STATISTICAL SCIENCE

Volume 36, Number 3 August 2021

***Statistical Science* [ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 36, Number 3, August 2021. 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.

Statistical Science **Volume 36, Number 3 (339–492) August 2021**

Volume 36 **Number 3** **August 2021**

Statistical Science
2021, Vol. 36, No. 3, 339–343
https://doi.org/10.1214/20-STS798

Khinchin's 1929 Paper on Von Mises' Frequency Theory of Probability

Lukas M. Verburgt

Abstract. In 1929, a few years prior to his colleague Kolmogorov's *Grundbegriffe*, the leading Russian probabilist Khinchin published a paper in which he commented on the foundational ambitions of von Mises' frequency theory of probability. This brief introduction provides background and context for the English translation of Khinchin's historically revealing paper, published as an online supplement.

Key words and phrases: Khinchin, Kolmogorov, von Mises, foundations of probability theory, frequency theory of probability, history of mathematics.

REFERENCES

BERNHARDT, H. (1984). *Richard Von Mises und Sein Beitrag zur Grundlegung der Wahrscheinlichkeitsrechnung Im* 20. *Jahrhundert*. Habilitationsschrift, Berlin, Humboldt-Univ.

BINGHAM, N. H. (2001). Probability and statistics: Some thoughts at the turn of the millennium. In *Probability Theory* (*Roskilde*, 1998) (V. F. Hendricks, S. A. Pedersen and K. F. Jorgensen, eds.). *Synthese Lib*. **297** 15–49. Kluwer Academic, Dordrecht. MR1894707

CARNAP, R. (1934). *Logische Syntax der Sprache*. Springer, Vienna.

CHAUMONT, L., MAZLIAK, L. and YOR, M. (2007). Some aspects of the probabilistic work. In *Kolmogorov's Heritage in Mathematics* (E. Charpentier, A. Lesne and N. Nikolski, eds.) 41–66. Springer, Berlin. MR2376738 https://doi.org/10.1007/978-3-540-36351-4_3

CRAMÉR, H. (1962). A. I. Khinchin's work in mathematical probability. *Ann*. *Math*. *Stat*. **33** 1227–1237. MR0139512 https://doi.org/10.1214/aoms/1177704355

DOOB, J. L. (1996). The development of rigor in mathematical probability (1900–1950). *Amer. Math. Monthly* **103** 586–595. MR1404084 https://doi.org/10.2307/2974673

DÖRGE, K. (1930). Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. Erste Mitteilung: Theorie des Glücksspiels. *Math. Z.* **32** 232–258. MR1545164 https://doi.org/10.1007/BF01194632

GNEDENKO, B. V. (1961). Alexander Iacovlevich Khinchin. In *Proc*. 4*th Berkeley Sympos*. *Math*. *Statist*. *and Prob*., *Vol. II* 1–15. (1 plate). Univ. California Press, Berkeley, CA. MR0131956

HAUSDORFF, F. (2006). *Gesammelte Werke. Band V. Astronomie*, *Optik und Wahrscheinlichkeitstheorie*. Springer, Berlin. MR2229144

HOCHKIRCHEN, T. (1999). *Die Axiomatisierung der Wahrscheinlichkeitsrechnung und Ihre Kontexte. Von Hilberts sechstem Problem zu Kolmogoroffs Grundbegriffen. Studien zur Wissenschafts-, Sozial- und Bildungsgeschichte der Mathematik* [*Studies on the Scientific*, *Social and Educational History of Mathematics*] **13**. Vandenhoeck & Ruprecht, Göttingen. MR1738288

KAMKE, E. (1932). *Einführung in die Wahrscheinlichkeitstheorie*. Hirzel, Leipzig. MR0010917

KHINCHIN, A. YA. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund*. *Math*. **6** 9–20.

KHINCHIN, A. YA. (1926). Ideas of intuitionism and the struggle for a subject matter in modern mathematics (Russian). *Vestnik Kommunisticheskoy Akademii* **16** 184–192. English translation in [Verburgt & Hoppe-Kondrikova, 2016].

KHINTCHINE, A. YA. (1927). Über das Gesetz der großen Zahlen. *Math*. *Ann*. **96** 152–168. MR1512310 https://doi.org/10.1007/BF01209158

KHINCHIN, A. YA. (1929). The role and character of induction in mathematics (Russian). *Vestnik Kommunisticheskoy Academii* **1** 5–7.

KHINCHIN, A. YA. (1933). Zur mathematischen Begründung der statistischen Mechanik. *ZAMM Z. Angew. Math*. *Mech*. **13** 101–103.

KHINCHIN, A. YA. (1952). The method of arbitrary functions and the struggle against idealism in the theory of probability. (Russian). *Filosofskiye Sovremennoy Fiziki* 522–538.

KHINCHIN, A. YA. (1961). Mises' frequentist theory and the modern ideas in the theory of probability (Russian). *Voprosy Filosofii*. *Nauchno-tekhnicheskiy Zhurnal* **15** (1), 92–102, **15** (2): 77–89.

KHINCHIN, A. YA. (2004 [1936–1944]). Mises' frequentist theory and the modern ideas in the theory of probability. Translated by O. Sheynin, with footnotes by Reinhard Siegmund-Schultze. *Sci*. *Context* **17** 319–422. MR2106024

KHINCHIN, A. YA. (1929 [2020]). Mises' teachings on probability and the principles of statistical physics'. Translated from the original Russian by Laurent Mazliak, Lukas M. Verburgt and Olga Hoppe-Kondrikova.

LYUSTERNIK, L. A. (1967). The early years of the Moscow mathematics school. III. *Russian Math*. *Surveys* **22** 55–91.

MARTIN-LÖF, P. (1969). The literature on von Mises' Kollektivs revisited. *Theoria* **35** 12–37. MR0240841 https://doi.org/10.1111/j.1755-2567.1969.tb00357.x

PECHENKIN, A. (2014). *Leonid Isaakovich Mandelstam*: *Research*, *Teaching*, *Life*. Springer, Cham. MR3235276 https://doi.org/10.1007/978-3-319-00572-0

PHILLIPS, E. R. (1978). Nicolai Nicolaevich Luzin and the Moscow school of the theory of functions. *Historia Math*. **5** 275–305. MR0500530 https://doi.org/10.1016/0315-0860(78)90114-3

Lukas M. Verburgt is Postdoctoral Researcher, Freudenthal Institute, History and Philosophy of Science, Utrecht University, Princetonplein 5, 3584 CC Utrecht, Netherlands (e-mail: l.m.verburgt@uu.nl).

SENETA, E. (1994). Russian probability and statistics before Kolmogorov. In *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*. *Vol*. 2 (I. Grattan-Guinness, ed.) 1325–1334. Routledge, London. MR1469978

SENETA, E. (2004). Mathematics, religion, and Marxism in the Soviet Union in the 1930s. *Historia Math*. **31** 337–367. MR2079595 https://doi.org/10.1016/S0315-0860(03)00046-6

SHAFER, G. and VOVK, V. (2006). The sources of Kolmogorov's *Grundbegriffe*. *Statist*. *Sci*. **21** 70–98. MR2275967 https://doi.org/10.1214/088342305000000467

SIEGMUND-SCHULTZE, R. (2004). Mathematicians forced to philosophize: An introduction to Khinchin's paper on von Mises' theory of probability. *Sci*. *Context* **17** 373–390. MR2106023 https://doi.org/10.1017/S0269889704000171

SIEGMUND-SCHULTZE, R. (2006). Probability in 1919/20: The von Mises–Pólya controversy. *Arch*. *Hist*. *Exact Sci*. **60** 431–515. MR2238933 https://doi.org/10.1007/s00407-006-0112-x

SIEGMUND-SCHULTZE, R. (2010). Sets versus trial sequences, Hausdorff versus von Mises: "pure" mathematics prevails in the foundations of probability around 1920. *Hist*. *Math*. **37** 204–241. MR2609068 https://doi.org/10.1016/j.hm.2009.11.007

TORNIER, E. (1930). Eine neue Grundlegung der Wahrscheinlichkeitsrechnung. *Z. Phys*. **63** 697–705.

VAN ATTEN, M. (2015). *Brouwer Meets Husserl*: *On the Phenomenology of Choice Sequences*. Springer, Cham.

VAN LAMBALGEN, M. (1996). Randomness and foundations of probability: Von Mises' axiomatisation of random sequences. In *Statistics*, *Probability and Game Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 347–367. IMS, Hayward, CA. MR1481789 https://doi.org/10.1214/lnms/1215453582

VERBURGT, L. M. (2021). Supplement to "Khinchin's 1929 Paper on Von Mises' Frequency Theory of Probability." https://doi.org/10.1214/20-STS798SUPP

VERBURGT, L. M. and HOPPE-KONDRIKOVA, O. (2016). On A. Ya. Khinchin's paper 'Ideas of intuitionism and the struggle for a subject matter in contemporary mathematics' (1926): A translation with introduction and commentary. *Historia Math*. **43** 369–398. MR3567936 https://doi.org/10.1016/j.hm.2016.07.002

VERE-JONES, D. (2005). Moscow school of probability theory. In *Encyclopedia of Statistical Sciences*. *Vol*. 6 (S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic and N. L. Johnson, eds.) 1–16. Wiley, Hoboken, NJ.

VON MISES, R. (1928). *Wahrscheinlichkeit*, *Statistik und Wahrheit*. Springer, Vienna.

VON MISES, R. (1931). *Wahrscheinlichkeit und Ihre Anwendung in der Statistik und Theoretischen Physik*. Deuticke, Vienna.

VON MISES, R. (1936). *Wahrscheinlichkeit*, *Statistik und Wahrheit*, 2nd ed. Springer, Vienna.

VON PLATO, J. (1994). *Creating Modern Probability*: *Its Mathematics*, *Physics and Philosophy in Historical Perspective*. *Cambridge Studies in Probability*, *Induction*, *and Decision Theory*. Cambridge Univ. Press, Cambridge. MR1265718 https://doi.org/10.1017/CBO9780511609107

VUCINICH, A. (1970). *Science in Russian Culture*, 1861–1917. Stanford Univ. Press, Stanford, CA.

VUCINICH, A. (1999). Mathematics and dialectics in the Soviet Union: The pre-Stalin period. *Historia Math*. **26** 107–124. MR1687137 https://doi.org/10.1006/hmat.1999.2229

YUSHKEVICH, A. P. (1993). Encounters with mathematicians. In *Golden Years of Moscow Mathematics* (S. Zdravkovska and P. L. Duren, eds.). *Hist*. *Math*. **6** 1–33. Amer. Math. Soc., Providence, RI. MR1246564

Statistical Science
2021, Vol. 36, No. 3, 344–359
https://doi.org/10.1214/20-STS805

A Problem in Forensic Science Highlighting the Differences between the Bayes Factor and Likelihood Ratio

Danica M. Ommen and Christopher P. Saunders

Abstract. This article is aimed at the growing number of statisticians interested in the important problem of interpreting evidence within the forensic identification of source problems. Our purpose is to formalize these forensic problems as statistical model selection problems. We use two different classes of statistics for quantifying the evidential value, the likelihood ratio and Bayes Factor. In forensics, both are commonly called the "likelihood ratio approach" and "the value of evidence" despite using different definitions of probability. In statistics, they are closely related to the traditional likelihood ratio from pattern recognition and the Bayes Factor used in model selection. For two different problem frameworks typical in forensic science, the common source and the specific source problems, we show the Bayes Factor and likelihood ratio are not equivalent, and highlight several interesting links between them. These contributions will help to elucidate the effects of choosing different definitions of probability when addressing the forensic identification of source problems. The broader population of statisticians may find this paper interesting as an introduction to forensic applications and for illuminating the connections between model selection methods from two different paradigms of statistics, particularly in view of the active recent discussions on the connections among Bayesian, Fiducial, Frequentist (BFF) approaches.

Key words and phrases: Bayesian, Frequentist, model selection, consistency, credible interval, forensics, common source, specific source.

REFERENCES

[1] AITKEN, C. G. G., ROBERTS, P. and JACKSON, G. (2010). *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*; *Guidance for Judges*, *Lawyers*, *Forensic Scientists and Expert Witnesses*, 1st ed. Royal Statistical Society's Working Group on Statistics and the Law, London, UK.

[2] AITKEN, C. G. G. and TARONI, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd ed. Wiley, West Sussex, UK.

[3] ASH, R. B. (2000). *Probability and Measure Theory*, 2nd ed. Harcourt/Academic Press, Burlington, MA. MR1810041

[4] BERGER, C. E. H. and SLOOTEN, K. (2016). The LR does not exist. *Sci*. *Justice* **56** 388–391.

[5] BIEDERMANN, A., BOZZA, S., TARONI, F. and AITKEN, C. G. G. (2016). Reframing the debate: A question of probability, not of likelihood ratio. *Sci*. *Justice* **56** 392–396.

[6] CHIB, S. and KUFFNER, T. A. (2016). Bayes Factor consistency. Technical report. https://arxiv.org/abs/1607.00292v1.

[7] CURRAN, J. M. (2016). Admitting to uncertainty in the LR. *Sci*. *Justice* **56** 380–382. https://doi.org/10.1016/j.scijus.2016.05.005

[8] DETTMAN, J. R., CASSABAUM, A. A., SAUNDERS, C. P., SNYDER, D. L. and BUSCAGLIA, J. (2014). Forensic discrimination of copper wire using trace element concentrations. *Analytical Chemistry* **86** 8176–8182.

[9] EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES (2015). *ENFSI Guideline for Evaluative Reporting in Forensic Science*.

[10] KAFADAR, K. (2018). The critical role of statistics in demonstrating the reliability of expert evidence. *Fordham Law Review* **86** 1617–1637.

[11] KASS, R. E. and RAFTERY, A. E. (1995). Bayes factors. *J*. *Amer*. *Statist*. *Assoc*. **90** 773–795. MR3363402 https://doi.org/10.1080/01621459.1995.10476572

[12] LINDLEY, D. V. (1977). A problem in forensic science. *Biometrika* **64** 207–213. MR0518935 https://doi.org/10.1093/

Danica M. Ommen is Assistant Professor, Department of Statistics, Iowa State University, Ames, Iowa 50011, USA (e-mail: dmommen@iastate.edu). Christopher P. Saunders is Associate Professor, Department of Mathematics & Statistics, South Dakota State University, Brookings, South Dakota 57007, USA (e-mail: christopher.saunders@sdstate.edu).

biomet/64.2.207

[13] LUND, S. P. and IYER, H. (2017). Likelihood ratio as weight of forensic evidence: A closer look. *J. Res. Natl. Inst. Stand. Technol.* **122** 1–32.

[14] MORRISON, G. S. and ENZINGER, E. (2016). What should a forensic practitioner's likelihood ratio be? *Sci. Justice* **56** 374–379.

[15] NATIONAL RESEARCH COUNCIL COMMITTEE ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIENCES COMMUNITY (2009). *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, D.C.

[16] NEWTON, M. A. and RAFTERY, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* **56** 3–48. MR1257793

[17] OMMEN, D. M. and SAUNDERS, C. P. (2018). Building a unified statistical framework for the forensic identification of source problems. *Law Probab. Risk* **17** 179–197.

[18] OMMEN, D. M. and SAUNDERS, C. P. (2021). Supplement to "A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio." https://doi.org/10.1214/20-STS805SUPP

[19] OMMEN, D. M., SAUNDERS, C. P. and NEUMANN, C. (2016). An argument against presenting interval quantifications as a surrogate for the value of evidence. *Sci. Justice* **56** 383–387. https://doi.org/10.1016/j.scijus.2016.07.001

[20] OMMEN, D. M., SAUNDERS, C. P. and NEUMANN, C. (2017). The characterization of Monte Carlo errors for the quantification of the value of forensic evidence. *J. Stat. Comput. Simul.* **87** 1608–1643. MR3625235 https://doi.org/10.1080/00949655.2017.1280036

[21] PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Executive Office of the President of the United States, Sept. 2016.

[22] ROYALL, R. M. (1997). *Statistical Evidence: A Likelihood Paradigm. Monographs on Statistics and Applied Probability* **71**. CRC Press, London. MR1629481

[23] SJERPS, M. J., ALBERINK, I., BOLCK, A., STOEL, R., VERGEER, P. and VAN ZANTEN, J. H. (2016). Uncertainty and LR; to integrate or not to integrate, that's the question. *Law Probab. Risk* **15** 23–29.

[24] SWANENBURG, J. (2020). Frequentistische en bayesiaanse statistiek in de forensische statistiek. Master's thesis, Technical Univ. Delft. http://resolver.tudelft.nl/uuid:11b29c61-8158-46ca-b74f-00e4d3905d7f.

[25] SWOFFORD, H. J., KOERNTNER, A. J., ZEMP, F., AUSDEMORE, M., LIU, A. and SALYARDS, M. J. (2018). A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation. *Forensic Sci. Int.* **287** 113–126.

[26] TARONI, F., BOZZA, S., BIEDERMANN, A. and AITKEN, C. G. G. (2016). Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. *Law Probab. Risk* **15** 1–16.

[27] TAYLOR, D., HICKS, T. and CHAMPOD, C. (2016). Using sensitivity analyses in Bayesian networks to highlight the impact of data paucity and direct future analyses: A contribution to the debate on measuring and reporting the precision of likelihood ratios. *Sci. Justice* **56** 402–410. https://doi.org/10.1016/j.scijus.2016.06.010

[28] VAN DEN HOUT, A. and ALBERINK, I. (2016). Posterior distributions for likelihood ratios in forensic science. *Sci. Justice* **56** 397–401. https://doi.org/10.1016/j.scijus.2016.06.011

[29] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[30] VUONG, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica* **57** 307–333. MR0996939 https://doi.org/10.2307/1912557

Statistical Science
2021, Vol. 36, No. 3, 360–378
https://doi.org/10.1214/20-STS795

A Horse Race between the Block Maxima Method and the Peak–over–Threshold Approach

Axel Bücher and Chen Zhou

Abstract. Classical extreme value statistics consists of two fundamental approaches: the block maxima (BM) method and the peak-over-threshold (POT) approach. It seems to be general consensus among researchers in the field that the POT approach makes use of extreme observations more efficiently than the BM method. We shed light on this discussion from three different perspectives. First, based on recent theoretical results for the BM method, we provide a theoretical comparison in i.i.d. scenarios. We argue that the data generating process may favour either one or the other approach. Second, if the underlying data possesses serial dependence, we argue that the choice of a method should be primarily guided by the ultimate statistical interest: for instance, POT is preferable for quantile estimation, while BM is preferable for return level estimation. Finally, we discuss the two approaches for multivariate observations and identify various open ends for future research.

Key words and phrases: Extreme value statistics, extreme value index, extremal index, stationary time series.

REFERENCES

BALKEMA, A. A. and DE HAAN, L. (1974). Residual life time at great age. *Ann. Probab.* **2** 792–804. MR0359049 https://doi.org/10.1214/aop/1176996548

BALKEMA, A. A. and RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Probab.* **14** 309–319. MR0438425 https://doi.org/10.1017/s002190020010498x

BEIRLANT, J., CAEIRO, F. and GOMES, M. I. (2012). An overview and open research topics in statistics of univariate extremes. *REVSTAT* **10** 1–31. MR2912369

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

BEIRLANT, J., ESCOBAR-BACH, M., GOEGEBEUR, Y. and GUILLOU, A. (2016). Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* **143** 453–466. MR3431445 https://doi.org/10.1016/j.jmva.2015.10.006

BERGHAUS, B. and BÜCHER, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* **46** 2307–2335. MR3845019 https://doi.org/10.1214/17-AOS1621

BÜCHER, A. (2014). A note on nonparametric estimation of bivariate tail dependence. *Stat. Risk Model.* **31** 151–162. MR3213603 https://doi.org/10.1515/strm-2013-1143

BÜCHER, A. and SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528. MR3252823 https://doi.org/10.1007/s10687-014-0195-8

BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. MR3737387 https://doi.org/10.1007/s10687-017-0292-6

BÜCHER, A. and SEGERS, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Stat.* **12** 1098–1125. MR3780041 https://doi.org/10.1214/18-EJS1415

BÜCHER, A. and SEGERS, J. (2018b). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. MR3706798 https://doi.org/10.3150/16-BEJ903

BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivariate Anal.* **173** 604–619. MR3959186 https://doi.org/10.1016/j.jmva.2019.04.011

BÜCHER, A. and ZHOU, C. (2021). Supplement to "A horse race between the block maxima method and the peak–over–threshold approach." https://doi.org/10.1214/20-STS795SUPP

BUISHAND, T. A., DE HAAN, L. and ZHOU, C. (2008). On spatial extremes: With application to a rainfall problem. *Ann. Appl. Stat.* **2** 624–642. MR2524349 https://doi.org/10.1214/08-AOAS159

Axel Bücher is Professor of Mathematical Statistics, Heinrich-Heine-Universität Düsseldorf, Mathematisches Institut, Universitätsstr. 1, 40225 Düsseldorf, Germany (e-mail: axel.buecher@hhu.de). Chen Zhou is Professor of Mathematical Statistics and Risk Management, Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands (e-mail: zhou@ese.eur.nl; c.zhou@dnb.nl).

CAIRES, S. (2009). A comparative simulation study of the annual maxima and the peaks-over-threshold methods Technical report, SBW-Belastingen: Subproject "Statistics". Deltares Report 1200264-002.

CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985 https://doi.org/10.1093/biomet/84.3.567

CHAVEZ-DEMOULIN, V. and GUILLOU, A. (2018). Extreme quantile estimation for β-mixing time series and applications. *Insurance Math. Econom.* **83** 59–74. MR3886487 https://doi.org/10.1016/j.insmatheco.2018.09.004

COLES, S. G. and TAWN, J. A. (1996). Modelling extremes of the areal rainfall process. *J. Roy. Statist. Soc. Ser. B* **58** 329–347. MR1377836

DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* **76** 226–248. MR1821820 https://doi.org/10.1006/jmva.2000.1903

DE FONDEVILLE, R. and DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105** 575–592. MR3842886 https://doi.org/10.1093/biomet/asy026

DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12** 1194–1204. MR0757776

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L. and LIN, T. (2003). Weak consistency of extreme value estimators in $C[0, 1]$. *Ann. Statist.* **31** 1996–2012. MR2036397 https://doi.org/10.1214/aos/1074290334

DE HAAN, L., MERCADIER, C. and ZHOU, C. (2016). Adapting extreme value statistics to financial time series: Dealing with bias and serial dependence. *Finance Stoch.* **20** 321–354. MR3479324 https://doi.org/10.1007/s00780-015-0287-6

DE HAAN, L., NEVES, C. and PENG, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99** 1260–1275. MR2419346 https://doi.org/10.1016/j.jmva.2007.08.003

DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290 https://doi.org/10.1007/BF00533086

DE HAAN, L. and STADTMÜLLER, U. (1996). Generalized regular variation of second order. *J. Aust. Math. Soc. A* **61** 381–395. MR1420345

DIEKER, A. B. and MIKOSCH, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes* **18** 301–314. MR3351818 https://doi.org/10.1007/s10687-015-0214-4

DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. MR3322325 https://doi.org/10.3150/13-BEJ573

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103** 303–317. MR3509888 https://doi.org/10.1093/biomet/asw008

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. ArXiv e-prints.

DOMBRY, C., ÉYI-MINKO, F. and RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124. MR3034327 https://doi.org/10.1093/biomet/ass067

DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. MR3961227 https://doi.org/10.3150/18-BEJ1032

DRAISMA, G., DE HAAN, L., PENG, L. and PEREIRA, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2** 367–404. MR1776855 https://doi.org/10.1023/A:1009900215680

DREES, H. (1998). On smooth statistical tail functionals. *Scand. J. Stat.* **25** 187–210. MR1614276 https://doi.org/10.1111/1467-9469.00097

DREES, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875 https://doi.org/10.1214/aoap/1019487617

DREES, H. and DE HAAN, L. (2015). Estimating failure probabilities. *Bernoulli* **21** 957–1001. MR3338653 https://doi.org/10.3150/13-BEJ594

DREES, H., DE HAAN, L. and LI, D. (2003). On large deviation for extremes. *Statist. Probab. Lett.* **64** 51–62. MR1995809 https://doi.org/10.1016/S0167-7152(03)00130-5

DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. MR1762911 https://doi.org/10.1214/aos/1016120372

DREES, H., FERREIRA, A. and DE HAAN, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14** 1179–1201. MR2071420 https://doi.org/10.1214/105051604000000279

DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. MR1619974 https://doi.org/10.1006/jmva.1997.1708

DREES, H. and KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.* **75** 149–172. MR1632189 https://doi.org/10.1016/S0304-4149(98)00017-9

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886 https://doi.org/10.1214/09-AOS788

EINMAHL, J. H. J., DE HAAN, L. and PITERBARG, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29** 1401–1423. MR1873336 https://doi.org/10.1214/aos/1013203459

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

EINMAHL, J. H. J. and LIN, T. (2006). Asymptotic normality of extreme value estimators on $C[0, 1]$. *Ann. Statist.* **34** 469–492. MR2275250 https://doi.org/10.1214/009053605000000831

EINMAHL, J. H. J. and SEGERS, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* **37** 2953–2989. MR2541452 https://doi.org/10.1214/08-AOS677

FALK, M. and GUILLOU, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivariate Anal.* **99** 715–734. MR2406079 https://doi.org/10.1016/j.jmva.2007.03.009

FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737. MR3263087 https://doi.org/10.3150/13-BEJ538

FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298. MR3285607 https://doi.org/10.1214/14-AOS1280

FERRO, C. A. T. and SEGERS, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 545–556. MR1983763 https://doi.org/10.1111/1467-9868.00401

FOUGÈRES, A.-L., DE HAAN, L. and MERCADIER, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43** 903–934. MR3325714 https://doi.org/10.1214/14-AOS1305

GENEST, C., GHOUDI, K. and RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. MR1366280 https://doi.org/10.1093/biomet/82.3.543

GENEST, C. and SEGERS, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453 https://doi.org/10.1214/08-AOS672

GENEST, C. and SEGERS, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.* **101** 1837–1845. MR2651959 https://doi.org/10.1016/j.jmva.2010.03.018

GINÉ, E., HAHN, M. G. and VATAN, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields* **87** 139–165. MR1080487 https://doi.org/10.1007/BF01198427

GOMES, M. I., DE HAAN, L. and RODRIGUES, L. H. (2008). Tail index estimation for heavy-tailed models: Accommodation of bias in weighted log-excesses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 31–52. MR2412630

GUDENDORF, G. and SEGERS, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Plann. Inference* **142** 3073–3085. MR2956794 https://doi.org/10.1016/j.jspi.2012.05.007

GUILLOU, A. and HALL, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 293–305. MR1841416 https://doi.org/10.1111/1467-9868.00286

GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press, New York. MR0096342

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084. MR0751294 https://doi.org/10.1214/aos/1176346723

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. MR0797563 https://doi.org/10.2307/1269706

HSING, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569. MR1126337 https://doi.org/10.1214/aos/1176348261

HSING, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* **21** 2043–2071. MR1245780 https://doi.org/10.1214/aos/1176349409

HUANG, X. (1992). Statistics of bivariate extreme values Ph. D. thesis, Tinbergen Institute Research Series, Netherlands.

HUSER, R. and DAVISON, A. C. (2014). Space-time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 439–461. MR3164873 https://doi.org/10.1111/rssb.12035

HUSER, R., DAVISON, A. C. and GENTON, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes* **19** 79–103. MR3454032 https://doi.org/10.1007/s10687-015-0230-4

HUSER, R. and WADSWORTH, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *J. Amer. Statist. Assoc.* **114** 434–444. MR3941266 https://doi.org/10.1080/01621459.2017.1411813

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440 https://doi.org/10.1214/09-AOP455

LEADBETTER, M. R. (1973/74). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **28** 289–303. MR0362465 https://doi.org/10.1007/BF00532947

LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. MR0722133 https://doi.org/10.1007/BF00532484

NANDAGOPALAN, S. (1994). On the multivariate extremal index. *J. Res. Natl. Inst. Stand. Technol.* **99** 543–543.

NAVEAU, P., GUILLOU, A., COOLEY, D. and DIEBOLT, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika* **96** 1–17. MR2482131 https://doi.org/10.1093/biomet/asp001

NORTHROP, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18** 585–603. MR3418769 https://doi.org/10.1007/s10687-015-0221-5

OESTING, M., SCHLATHER, M. and ZHOU, C. (2018). Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli* **24** 1497–1530. MR3706800 https://doi.org/10.3150/16-BEJ905

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202 https://doi.org/10.1198/jasa.2009.tm08577

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

PICKANDS, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the* 43rd *Session of the International Statistical Institute*, *Vol.* 2 (*Buenos Aires*, 1981) **49** 859–878, 894–902. MR0820979

PRESCOTT, P. and WALDEN, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. MR0601119 https://doi.org/10.1093/biomet/67.3.723

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics*: *With Applications to Nonparametric Statistics*. *Springer Series in Statistics*. Springer, New York. MR0988164 https://doi.org/10.1007/978-1-4613-9620-8

RESNICK, S. I. (1987). *Extreme Values*, *Regular Variation*, *and Point Processes*. *Applied Probability. a Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

RESNICK, S. and STĂRICĂ, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Probab.* **8** 1156–1183. MR1661160 https://doi.org/10.1214/aoap/1028903376

ROBERT, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.* **37** 271–310. MR2488352 https://doi.org/10.1214/07-AOS551

ROBERT, C. Y., SEGERS, J. and FERRO, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. MR2540849 https://doi.org/10.1214/08-EJS345

ROOTZÉN, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.* **119** 468–490. MR2494000 https://doi.org/10.1016/j.spa.2008.03.003

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018a). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *J. Multivariate Anal.* **165** 117–131. MR3768756 https://doi.org/10.1016/j.jmva.2017.12.003

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018b). Multivariate peaks over thresholds models. *Extremes* **21** 115–145. MR3764613 https://doi.org/10.1007/s10687-017-0294-4

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668 https://doi.org/10.3150/bj/1161614952

SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Stat.* **33** 307–335. MR2279645 https://doi.org/10.1111/j.1467-9469.2005.00483.x

SMITH, R. L. and WEISSMAN, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56** 515–528. MR1278224

STEPHENSON, A. (2018). evd: Functions for Extreme Value Distributions. R package version 2.3-3.

SÜVEGES, M. (2007). Likelihood estimation of the extremal index. *Extremes* **10** 41–55. MR2407640 https://doi.org/10.1007/s10687-007-0034-2

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870. MR3431558 https://doi.org/10.1093/biomet/asv045

WEISSMAN, I. and NOVAK, S. YU. (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inference* **66** 281–288. MR1614480 https://doi.org/10.1016/S0378-3758(97)00095-5

ZHANG, D., WELLS, M. T. and PENG, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J. Multivariate Anal.* **99** 577–588. MR2406072 https://doi.org/10.1016/j.jmva.2006.09.011

ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* **49** 295–320. MR4206679 https://doi.org/10.1214/20-AOS1957

Statistical Science
2021, Vol. 36, No. 3, 379–399
https://doi.org/10.1214/20-STS788

A Hybrid Scan Gibbs Sampler for Bayesian Models with Latent Variables

Grant Backlund, James P. Hobert, Yeun Ji Jung and Kshitij Khare

Abstract. Gibbs sampling is a widely popular Markov chain Monte Carlo algorithm that can be used to analyze intractable posterior distributions associated with Bayesian hierarchical models. There are two standard versions of the Gibbs sampler: The systematic scan (SS) version, where all variables are updated at each iteration, and the random scan (RS) version, where a single, randomly selected variable is updated at each iteration. The literature comparing the theoretical properties of SS and RS Gibbs samplers is reviewed, and an alternative *hybrid scan* Gibbs sampler is introduced, which is particularly well suited to Bayesian models with latent variables. The word "hybrid" reflects the fact that the scan used within this algorithm has both systematic and random elements. Indeed, at each iteration, one updates the entire set of latent variables, along with a randomly chosen block of the remaining variables. The hybrid scan (HS) Gibbs sampler has important advantages over the two standard scan Gibbs samplers. First, the HS algorithm is often easier to analyze from a theoretical standpoint. In particular, it can be much easier to establish the geometric ergodicity of a HS Gibbs Markov chain than to do the same for the corresponding SS and RS versions. Second, the sandwich methodology developed in (*Ann. Statist.* **36** (2008) 532–554), which is also reviewed, can be applied to the HS Gibbs algorithm (but not to the standard scan Gibbs samplers). It is shown that, under weak regularity conditions, adding sandwich steps to the HS Gibbs sampler always results in a theoretically superior algorithm. Three specific Bayesian hierarchical models of varying complexity are used to illustrate the results. One is a simple location-scale model for data from the Student's t distribution, which is used as a pedagogical tool. The other two are sophisticated, yet practical Bayesian regression models.

Key words and phrases: Geometric ergodicity, heavy-tailed errors, linear mixed model, Markov chain Monte Carlo, sandwich algorithm, shrinkage prior.

REFERENCES

ABRAHAMSEN, T. and HOBERT, J. P. (2019). Fast Monte Carlo Markov chains for Bayesian shrinkage models with random effects. *J. Multivariate Anal.* **169** 61–80. MR3875587 https://doi.org/10.1016/j.jmva.2018.08.014

ANDRIEU, C. (2016). On random- and systematic-scan samplers. *Biometrika* **103** 719–726. MR3551794 https://doi.org/10.1093/biomet/asw019

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

BACKLUND, G. (2020). Analysis of Markov Chain Monte Carlo Algorithms For Bayesian Regression Models with Heavy-tailed and Skewed Error Distributions Ph.D. thesis, University of Florida.

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2012). Bayesian shrinkage. Available at arXiv:1212.6088.

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *J. Amer. Statist. Assoc.* **110** 1479–1490. MR3449048 https://doi.org/10.

Grant Backlund is a PhD Student, Department of Statistics, University of Florida, Gainesville, Florida, USA 32611 (e-mail: grantback21@ufl.edu). James P. Hobert is Professor, Department of Statistics, University of Florida, Gainesville, Florida, USA 32611 (e-mail: jhobert@stat.ufl.edu). Yeun Ji Jung is Associate, Model Governance Group, JP Morgan Chase & Co, New York, New York, USA 10172 (e-mail: snow0907@ufl.edu). Kshitij Khare is Associate Professor, Department of Statistics, University of Florida, Gainesville, Florida, USA (e-mail: kdkhare@stat.ufl.edu).

1080/01621459.2014.960967

DA SILVA FERREIRA, C., BOLFARINE, H. and LACHOS, V. H. (2011). Skew scale mixtures of normal distributions: Properties and estimation. *Stat. Methodol.* **8** 154–171. MR2769277 https://doi.org/10.1016/j.stamet.2010.09.001

FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823 https://doi.org/10.1214/08-STS257

GEYER, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 3–48. CRC Press, Boca Raton, FL. MR2858443

GREENWOOD, P. E., MCKEAGUE, I. W. and WEFELMEYER, W. (1998). Information bounds for Gibbs samplers. *Ann. Statist.* **26** 2128–2156. MR1700224 https://doi.org/10.1214/aos/1024691464

GRIFFIN, J. E. and BROWN, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Anal.* **5** 171–188. MR2596440 https://doi.org/10.1214/10-BA507

HE, B., DE SA, C., MITLIAGKAS, I. and RÉ, C. (2016). Scan order in Gibbs sampling: Models in which it matters and bounds on how much. *Adv. Neural Inf. Process. Syst.* **29** 1–29.

HOBERT, J. P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 253–293. CRC Press, Boca Raton, FL. MR2858452

HOBERT, J. P. and MARCHEV, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *Ann. Statist.* **36** 532–554. MR2396806 https://doi.org/10.1214/009053607000000569

HOBERT, J. P., JUNG, Y. J., KHARE, K. and QIN, Q. (2018). Convergence analysis of MCMC algorithms for Bayesian multivariate linear regression with non-Gaussian errors. *Scand. J. Stat.* **45** 513–533. MR3858944 https://doi.org/10.1111/sjos.12310

JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447 https://doi.org/10.1214/ss/1015346317

JONES, G. L., ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Convergence of conditional Metropolis–Hastings samplers. *Adv. in Appl. Probab.* **46** 422–445. MR3215540 https://doi.org/10.1239/aap/1401369701

JUNG, Y. J. (2015). Convergence Analysis of Markov Chain Monte Carlo Algorithms For Bayesian Regression Models with Non-Gaussian Errors Ph.D. thesis, Univ. Florida, Gainesville, FL.

JUNG, Y. J. and HOBERT, J. P. (2014). Spectral properties of MCMC algorithms for Bayesian linear regression with generalized hyperbolic errors. *Statist. Probab. Lett.* **95** 92–100. MR3262955 https://doi.org/10.1016/j.spl.2014.07.034

KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Ann. Statist.* **39** 2585–2606. MR2906879 https://doi.org/10.1214/11-AOS916

ŁATUSZYŃSKI, K., MIASOJEDOW, B. and NIEMIRO, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli* **19** 2033–2066. MR3129043 https://doi.org/10.3150/12-BEJ442

ŁATUSZYŃSKI, K., ROBERTS, G. O. and ROSENTHAL, J. S. (2013). Adaptive Gibbs samplers and related MCMC methods. *Ann. Appl. Probab.* **23** 66–98. MR3059204 https://doi.org/10.1214/11-AAP806

LEVINE, R. A. (2005). A note on Markov chain Monte Carlo sweep strategies. *J. Stat. Comput. Simul.* **75** 253–262. MR2134639 https://doi.org/10.1080/0094965042000223671

LIU, J. S., WONG, W. H. and KONG, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* **57** 157–169. MR1325382

LIU, J. S. and WU, Y. N. (1999). Parameter expansion for data augmentation. *J. Amer. Statist. Assoc.* **94** 1264–1274. MR1731488 https://doi.org/10.2307/2669940

MAIRE, F., DOUC, R. and OLSSON, J. (2014). Comparison of asymptotic variances of inhomogeneous Markov chains with application to Markov chain Monte Carlo methods. *Ann. Statist.* **42** 1483–1510. MR3262458 https://doi.org/10.1214/14-AOS1209

MENG, X.-L. and VAN DYK, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika* **86** 301–320. MR1705351 https://doi.org/10.1093/biomet/86.2.301

MEYN, S. P. and TWEEDIE, R. L. (2012). *Markov Chains and Stochastic Stability*, 2nd ed. ed. Springer, London.

MIRA, A. and GEYER, C. J. (1999). Ordering Monte Carlo Markov chains Tech. Rep. No. 632, School of Statistics, Univ. Minnesota, Minneapolis, MN.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

PAL, S., KHARE, K. and HOBERT, J. P. (2015). Improving the data augmentation algorithm in the two-block setup. *J. Comput. Graph. Statist.* **24** 1114–1133. MR3432932 https://doi.org/10.1080/10618600.2014.955177

ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Markov-chain Monte Carlo: Some practical implications of theoretical results. *Canad. J. Statist.* **26** 5–31. MR1624414 https://doi.org/10.2307/3315667

ROBERTS, G. O. and ROSENTHAL, J. S. (2015). Surprising convergence properties of some simple Gibbs samplers under various scans. *Int. J. Stat. Probab.* **5** 51–60.

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

VAN DYK, D. A. and MENG, X.-L. (2001). The art of data augmentation (with discussion). *J. Comput. Graph. Statist.* **10** 1–111. MR1936358 https://doi.org/10.1198/10618600152418584

Statistical Science
2021, Vol. 36, No. 3, 400–420
https://doi.org/10.1214/20-STS793

Maximum Likelihood Multiple Imputation: Faster Imputations and Consistent Standard Errors Without Posterior Draws

Paul T. von Hippel and Jonathan W. Bartlett

Abstract. Multiple imputation (MI) is a method for repairing and analyzing data with missing values. MI replaces missing values with a sample of random values drawn from an *imputation model*. The most popular form of MI, which we call *posterior draw multiple imputation* (PDMI), draws the parameters of the imputation model from a Bayesian posterior distribution. An alternative, which we call *maximum likelihood multiple imputation* (MLMI), estimates the parameters of the imputation model using maximum likelihood (or equivalent). Compared to PDMI, MLMI is faster and yields slightly more efficient point estimates.

A past barrier to using MLMI was the difficulty of estimating the standard errors of MLMI point estimates. We derive, implement and evaluate three consistent standard error formulas: (1) one combines variances within and between the imputed datasets, (2) one uses the score function and (3) one uses the bootstrap with two imputations of each bootstrapped sample. Formula (1) modifies for MLMI a formula that has long been used under PDMI, while formulas (2) and (3) can be used without modification under either PDMI or MLMI. We have implemented MLMI and the standard error estimators in the *mlmi* and *bootImpute* packages for R.

Key words and phrases: Missing data, incomplete data.

REFERENCES

[1] ANDERSON, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52** 200–203. MR0087286

[2] ARBUCKLE, J. L. (1996). Full information estimation in the presence of incomplete data. In *Advanced Structural Equation Modeling: Issues and Techniques* 243–277.

[3] BARNARD, J. and RUBIN, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika* **86** 948–955. MR1741991 https://doi.org/10.1093/biomet/86.4.948

[4] BARTLETT, J. W. (2019). bootImpute: Bootstrap inference for multiple imputation. Comprehensive R Archive Network. https://CRAN.R-project.org/package=bootImpute.

[5] BARTLETT, J. W. (2019). mlmi: Maximum likelihood multiple imputation. Comprehensive R Archive Network. https://CRAN.R-project.org/package=mlmi.

[6] BARTLETT, J. W. and KEOGH, R. H. (2019). smcfcs: Multiple imputation of covariates by substantive model compatible fully conditional specification. Comprehensive R Archive Network. https://CRAN.R-project.org/package=smcfcs.

[7] BARTLETT, J. W., SEAMAN, S. R., WHITE, I. R. and CARPENTER, J. R. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat. Methods Med. Res.* **24** 462–487. MR3372102 https://doi.org/10.1177/0962280214521348

[8] CENTRE FOR LONGITUDINAL STUDIES, INSTITUTE OF EDUCATION, UNIVERSITY OF LONDON (2017). Millennium cohort study: Second survey 2003-2005, UK Data Service, 9th edition. SN: 5350. https://doi.org/10.5255/UKDA-SN-5350-4

[9] 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. MR0501537

[10] EFRON, B. (1994). Missing data, imputation, and the bootstrap. *J. Amer. Statist. Assoc.* **89** 463–479. MR1294072

[11] EID, S. (2016). Mult[i]ple Imputation taking forever!! Retrieved May 8, 2017, from http://www.statalist.org/forums/forum/general-stata-discussion/general/1330365-multple-imputation-taking-forever.

[12] ENDERS, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Struct. Equ. Model.* **8** 128–141.

[13] FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, London. MR0346954

[14] HARRIS, J. A. (1913). On the calculation of intra-class and inter-class coefficients of correlation from class moments when the

Paul T. von Hippel is Associate Professor, LBJ School of Public Affairs, University of Texas, Austin, Texas, USA 78712, (e-mail: paulvonhippel@utexas.edu). Jonathan W. Bartlett is Reader in Statistics, University of Bath, BA2 7AY, UK (e-mail: j.w.bartlett@bath.ac.uk).

number of possible combinations is large. *Biometrika* **9** 446–472. https://doi.org/10.2307/2331901

[15] HEITJAN, D. F. and LITTLE, R. J. A. (1991). Multiple imputation for the fatal accident reporting system. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **40** 13–29.

[16] HONAKER, J. and KING, G. (2010). What to do about missing values in time-series cross-section data. *Amer. J. Polit. Sci.* **54** 561–581.

[17] HONAKER, J., KING, G. and BLACKWELL, M. (2015). AMELIA II: A Program for missing data, version 1.7.4.

[18] HUANG, J. (2015). How to speed up multiple imputation process. Retrieved May 8, 2017, from http://www.statalist.org/forums/forum/general-stata-discussion/general/1305705-how-to-speed-up-multiple-imputation-process.

[19] SAS INSTITUTE (2000). The MI procedure for SAS version 8.1. Cary, NC.

[20] KIM, J. K. (2004). Finite sample properties of multiple imputation estimators. *Ann. Statist.* **32** 766–783. MR2060177 https://doi.org/10.1214/009053604000000175

[21] KIM, J. K. and RAO, J. N. K. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96** 917–932. MR2767279 https://doi.org/10.1093/biomet/asp041

[22] KING, G., HONAKER, J., JOSEPH, A. and SCHEVE, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *Am. Polit. Sci. Rev.* **95** 49–69.

[23] LANNING, D. and BERRY, D. (2003). An alternative to PROC MI for large samples (SUGI 28-271). In 28*th Meeting of the SAS Users Group International*, *Seattle*, *WA*. Retrieved from http://www2.sas.com/proceedings/sugi28/271-28.pdf.

[24] ROBINS, J. M. and WANG, N. (2000). Inference for imputation estimators. *Biometrika* **87** 113–124. MR1766832 https://doi.org/10.1093/biomet/87.1.113

[25] ROJAS, F. (2012). mi impute: a stata command review. Retrieved May 8, 2017, from https://orgtheory.wordpress.com/2012/02/17/mi-impute-a-stata-command-review/.

[26] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

[27] RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0899519 https://doi.org/10.1002/9780470316696

[28] SCHAFER, J. L. (1997). *Analysis of Incomplete Multivariate Data*. *Monographs on Statistics and Applied Probability* **72**. CRC Press, London. MR1692799 https://doi.org/10.1201/9781439821862

[29] SOCIAL SCIENCE COMPUTING COOPERATIVE, UNIVERSITY OF WISCONSIN (2012). Speeding up Multiple Imputation in Stata using Parallel Processing. Retrieved May 8, 2017, from https://www.ssc.wisc.edu/sscc/pubs/stata_mi_condor.htm.

[30] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. *Springer Series in Statistics*. Springer, New York. MR2233926

[31] VAN BUUREN, S. (2018). *Flexible Imputation of Missing Data*, 2nd ed. CRC Press/CRC, Boca Raton, FL.

[32] VAN BUUREN, S., GROOTHUIS-OUDSHOORN, K., ROBITZSCH, A., VINK, G., DOOVE, L., JOLANI, S., SCHOUTEN, R., GAFFERT, P., MEINFELDER, F. et al. (2018). Mice: Multivariate imputation by chained equations. Comprehensive R Archive Network. https://CRAN.R-project.org/package=mice.

[33] VON HIPPEL, P. T. (2013). The bias and efficiency of incomplete-data estimators in small univariate normal samples. *Sociol. Methods Res.* **42** 531–558. MR3190739 https://doi.org/10.1177/0049124113494582

[34] VON HIPPEL, P. T. (2016). New confidence intervals and bias comparisons show that maximum likelihood can beat multiple imputation in small samples. *Struct. Equ. Model.* **23** 422–437. MR3488832 https://doi.org/10.1080/10705511.2015.1047931

[35] VON HIPPEL, P. T. (2018). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociol. Methods Res.* https://doi.org/10.1177/0049124117747303

[36] VON HIPPEL, P. T. and LYNCH, J. (2013). Efficiency gains from using auxiliary variables in imputation. arXiv preprint, arXiv:1311.5249.

[37] WANG, N. and ROBINS, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85** 935–948. MR1666715 https://doi.org/10.1093/biomet/85.4.935

[38] YANG, S. and KIM, J. K. (2016). Fractional imputation in survey sampling: A comparative review. *Statist. Sci.* **31** 415–432. MR3552742 https://doi.org/10.1214/16-STS569

Statistical Science
2021, Vol. 36, No. 3, 421–442
https://doi.org/10.1214/20-STS799

Random Matrix Theory and Its Applications

Alan Julian Izenman

Abstract. This article reviews the important ideas behind random matrix theory (RMT), which has become a major tool in a variety of disciplines, including mathematical physics, number theory, combinatorics and multivariate statistical analysis. Much of the theory involves ensembles of random matrices that are governed by some probability distribution. Examples include Gaussian ensembles and Wishart–Laguerre ensembles. Interest has centered on studying the spectrum of random matrices, especially the extreme eigenvalues, suitably normalized, for a single Wishart matrix and for two Wishart matrices, for finite and infinite sample sizes in the real and complex cases. The Tracy–Widom Laws for the probability distribution of a normalized largest eigenvalue of a random matrix have become very prominent in RMT. Limiting probability distributions of eigenvalues of a certain random matrix lead to Wigner's Semicircle Law and Marčenko–Pastur's Quarter-Circle Law. Several applications of these results in RMT are described in this article.

Key words and phrases: Eigenvalue density, Gaussian ensembles, Jacobi ensembles, Marčenko–Pastur's quarter-circle law, Spiked covariance model, Tracy–Widom laws, Wigner matrix, Wigner's semicircle law, Wishart matrix, Wishart–Laguerre ensembles.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1970). *Handbook of Mathematical Functions*, *with Formulas*, *Graphs*, *and Mathematical Tables*. Dover Publications, New York.

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) (B. N. Petrov and F. Csáki, eds.) 267–281. Akadémia Kiado, Budapest. MR0483125

AKEMANN, G. BAIK, J. and DI FRANCESCO, P. (2011). *The Oxford Handbook of Random Matrix Theory*. Oxford Univ. Press, London.

ALDOUS, D. and DIACONIS, P. (1999). Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem. *Bull*. *Amer*. *Math*. *Soc*. (*N*.*S*.) **36** 413–432. MR1694204 https://doi.org/10.1090/S0273-0979-99-00796-X

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann*. *Math*. *Stat*. **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0771294

ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

ARNOLD, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J*. *Math*. *Anal*. *Appl*. **20** 262–268. MR0217833 https://doi.org/10.1016/0022-247X(67)90089-3

BACHMAT, E., BEREND, D., SAPIR, L., SKIENA, S. and STOLYAROV, N. (2006). Letter to the editor: Analysis of aeroplane boarding via spacetime geometry and random matrix theory. *J*. *Phys*. *A*

BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist*. *Sinica* **9** 611–677. MR1711663

BAI, Z., CHOI, K. P. and FUJIKOSHI, Y. (2018a). Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis. *Ann*. *Statist*. **46** 1050–1076. MR3797996 https://doi.org/10.1214/17-AOS1577

BAI, Z., CHOI, K. P. and FUJIKOSHI, Y. (2018b). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT. *Ann*. *Statist*. **46** 2985–3013. MR3851762 https://doi.org/10.1214/17-AOS1646

BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl*. **1** 1150011. MR2934717 https://doi.org/10.1142/S2010326311500110

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J*. *Multivariate Anal*. **26** 166–168. MR0963829 https://doi.org/10.1016/0047-259X(88)90078-4

BAI, Z. D. and YIN, Y. Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann*. *Probab*. **16** 1729–1741. MR0958213

Alan J. Izenman is Professor, Department of Statistical Science, Temple University, 1810 Liacouras Walk, Philadelphia, Pennsylvania 19122, USA (e-mail: alan@temple.edu).

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

BAIK, J. (2005). Limiting distribution of last passage percolation models. In *XIVth International Congress on Mathematical Physics* 339–346. World Sci. Publ., Hackensack, NJ. MR2227847

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

BAIK, J., DEIFT, P. and JOHANSSON, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** 1119–1178. MR1682248 https://doi.org/10.1090/S0894-0347-99-00307-0

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

BAIK, J., BORODIN, A., DEIFT, P. and SUIDAN, T. (2006). A model for the bus system in Cuernavaca (Mexico). *J. Phys. A* **39** 8965–8975. MR2240467 https://doi.org/10.1088/0305-4470/39/28/S11

BEENAKKER, C. W. J. (1997). Random-matrix theory of quantum transport. *Rev. Modern Phys.* **69** 731–816.

BEJAN, A. I. (2005). Largest eigenvalues and sample covariance matrices. Tracy–Widom and Painleve II: Computational aspects and realization in S-Plus with applications. M.Sc. dissertation, Univ. Warwick.

BELLMAN, R. (1960). *Introduction to Matrix Analysis*. McGraw-Hill, Inc., New York. MR0122820

BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803 https://doi.org/10.1214/12-AOS1014

BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

BOHIGAS, O., GIANNONI, M.-J. and SCHMIT, C. (1984). Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Phys. Rev. Lett.* **52** 1–4. MR0730191 https://doi.org/10.1103/PhysRevLett.52.1

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart & Winston, New York. MR0443257

BURSTEIN, A. and LANKHAM, I. (2005). Combinatorics of patience sorting piles. *Sém. Lothar. Combin.* **54A** Art. B54Ab, 19. MR2223026

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

CONSTANTINE, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Ann. Math. Stat.* **34** 1270–1285. MR0181056 https://doi.org/10.1214/aoms/1177703863

DEIFT, P. (2007). Universality for mathematical and physical systems. In *International Congress of Mathematicians. Vol. I* 125–152. Eur. Math. Soc., Zürich. MR2334189 https://doi.org/10.4171/022-1/7

DELVAUX, S. and KUIJLAARS, A. B. J. (2010). A graph-based equilibrium problem for the limiting distribution of nonintersecting Brownian motions at low temperature. *Constr. Approx.* **32** 467–512. MR2726442 https://doi.org/10.1007/s00365-010-9106-7

DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. *IEEE International Symposium on Information Theory* 2197–2201.

DIACONIS, P. (2003). Patterns in eigenvalues: The 70th Josiah Willard Gibbs lecture. *Bull. Amer. Math. Soc.* (*N.S.*) **40** 155–178. MR1962294 https://doi.org/10.1090/S0273-0979-03-00975-3

DOBRIBAN, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. MR3670197 https://doi.org/10.1214/16-AOS1514

DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601

DYSON, F. J. (1962). The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics. *J. Math. Phys.* **3** 1199–1215. MR0177643 https://doi.org/10.1063/1.1703863

EDELMAN, A. and RAO, N. R. (2005). Random matrix theory. *Acta Numer.* **14** 233–297. MR2168344 https://doi.org/10.1017/S0962492904000236

EL KAROUI, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. Technical report, Dept. Statistics, Stanford Univ. Available at arXiv:math/0309355.

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581

ERDËSH, L. (2011). Universality of Wigner random matrices: A survey of recent results. *Uspekhi Mat. Nauk* **66** 67–198. MR2859190 https://doi.org/10.1070/RM2011v066n03ABEH004749

ERDŐS, L. and YAU, H.-T. (2017). *A Dynamical Approach to Random Matrix Theory*. *Courant Lecture Notes in Mathematics* **28**. Amer. Math. Soc., Providence, RI. MR3699468

ERDŐS, L., PÉCHÉ, S., RAMÍREZ, J. A., SCHLEIN, B. and YAU, H.-T. (2010). Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.* **63** 895–925. MR2662426 https://doi.org/10.1002/cpa.20317

FAN, Z. and JOHNSTONE, I. M. (2019). Eigenvalue distributions of variance components estimators in high-dimensional random effects models. *Ann. Statist.* **47** 2855–2886. MR3988775 https://doi.org/10.1214/18-AOS1767

FORRESTER, P. J., SNAITH, N. C. and VERBAARSCHOT, J. J. M. (2003). Developments in random matrix theory. *Journal of Physics A, General Physics, Special Edition on Random Matrix Theory*, **36**, 22pp.

GEMAN, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592

GRENANDER, U. (1963). *Probabilities on Algebraic Structures*. Wiley, New York. MR0206994

GUIONNET, A. (2008). *On Random Matrices*. *Lecture Notes in Math*. Springer, New York.

HAMMERSLEY, J. M. (1972). A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. I*: *Theory of Statistics* (L. M. LeCam, J. Neymanand E. L. Scott, eds.) 345–394. Univ. California Press, Berkeley, CA. MR0405665

HAN, X., PAN, G. and ZHANG, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933 https://doi.org/10.1214/15-AOS1427

IZENMAN, A. J. (2013). *Modern Multivariate Statistical Techniques*: *Regression*, *Classification*, *and Manifold Learning*. *Springer Texts in Statistics*. Springer, New York. https://doi.org/10.1007/978-0-387-78189-1_1.

JAMES, A. T. (1960). The distribution of the latent roots of the covariance matrix. *Ann. Math. Stat.* **31** 151–158. MR0126901 https://doi.org/10.1214/aoms/1177705994

JAMES, A. T. (1961). Zonal polynomials of the real positive definite symmetric matrices. *Ann. of Math.* (2) **74** 456–469. MR0140741 https://doi.org/10.2307/1970291

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. MR0181057 https://doi.org/10.1214/aoms/1177703550

JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991 https://doi.org/10.1007/s002200050027

JOHANSSON, K. (2002). Non-intersecting paths, random tilings and random matrices. *Probab. Theory Related Fields* **123** 225–280. MR1900323 https://doi.org/10.1007/s004400100187

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195 https://doi.org/10.4171/022-1/13

JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010 https://doi.org/10.1214/08-AOS605

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

JOHNSTONE, I. M. and MA, Z. (2012). Fast approach to the Tracy–Widom law at the edge of GOE and GUE. *Ann. Appl. Probab.* **22** 1962–1988. MR3025686 https://doi.org/10.1214/11-AAP819

JOHNSTONE, I. M., MA, Z., PERRY, P. O. SHAHRAM, M. (2014). RMTstat: Distributions, statistics and tests derived from random matrix theory. Available at cran.r-project.org/web/packages/RMTstat/index.html.

JOHNSTONE, I. M. and ONATSKI, A. (2020). Testing in high-dimensional spiked models. *Ann. Statist.* **48** 1231–1254. MR4124321 https://doi.org/10.1214/18-AOS1697

KARGIN, V. (2015). On estimation in the reduced-rank regression with a large number of responses and predictors. *J. Multivariate Anal.* **140** 377–394. MR3372575 https://doi.org/10.1016/j.jmva.2015.06.004

KHATRI, C. G. (1964). Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations. *Ann. Math. Stat.* **35** 1807–1810. MR0175244 https://doi.org/10.1214/aoms/1177700403

KRBALEK, M. and SEBA, P. (2000). The statistical properties of the city transport in Cuernavaca (Mexico) and random matrix ensembles. *J. Phys. A: Math. Gen.* **33** L229–L234.

KRIECHERBAUER, T., MARKLOF, J. and SOSHNIKOV, A. (2001). Random matrices and quantum chaos. *Proc. Natl. Acad. Sci. USA* **98** 10531–10532. MR1854546 https://doi.org/10.1073/pnas.191366198

KULESZA, A. and TASKAR, B. (2012). Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **5** 123–286.

LIVSHYTS, G. V., TIKHOMIROV, K. and VERSHYNIN, R. (2019). The smallest singular value of inhomogeneous square random matrices. Unpublished manuscript. Available at arXiv:1909.04219.

MALLOWS, C. L. (1973). Patience sorting. *Bulletin of the Institute of Mathematics and Its Applications* **9** 216–224. Originally introduced by Mallows in 1962 as “Problem 62-2. Patience Sorting,” *SIAM Review*, **4**, 148–149.

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72** 507–536. MR0208649

MEHTA, M. L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics* (*Amsterdam*) **142**. Elsevier/Academic Press, Amsterdam. MR2129906

MEZZADRI, F. and SNAITH, N. C., eds. (2005) *Recent Perspectives in Random Matrix Theory and Number Theory*. *London Mathematical Society Lecture Note Series* **322**. Cambridge Univ. Press, Cambridge. MR2145172 https://doi.org/10.1017/CBO9780511550492

MINGO, J. A. and SPEICHER, R. (2017). *Free Probability and Random Matrices*. *Fields Institute Monographs* **35**. Springer, New York. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

MIOLANE, L. (2019). Phase transitions in spiked matrix estimation: Information-theoretic analysis. Unpublished manuscript. Available at arXiv:1806.04343.

MO, Y. (2012). The rank 1 real Wishart spiked model. *Comm. Pure. Appl. Math.* **65** 1528–1638. MR2969495 https://doi.org/10.1002/cpa.21415

MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. MR3472260 https://doi.org/10.1109/TIT.2015.2457942

MONTGOMERY, H. L. (1973). The pair correlation of zeros of the zeta function. In *Analytic Number Theory* (*Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo.*, 1972) 181–193. MR0337821

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

ODLYZKO, A. M. (1987). On the distribution of spacings between zeros of the zeta function. *Math. Comp.* **48** 273–308. MR0866115 https://doi.org/10.2307/2007890

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

PAUL, D. (2004). Asymptotics of the leading sample eigenvalues for a spiked covariance model. Technical report, Stanford Univ. Available at anson.ucdavis.edu/~debashis/techrep/eigenlimit.pdf.

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

PÉCHÉ, S. (2003). Universality of local eigenvalue statistics for random sample covariance matrices, Thèse No. 2716. École Polytechnique Fédérale de Lausanne.

PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670 https://doi.org/10.1007/s00440-007-0133-7

PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625

PORTER, C. E. (1965). *Statistical Theory of Spectra: Fluctuations*. Academic Press, New York.

PORTER, C. E. and ROSENZWEIG, N. (1960). Statistical properties of atomic and nuclear spectra. *Ann. Acad. Sci. Fenn. Ser. A VI Phys.* **44** 66. MR0118377

RAMÍREZ, J. A., RIDER, B. and VIRÁG, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.* **24** 919–944. MR2813333 https://doi.org/10.1090/S0894-0347-2011-00703-0

RAO, J. R. and EDELMAN, A. (2007). The polynomial method for random matrices. *Found. Comput. Math.* Available at web.eecs.umich.edu/~rajnrao/rmtool/.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

ŠEBA, P. (2009). Parking and the visual perception of space. *J. Stat. Mech. Theory Exp.* L10002.

SILVERSTEIN, J. W. (1985). The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.* **13** 1364–1368. MR0806232

SINAĬ, Y. G. and SOSHNIKOV, A. B. (1998). A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *Funktsional. Anal. i Prilozhen.* **32** 56–79, 96. MR1647832 https://doi.org/10.1007/BF02482597

SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** 697–733. MR1727234 https://doi.org/10.1007/s002200050743

SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056.

TAO, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. MR2906465 https://doi.org/10.1090/gsm/132

TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449 https://doi.org/10.1007/s00220-010-1044-5

TAO, T. and VU, V. (2011a). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665 https://doi.org/10.1007/s11511-011-0061-3

TAO, T. and VU, V. (2011b). The Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices. *Electron. J. Probab.* **16** 2104–2121. MR2851058 https://doi.org/10.1214/EJP.v16-944

TAO, T. and VU, V. (2012). Random covariance matrices: Universality of local statistics of eigenvalues. *Ann. Probab.* **40** 1285–1315. MR2962092 https://doi.org/10.1214/11-AOP648

TIKHOMIROV, K. (2015). The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math.* **284** 1–20. MR3391069 https://doi.org/10.1016/j.aim.2015.07.020

TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

TRACY, C. A. and WIDOM, H. (2000). Universality of the distribution functions of random matrix theory. In *Integrable Systems: From Classical to Quantum* (*Montréal, QC*, 1999). *CRM Proc. Lecture Notes* **26** 251–264. Amer. Math. Soc., Providence, RI. MR1791893

TULINO, A. M. and VERDÚ, S. (2004). Random matrix theory and wireless communications. *Found. Trends Commun. Inf. Theory* **1** 1–182.

ULAM, S. M. (1961). Monte Carlo calculations in problems of mathematical physics. In *Modern Mathematics for the Engineer: Second Series* (E. F. Beckenbach, eds.) 261–281. McGraw-Hill, New York. MR0129165

VERŠIK, A. M. and KEROV, S. V. (1977). Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Dokl. Akad. Nauk SSSR* **233** 1024–1027. MR0480398

VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

VOICULESCU, D. (1995). Free probability theory: Random matrices and von Neumann algebras. In *Proceedings of the International Congress of Mathematicians, Vol.* 1, 2 (*Zürich*, 1994) 227–241. Birkhäuser, Basel. MR1403924

VON NEUMANN, J. (1937). Some matrix inequalities and metrization of matric space. *Tomsk University Review* **1** 286–300.

WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18. MR0467894 https://doi.org/10.1214/aop/1176995607

WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

WANG, Q. and YAO, J. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.* **45** 415–460. MR3611497 https://doi.org/10.1214/16-AOS1463

WIGNER, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* (2) **62** 548–564. MR0077805 https://doi.org/10.2307/1970079

WIGNER, P. (1957). Random matrices in physics. *SIAM Rev.* **9** 1–23.

WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2) **67** 325–327. MR0095527 https://doi.org/10.2307/1970008

WISHART, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344 https://doi.org/10.1007/BF00353874

Statistical Science
2021, Vol. 36, No. 3, 443–464
https://doi.org/10.1214/20-STS802

The GENIUS Approach to Robust Mendelian Randomization Inference

Eric Tchetgen Tchetgen, BaoLuo Sun and Stefan Walter

Abstract. Mendelian randomization (MR) is a popular instrumental variable (IV) approach, in which one or several genetic markers serve as IVs that can sometimes be leveraged to recover valid inferences about a given exposure-outcome causal association subject to unmeasured confounding. A key IV identification condition known as the exclusion restriction states that the IV cannot have a direct effect on the outcome which is not mediated by the exposure in view. In MR studies, such an assumption requires an unrealistic level of prior knowledge about the mechanism by which genetic markers causally affect the outcome. As a result, possible violation of the exclusion restriction can seldom be ruled out in practice. To address this concern, we introduce a new class of IV estimators which are robust to violation of the exclusion restriction under data generating mechanisms commonly assumed in MR literature. The proposed approach named “MR G-Estimation under No Interaction with Unmeasured Selection” (MR GENIUS) improves on Robins’ G-estimation by making it robust to both additive unmeasured confounding and violation of the exclusion restriction assumption. In certain key settings, MR GENIUS reduces to the estimator of Lewbel (*J. Bus. Econom. Statist.* **30** (2012) 67–80) which is widely used in econometrics but appears largely unappreciated in MR literature. More generally, MR GENIUS generalizes Lewbel’s estimator to several key practical MR settings, including multiplicative causal models for binary outcome, multiplicative and odds ratio exposure models, case control study design and censored survival outcomes.

Key words and phrases: Additive model, confounding, exclusion restriction, G-estimation, instrumental variable, robustness.

REFERENCES

ALBERT, J. M. (2008). Mediation analysis *via* potential outcomes models. *Stat. Med.* **27** 1282–1304. MR2420158 https://doi.org/10.1002/sim.3016

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44** 512–525.

BOWDEN, J. and VANSTEELANDT, S. (2011). Mendelian randomization analysis of case-control data using structural mean models. *Stat. Med.* **30** 678–694. MR2767465 https://doi.org/10.1002/sim.4138

BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** 304–314.

BREUSCH, T. S. and PAGAN, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47** 1287–1294. MR0545960 https://doi.org/10.2307/1911963

CHAO, J. C. and SWANSON, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica* **73** 1673–1692. MR2156676 https://doi.org/10.1111/j.1468-0262.2005.00632.x

CONLEY, T. G., HANSEN, C. B. and ROSSI, P. E. (2012). Plausibly exogenous. *Rev. Econ. Stat.* **94** 260–272.

DAVEY SMITH, G. and EBRAHIM, S. (2003). ‘Mendelian randomization’: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32** 1–22.

Eric Tchetgen Tchetgen is Professor, Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: ett@wharton.upenn.edu). BaoLuo Sun is Assistant Professor, Department of Statistics and Applied Probability, National University of Singapore, Singapore, Republic of Singapore (e-mail: stasb@nus.edu.sg). Stefan Walter is Principal Investigator, Department of Medicine and Public Health, Rey Juan Carlos University, Madrid, Spain (e-mail: stefan.walter@urjc.es).

DAVEY SMITH, G. and EBRAHIM, S. (2004). Mendelian randomization: Prospects, potentials, and limitations. *Int. J. Epidemiol.* **33** 30–42.

DIDELEZ, V., MENG, S. and SHEEHAN, N. A. (2010). Assumptions of IV methods for observational epidemiology. *Statist. Sci.* **25** 22–40. MR2741813 https://doi.org/10.1214/09-STS316

EMSLEY, R., DUNN, G. and WHITE, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat. Methods Med. Res.* **19** 237–270. MR2757115 https://doi.org/10.1177/0962280209105014

FLORES, C. A. and FLORES-LAGUNES, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *J. Bus. Econom. Statist.* **31** 534–545. MR3173699 https://doi.org/10.1080/07350015.2013.822760

GUO, Z., KANG, H., CAI, T. T. and SMALL, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 793–815. MR3849344 https://doi.org/10.1111/rssb.12275

HAN, C. (2008). Detecting invalid instruments using L_1-GMM. *Econom. Lett.* **101** 285–287. MR2477476 https://doi.org/10.1016/j.econlet.2008.09.004

HECKMAN, J. and VYTLACIL, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *J. Hum. Resour.* 974–987.

HIRANO, K., IMBENS, G. W., RUBIN, D. B. and ZHOU, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1** 69–88.

HUBER, M. (2014). Sensitivity checks for the local average treatment effect. *Econom. Lett.* **123** 220–223.

JUSTER, F. T. and SUZMAN, R. (1995). An overview of the health and retirement study. *J. Hum. Resour.* **30** S7–S56.

KANG, H. (2017). sisVIVE: Some Invalid Some Valid Instrumental Variables Estimator. R package version 1.4.

KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J. Amer. Statist. Assoc.* **111** 132–144. MR3494648 https://doi.org/10.1080/01621459.2014.994705

KLEIBER, C. and ZEILEIS, A. (2008). *Applied Econometrics with R*. Springer, New York.

KOLESÁR, M., CHETTY, R., FRIEDMAN, J., GLAESER, E. and IMBENS, G. W. (2015). Identification and inference with many invalid instruments. *J. Bus. Econom. Statist.* **33** 474–484. MR3416595 https://doi.org/10.1080/07350015.2014.978175

LAWLOR, D. A., HARBORD, R. M., STERNE, J. A. C., TIMPSON, N. and SMITH, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27** 1133–1163. MR2420151 https://doi.org/10.1002/sim.3034

LEEB, H. and PÖTSCHER, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. *J. Econometrics* **142** 201–211. MR2394290 https://doi.org/10.1016/j.jeconom.2007.05.017

LEWBEL, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econom. Statist.* **30** 67–80. MR2899185 https://doi.org/10.1080/07350015.2012.643126

LEWBEL, A. (2018). Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Econom. Lett.* **165** 10–12. MR3768154 https://doi.org/10.1016/j.econlet.2018.01.003

LI, J., FINE, J. and BROOKHART, A. (2015). Instrumental variable additive hazards models. *Biometrics* **71** 122–130. MR3335356 https://doi.org/10.1111/biom.12244

LITTLE, J. and KHOURY, M. J. (2003). Mendelian randomisation: A new spin or real progress? *Lancet* **362** 930–931. https://doi.org/10.1016/S0140-6736(03)14396-6

MA, Z., XIE, X. and GENG, Z. (2006). Collapsibility of distribution dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 127–133. MR2212578 https://doi.org/10.1111/j.1467-9868.2005.00536.x

MAILMAN, M. D., FEOLO, M., JIN, Y., KIMURA, M., TRYKA, K., BAGOUTDINOV, R., HAO, L., KIANG, A., PASCHALL, J. et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39** 1181.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461** 747.

MANSKI, C. F. and PEPPER, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* **68** 997–1010. MR1771587 https://doi.org/10.1111/1468-0262.00144

MARTINUSSEN, T., VANSTEELANDT, S., TCHETGEN TCHETGEN, E. J. and ZUCKER, D. M. (2017). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics* **73** 1140–1149. MR3744528 https://doi.org/10.1111/biom.12699

MEALLI, F. and PACINI, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *J. Amer. Statist. Assoc.* **108** 1120–1131. MR3174688 https://doi.org/10.1080/01621459.2013.802238

MORRIS, A. P., VOIGHT, B. F., TESLOVICH, T. M., FERREIRA, T., SEGRE, A. V., STEINTHORSDOTTIR, V., STRAWBRIDGE, R. J., KHAN, H., GRALLERT, H. et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44** 981.

NEVO, A. and ROSEN, A. M. (2012). Identification with imperfect instruments. *Rev. Econ. Stat.* **94** 659–671.

NEWEY, W. K. and WINDMEIJER, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica* **77** 687–719. MR2531359 https://doi.org/10.3982/ECTA6224

NIE, H., CHENG, J. and SMALL, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* **67** 1397–1405. MR2872390 https://doi.org/10.1111/j.1541-0420.2011.01575.x

OFSTEDAL, M. B., FISHER, G. G. and HERZOG, A. R. (2005). Documentation of cognitive functioning measures in the health and retirement study.

ROBINS, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: *Health Service Research Methodology: A Focus on AIDS* 113–159.

ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185 https://doi.org/10.1080/03610929408831393

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality (Los Angeles, CA, 1994). Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

SHARDELL, M. and FERRUCCI, L. (2016). Instrumental variable analysis of multiplicative models with potentially invalid instruments. *Stat. Med.* **35** 5430–5447. MR3573068 https://doi.org/10.1002/sim.7069

SMALL, D. S. (2012). Mediation analysis without sequential ignorability: Using baseline covariates interacted with random assignment as instrumental variables. *J. Statist. Res.* **46** 91–103. MR3114108

SPILLER, W., SLICHTER, D., BOWDEN, J. and DAVEY SMITH, G. (2018). Detecting and correcting for bias in Mendelian randomization analyses using gene-by-environment interactions. https://doi.org/10.1093/ije/dyy204

STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65** 557–586. MR1445622 https://doi.org/10.2307/2171753

STOCK, J. H. and WRIGHT, J. H. (2000). GMM with weak identification. *Econometrica* **68** 1055–1096. MR1779144 https://doi.org/10.1111/1468-0262.00151

STOCK, J. H., WRIGHT, J. H. and YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *J. Bus. Econom. Statist.* **20** 518–529. MR1973801 https://doi.org/10.1198/073500102288618658

TCHETGEN TCHETGEN, E. J., ROBINS, J. M. and ROTNITZKY, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika* **97** 171–180. MR2594425 https://doi.org/10.1093/biomet/asp062

TCHETGEN TCHETGEN, E., SUN, B. and WALTER, S. (2021). Supplement to "The GENIUS Approach to Robust Mendelian Randomization Inference." https://doi.org/10.1214/20-STS802SUPP

TCHETGEN TCHETGEN, E. J., WALTER, S., VANSTEELANDT, S., MARTINUSSEN, T. and GLYMOUR, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26** 402.

TEN HAVE, T. R., JOFFE, M. M., LYNCH, K. G., BROWN, G. K., MAISTO, S. A. and BECK, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics* **63** 926–934. MR2395813 https://doi.org/10.1111/j.1541-0420.2007.00766.x

VANSTEELANDT, S. and DIDELEZ, V. (2018). Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. *Scand. J. Stat.* **45** 941–961. MR3884895 https://doi.org/10.1111/sjos.12329

WANG, L. and TCHETGEN TCHETGEN, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 531–550. MR3798877 https://doi.org/10.1111/rssb.12262

WHITE, K., MONDESIR, F. L., BATES, L. M. and GLYMOUR, M. M. (2014). Diabetes risk, diagnosis, and control: Do psychosocial factors predict hemoglobin A1C defined outcomes or accuracy of self-reports? *Ethn. Dis.* **21** 19–27.

WINDMEIJER, F., FARBMACHER, H., DAVIES, N. and SMITH, G. D. (2019). On the use of the Lasso for instrumental variables estimation with some invalid instruments. *J. Amer. Statist. Assoc.* **114** 1339–1350. MR4011783 https://doi.org/10.1080/01621459.2018.1498346

WOOLDRIDGE, J. M. (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Econom. Lett.* **79** 185–191. MR1968966 https://doi.org/10.1016/S0165-1765(02)00318-X

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

WU, Q., TCHETGEN TCHETGEN, E., OSYPUK, T., WHITE, K., MUJAHID, M. and GLYMOUR, M. (2012). Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Dis. Assoc. Disord.* **27** 207–212.

YAVORSKA, O. and STALEY, J. (2019). MendelianRandomization: Mendelian Randomization Package. R package version 0.4.1.

ZHENG, C. and ZHOU, X.-H. (2015). Causal mediation analysis in the multilevel intervention and multicomponent mediator case. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 581–615. MR3351447 https://doi.org/10.1111/rssb.12082

Statistical Science
2021, Vol. 36, No. 3, 465–492
https://doi.org/10.1214/20-STS803

A General Framework for the Analysis of Adaptive Experiments

Ian C. Marschner

Abstract. Adaptive experiments have design features that adapt to the accumulating data and are therefore informative about the parameter of interest. As a consequence, the overall information in an adaptive experiment is a combination of information from two sources, the realized design and the observed outcomes. This paper presents a general framework for the analysis of adaptive experiments, based on the decomposition of overall information into design information and outcome information. Likelihood inference is discussed, beginning with assumptions that guarantee insensitivity of the likelihood to the adaptive design. We then focus on the relative merits of unconditional and conditional inference. Although conditional inference is inefficient due to the nonancillary design, unconditional inference may be biased conditional on the realized design. Identifying such conditional bias in a given experiment is a motivation of the proposed framework. We show that conditional bias stems from correlation between the total information and the design information, and that this bias is most pronounced in samples where the design information is inconsistent with the outcome information. Thus, by viewing the unconditional likelihood as the aggregation of information from a design likelihood and a conditional likelihood, we can use meta-analysis principles to assess heterogeneity between the two information sources. When such heterogeneity is detected, conditional inference may be more appropriate. Interpretation from a Bayesian perspective is also discussed.

Key words and phrases: Adaptive design, Bayesian inference, conditional inference, estimation bias, randomized experiment, sequential experiment.

REFERENCES

BAKSHY, E., DWORKIN, L., KARRER, B., KASHIN, K., LETHAM, B., MURTHY, A. and SINGH, S. (2018). AE: A domain-agnostic platform for adaptive experimentation. In *32nd Conference on Neural Processing Systems* (*NIPS* 2018), Montreal, Canada.

BASSLER, D., BRIEL, M., MONTORI, V. M., LANE, M., GLASZIOU, P., HEELS-ANSDELL, D., WALTER, S. D. and GUYATT, G. H. (2010). Stopping randomized trials early for benefit and estimation of treatment effects: Systematic review and meta-regression analysis. *J. Am. Med. Assoc.* **303** 1180–1187.

BASSLER, D., MONTORI, V. M., BRIEL, M., GLASZIOU, P., WALTER, S. D., RAMSAY, T. and GUYATT, G. (2013). Reflections on meta-analyses involving trials stopped early for benefit: Is there a problem and if so, what is it? *Stat. Methods Med. Res.* **22** 159–168. MR3190652 https://doi.org/10.1177/0962280211432211

BERRY, D. A. (1987). Interim analysis in clinical trials. The role of the likelihood principle. *Amer. Statist.* **41** 117–122. MR0898268 https://doi.org/10.2307/2684222

BERRY, S. M., CARLIN, B. P., LEE, J. J. and MÜLLER, P. (2010). *Bayesian Adaptive Methods for Clinical Trials. Chapman & Hall/CRC Biostatistics Series* **38**. CRC Press, Boca Raton, FL. MR2723582

BOTHWELL, L. E., AVORN, J., KHAN, N. F. and KESSELHEIM, A. S. (2018). Adaptive design clinical trials: A review of the literature and ClinicalTrials.gov. *BMJ Open* **8** e018320. https://doi.org/10.1136/bmjopen-2017-018320

BOWDEN, J. and GLIMM, E. (2014). Conditionally unbiased and near unbiased estimation of the selected treatment mean for multistage drop-the-losers trials. *Biom. J.* **56** 332–349. MR3249685 https://doi.org/10.1002/bimj.201200245

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

BRANNATH, W., MEHTA, C. R. and POSCH, M. (2009). Exact confidence bounds following adaptive group sequential tests. *Biometrics* **65** 539–546. MR2751478 https://doi.org/10.1111/j.1541-0420.2008.01101.x

CAVAGNARO, D. R., MYUNG, J. I., PITT, M. A. and KUJALA, J. V.

Ian C. Marschner is Professor of Biostatistics, NHMRC Clinical Trials Centre, The University of Sydney, Sydney, New South Wales, Australia (e-mail: ian.marschner@ctc.usyd.edu.au).

(2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Comput*. **22** 887–905. MR2654272 https://doi.org/10.1162/neco.2009.02-09-959

CHIU, Y.-D., KOENIG, F., POSCH, M. and JAKI, T. (2018). Design and estimation in clinical trials with subpopulation selection. *Stat*. *Med*. **37** 4335–4352. MR3879432 https://doi.org/10.1002/sim.7925

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

DAIMON, T., HIRAKAWA, A. and MATSUI, S. (2019). *Dose-Finding Designs for Early-Phase Cancer Clinical Trials*: *A Brief Guidebook to Theory and Practice*. *JSS Research Series in Statistics*. Springer, Tokyo. MR3932094 https://doi.org/10.1007/978-4-431-55585-8

DEEKS, J. J., ALTMAN, D. G. and BRADBURN, M. J. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In *Systematic Reviews in Health Care*: *Meta-Analysis in Context*, 2nd ed. (M. Egger, G. Davey Smith and D. G. Altman, eds.). BMJ Books, London.

EMERSON, S. S. and FLEMING, T. R. (1990). Parameter estimation following group sequential hypothesis testing. *Biometrika* **77** 875–892. MR1086697 https://doi.org/10.1093/biomet/77.4.875

ENSIGN, L. G., GEHAN, E. A., KAMEN, D. S. and THALL, P. (1994). An optimal three-stage design for phase II clinical trials. *Stat*. *Med*. **13** 1727–1736.

EVANS, M. and MOSHONOV, H. (2006). Checking for prior-data conflict. *Bayesian Anal*. **1** 893–914. MR2282210 https://doi.org/10.1214/06-BA129

FAN, X. and DEMETS, D. L. (2006). Conditional and unconditional confidence intervals following a group sequential test. *J. Biopharm*. *Statist*. **16** 107–122. MR2222027 https://doi.org/10.1080/10543400500406595

FAN, X., DEMETS, D. L. and LAN, K. K. G. (2004). Conditional bias of point estimates following a group sequential test. *J. Biopharm*. *Statist*. **14** 505–530. MR2190497 https://doi.org/10.1081/BIP-120037195

FDA (2018). *Adaptive Designs for Clinical Trials of Drugs and Biologics*: *Guidance for Industry*. Food and Drug Administration, MD, USA.

FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38. MR1225212 https://doi.org/10.1093/biomet/80.1.27

FLOURNOY, N. and ORON, A. (2019). Bias induced by adaptive dose-finding designs. *J. Appl*. *Stat*. https://doi.org/10.1080/02664763.2019.1649375

GAYDOS, B., KRAMS, M., PEREVOZSKAYA, I., BRETZ, F., LIU, Q., GALLO, P., BERRY, D., CHUANG-STEIN, C., PINHEIRO, J. et al. (2006). Adpative dose–response studies. *Drug Inf*. *J*. **40** 451–461.

HERITIER, S., CANTONI, E., COPT, S. and VICTORIA-FESER, M.-P. (2009). *Robust Methods in Biostatistics*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2604994 https://doi.org/10.1002/9780470740538

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

JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press/CRC, Boca Raton, FL. MR1710781

KIMANI, P. K., TODD, S. and STALLARD, N. (2015). Estimation after subpopulation selection in adaptive seamless trials. *Stat*. *Med*. **34** 2581–2601. MR3368404 https://doi.org/10.1002/sim.6506

KRAMAR, A., POTVIN, D. and HILL, C. (1996). Multistage designs for phase II clinical trials: Statistical issues in cancer research. *Br. J. Cancer* **74** 1317–1320.

LIU, A., TROENDLE, J. F., YU, K. F. and YUAN, V. W. (2004). Conditional maximum likelihood estimation following a group sequential test. *Biom*. *J*. **46** 760–768. MR2108617 https://doi.org/10.1002/bimj.200410076

MARSCHNER, I. C. and SCHOU, I. M. (2019). Underestimation of treatment effects in sequentially monitored clinical trials that did not stop early for benefit. *Stat*. *Methods Med*. *Res*. **28** 3027–3041. MR4002682 https://doi.org/10.1177/0962280218795320

MEHTA, C. R. and POCOCK, S. J. (2011). Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Stat*. *Med*. **30** 3267–3284. MR2861612 https://doi.org/10.1002/sim.4102

MEHTA, C. R., BAUER, P., POSCH, M. and BRANNATH, W. (2007). Repeated confidence intervals for adaptive group sequential trials. *Stat*. *Med*. **26** 5422–5433. MR2416837 https://doi.org/10.1002/sim.3062

MISTRY, P., DUNN, J. A. and MARSHALL, A. (2017). A literature review of applied adaptive design methodology within the field of oncology in randomised controlled trials and a proposed extension to the CONSORT guidelines. *BMC Med*. *Res*. *Methodol*. **17** 108.

MOLENBERGHS, G., KENWARD, M. G., AERTS, M., VERBEKE, G., TSIATIS, A. A., DAVIDIAN, M. and RIZOPOULOS, D. (2014). On random sample size, ignorability, ancillarity, completeness, separability, and degeneracy: Sequential trials, random sample sizes, and missing data. *Stat*. *Methods Med*. *Res*. **23** 11–41. MR3190685 https://doi.org/10.1177/0962280212445801

MONTORI, V. M., DEVEREAUX, P. J., ADHIKARI, N. K., BURNS, K. E. A., EGGERT, C. H., BRIEL, M., LACCHETTI, C., LEUNG, T. W., DARLING, E. et al. (2005). Randomized trials stopped early for benefit: A systematic review. *J. Am*. *Med*. *Assoc*. **294** 2203–2209.

MYUNG, J. I., CAVAGNARO, D. R. and PITT, M. A. (2013). A tutorial on adaptive design optimization. *J. Math*. *Psych*. **57** 53–67. MR3094807 https://doi.org/10.1016/j.jmp.2013.05.005

OHMAN STRICKLAND, P. A. and CASELLA, G. (2003). Conditional inference following group sequential testing. *Biom*. *J*. **45** 515–526. MR1998132 https://doi.org/10.1002/bimj.200390029

PALLMANN, P., BEDDING, A. W., CHOODARI-OSKOOEI, B., DIMAIRO, M., FLIGHT, L., HAMPSO, L. V., HOLMES, J., MANDER, A. P., ODONDI, L. et al. (2018). Adaptive designs in clinical trials: Why use them, and how to run them. *BMC Med*. **16** 29.

POCOCK, S. J. and SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103–115.

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., 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

ROSNER, G. L. and TSIATIS, A. A. (1988). Exact confidence intervals following a group sequential trial: A comparison of methods. *Biometrika* **75** 723–729.

SCHOU, I. M. and MARSCHNER, I. C. (2013). Meta-analysis of clinical trials with early stopping: An investigation of potential bias. *Stat*. *Med*. **32** 4859–4874. MR3127181 https://doi.org/10.1002/sim.5893

SCOTT, N. W., MCPHERSON, G. C., RAMSAY, C. R. and CAMPBELL, M. K. (1992). The method of minimization for allocation to clinical trials: A review. *Control*. *Clin*. *Trials* **23** 662–674.

SIGNORINI, D. F., LEUNG, O., SIMES, R. J., BELLER, E. and GEBSKI, V. J. (1993). Dynamic balanced randomization for clinical trials. *Stat*. *Med*. **12** 2343–2350.

SIMON, R. S. (1989). Optimal two-stage designs for phase II clinical trials. *Control*. *Clin*. *Trials* **10** 1–10.

SIMON, N. and SIMON, R. (2013). Adaptive enrichment designs for clinical trials. *Biostatistics* **14** 613–625. https://doi.org/10.1093/biostatistics/kxt010

TAVES, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clin*. *Pharmacol*. *Ther.* **15** 443–453.

THALL, P. F. and WATHEN, J. K. (2007). Practical Bayesian adaptive randomization in clinical trials. *Eur. J. Cancer* **43** 859–866.

TSIATIS, A. A., ROSNER, G. L. and MEHTA, C. R. (1984). Exact confidence intervals following a group sequential test. *Biometrics* **40** 797–803. MR0775387 https://doi.org/10.2307/2530924

WALD, A. (1947). *Sequential Analysis*. Wiley, New York. MR0020764

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., MAGIRR, D., LAW, M. and JAKI, T. (2016). Some recommendations for multi-arm multi-stage trials. *Stat*. *Methods Med*. *Res*. **25** 716–727. MR3489662 https://doi.org/10.1177/0962280212465498

WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist*. *Assoc*. **73** 840–843.

WEI, L. J., SMYTHE, R. T., LIN, D. Y. and PARK, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *J. Amer. Statist*. *Assoc*. **85** 156–162. MR1137361

WHITEHEAD, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika* **73** 573–581. MR0897848 https://doi.org/10.1093/biomet/73.3.573

ZELEN, M. (1969). Play the winner rule and the controlled clinical trial. *J. Amer. Statist*. *Assoc*. **64** 131–146. MR0240938

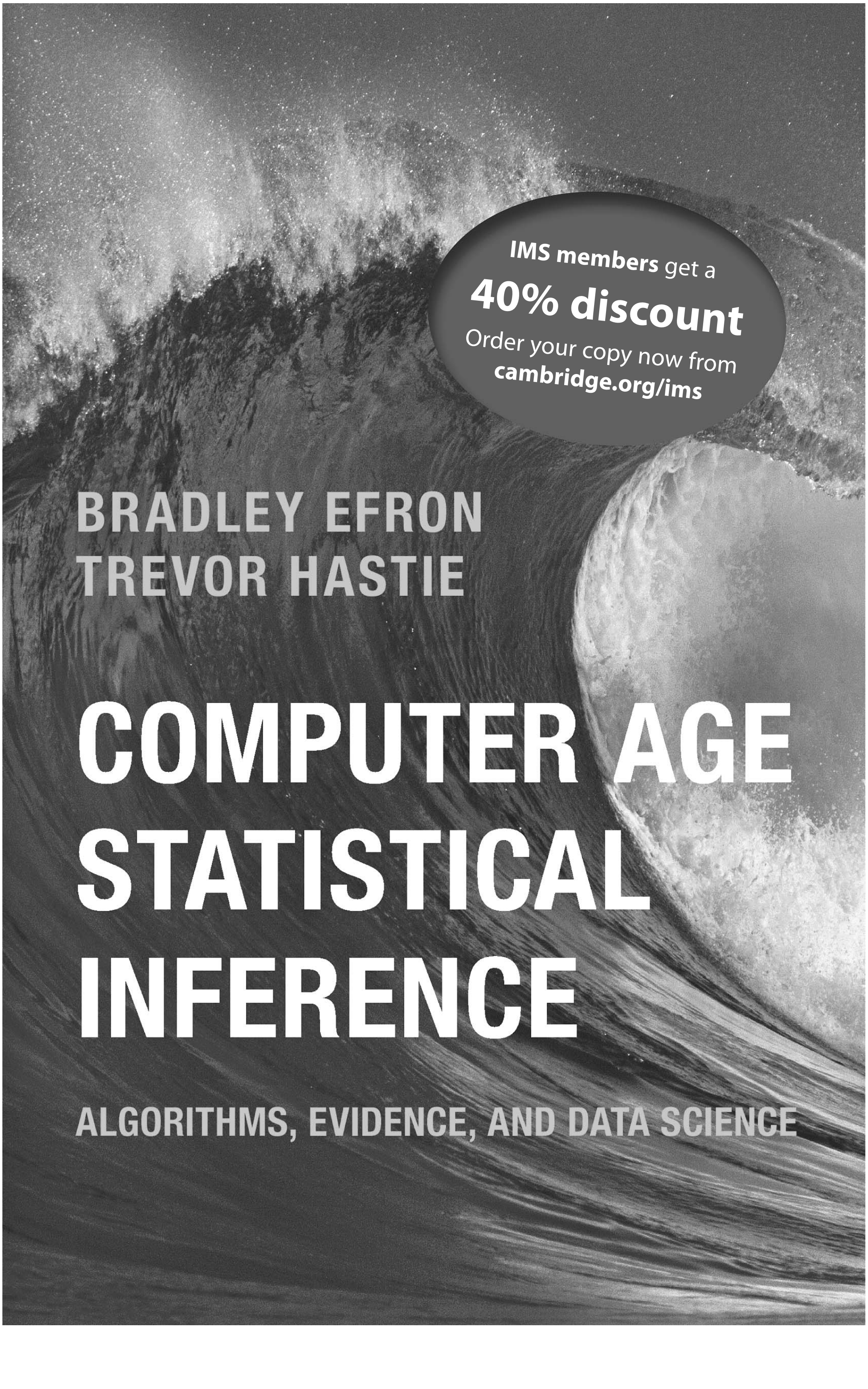
INSTITUTE OF MATHEMATICAL STATISTICS

(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

IMS EDITORS


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