CO;2-W)
- [7] HALÁSZ, G. and SZÉKELY, G. J. (1976). On the elementary symmetric polynomials of independent random variables. *Acta Math. Acad. Sci. Hung.* **28** 397–400. MR0423491 <https://doi.org/10.1007/BF01896806>
- [8] KERNS, G. J. and SZÉKELY, G. J. (2006). De Finetti's theorem for abstract finite exchangeable sequences. *J. Theoret. Probab.* **19** 589–608. MR2280511 <https://doi.org/10.1007/s10959-006-0028-z>

Yulia R. Gel is Professor, Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75080, USA (e-mail: ygl@utdallas.edu). Edsel A. Peña is Professor, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, USA (e-mail: pena@stat.sc.edu). Huixia Judy Wang is Professor and Chair, Department of Statistics, George Washington University, Washington, DC 20052, USA (e-mail: judywang@gwu.edu).

- [9] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 <https://doi.org/10.1214/12-AOP803>
- [10] MAJOR, P. (1999). The limit behavior of elementary symmetric polynomials of i.i.d. random variables when their order tends to infinity. *Ann. Probab.* **27** 1980–2010. MR1742897 <https://doi.org/10.1214/aop/1022677557>
- [11] MÓRI, T. F. and SZÉKELY, G. J. (2019). Four simple axioms of dependence measures. *Metrika* **82** 1–16. MR3897521 <https://doi.org/10.1007/s00184-018-0670-3>
- [12] MÓRI, T. F., SZÉKELY, G. J. and RIZZO, M. L. (2021). On energy tests of normality. *J. Statist. Plann. Inference* **213** 1–15. MR4182996 <https://doi.org/10.1016/j.jspi.2020.11.001>
- [13] PEÑA, E. A., ROHATGI, V. K. and SZÉKELY, G. J. (1992). On the nonexistence of ancillary statistics. *Statist. Probab. Lett.* **15** 357–360. MR1193892 [https://doi.org/10.1016/0167-7152\(92\)90153-V](https://doi.org/10.1016/0167-7152(92)90153-V)
- [14] RIZZO, M. L. and SZÉKELY, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Stat.* **4** 1034–1055. MR2758432 <https://doi.org/10.1214/09-AOAS245>
- [15] RUZSA, I. Z. and SZÉKELY, G. J. (1982). Intersections of traces of random walks with fixed sets. *Ann. Probab.* **10** 132–136. MR0637381
- [16] RUZSA, I. Z. and SZÉKELY, G. J. (1983). Convolution quotients of nonnegative functions. *Monatsh. Math.* **95** 235–239. MR0712423 <https://doi.org/10.1007/BF01352002>
- [17] RUZSA, I. Z. and SZÉKELY, G. J. (1985). No distribution is prime. *Z. Wahrsch. Verw. Gebiete* **70** 263–269. MR0799149 <https://doi.org/10.1007/BF02451431>
- [18] RUZSA, I. Z. and SZÉKELY, G. J. (1988). *Algebraic Probability Theory*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, Chichester. MR0974112
- [19] SZÉKELY, G. J. (1986). *Paradoxes in Probability Theory and Mathematical Statistics. Mathematics and Its Applications (East European Series)* **15**. D. Reidel Publishing Co., Dordrecht. Translated from the Hungarian by Márta Alpár and Éva Unger. MR0880020 <https://doi.org/10.1007/978-94-017-2533-0>
- [20] SZÉKELY, G. J. (2005). Half of a coin: Negative probabilities. Invited paper. *Wilmott Magazine* **66–68**.
- [21] SZÉKELY, G. J. (2006). Student’s t-test for scale mixtures. In *The Second Erich L. Lehmann Symposium, May 19–22, 2004, Rice Univ.* (J. Rojo, ed.). *IMS Lecture Notes—Monograph Series* **49** 10–18.
- [22] SZÉKELY, G. J. and RIZZO, M. L. *The Energy of Data and Distance Correlation*. Chapman & Hall/CRC. Expected to be available in 2023.
- [23] SZÉKELY, G. J. and RIZZO, M. L. (2007). The uncertainty principle of game theory. *Amer. Math. Monthly* **114** 688–702. MR2354439 <https://doi.org/10.1080/00029890.2007.11920460>
- [24] SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 <https://doi.org/10.1214/09-AOAS312>
- [25] SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J. Statist. Plann. Inference* **143** 1249–1272. MR3055745 <https://doi.org/10.1016/j.jspi.2013.03.018>
- [26] SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 <https://doi.org/10.1214/14-AOS1255>
- [27] SZÉKELY, G. J. and RIZZO, M. L. (2017). The energy of data, invited paper. *Annu. Rev. Stat. Appl.* **4** 447–479.
- [28] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 <https://doi.org/10.1214/009053607000000505>

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: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

President-Elect: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

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

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

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors:* Enno Mammen, Institute for Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

The Annals of Applied Statistics. *Editor-in-Chief:* Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. *Editors:* Alice Guionnet, Unité de Mathématiques Pures et Appliquées, ENS de Lyon, Lyon, France. Christophe Garban, Institut Camille Jordan, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

The Annals of Applied Probability. *Editors:* Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. *Editor:* Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The IMS Bulletin. *Editor:* Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France



IMS members get a
40% discount
Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE