

STATISTICAL SCIENCE

Volume 37, Number 3

August 2022

Gambler's Ruin and the ICM.....	<i>Persi Diaconis and Stewart N. Ethier</i>	289
Confidence Intervals for Seroprevalence	<i>Thomas J. DiCiccio, David M. Ritzwoller, Joseph P. Romano and Azeem M. Shaikh</i>	306
Statistical Aspects of the Quantum Supremacy Demonstration	<i>Yosef Rinott, Tomer Shoham and Gil Kalai</i>	322
Choosing Among Notions of Multivariate Depth Statistics	<i>Karl Mosler and Pavlo Mozharovskyi</i>	348
Power Calculations for Replication Studies.....	<i>Charlotte Micheloud and Leonhard Held</i>	369
Intention-to-Treat Comparisons in Randomized Trials	<i>Ross L. Prentice and Aaron K. Aragaki</i>	380
Modeling the Occurrence of Events Subject to a Reporting Delay via an EM Algorithm	<i>Roel Verbelen, Katrien Antonio, Gerda Claeskens and Jonas Crevecoeur</i>	394
A Unifying Framework of High-Dimensional Sparse Estimation with Difference-of-Convex (DC) Regularizations.....	<i>Shanshan Cao, Xiaoming Huo and Jong-Shi Pang</i>	411
Challenges in Markov Chain Monte Carlo for Bayesian Neural Networks	<i>Theodore Papamarkou, Jacob Hinkle, M. Todd Young and David Womble</i>	425
A Conversation with Stephen Portnoy.....	<i>Xuming He and Xiaofeng Shao</i>	443

Statistical Science [ISSN 0883-4237 (print); ISSN 2168-8745 (online)], Volume 37, Number 3, August 2022. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Copyright © 2022 by the Institute of Mathematical Statistics
Printed in the United States of America

Statistical Science

Volume 37, Number 3 (289–454) August 2022

Volume 37

Number 3

August 2022

Gambler's Ruin and the ICM

Persi Diaconis and Stewart N. Ethier

Confidence Intervals for Seroprevalence

Thomas J. DiCiccio, David M. Ritzwoller, Joseph P. Romano and Azeem M. Shaikh

Statistical Aspects of the Quantum Supremacy Demonstration

Yosef Rinott, Tomer Shoham and Gil Kalai

Choosing Among Notions of Multivariate Depth Statistics

Karl Mosler and Pavlo Mozharovskyi

Power Calculations for Replication Studies

Charlotte Micheloud and Leonhard Held

Intention-to-Treat Comparisons in Randomized Trials

Ross L. Prentice and Aaron K. Aragaki

Modeling the Occurrence of Events Subject to a Reporting Delay via an EM Algorithm

Roel Verbelen, Katrien Antonio, Gerda Claeskens and Jonas Crevecoeur

A Unifying Framework of High-Dimensional Sparse Estimation with Difference-of-Convex (DC) Regularizations

Shanshan Cao, Xiaoming Huo and Jong-Shi Pang

Challenges in Markov Chain Monte Carlo for Bayesian Neural Networks

Theodore Papamarkou, Jacob Hinkle, M. Todd Young and David Womble

A Conversation with Stephen Portnoy

Xuming He and Xiaofeng Shao

EDITOR

Sonia Petrone
Università Bocconi

ASSOCIATE EDITORS

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

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

Bodhisattva Sen
Columbia University
Glenn Shafer
*Rutgers Business
School–Newark and
New Brunswick
Royal Holloway College,
University of London*
Yuekai Sun
University of Michigan
Lorenzo Trippa
Harvard University
Jon Wellner
University of Washington
Bin Yu
*University of California,
Berkeley*
Giacomo Zanella
Università Bocconi
Cun-Hui Zhang
Rutgers University

MANAGING EDITOR

Robert Keener
University of Michigan

PRODUCTION EDITOR

Patrick Kelly

EDITORIAL COORDINATOR

Kristina Mattson

PAST EXECUTIVE EDITORS

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

Gambler's Ruin and the ICM

Persi Diaconis and Stewart N. Ethier

“In the case where there are three players with limited fortunes, the various problems appear to be of quite a different order of difficulty than in the case of two players.”

Louis Bachelier (1912)

Abstract. Consider gambler's ruin with three players, 1, 2, and 3, having initial capitals A , B , and C units. At each round a pair of players is chosen (uniformly at random) and a fair coin flip is made resulting in the transfer of one unit between these two players. Eventually, one of the players is eliminated and play continues with the remaining two. Let $\sigma \in S_3$ be the elimination order (e.g., $\sigma = 132$ means player 1 is eliminated first and player 3 is eliminated second, leaving player 2 with $A + B + C$ units).

We seek approximations (and exact formulas) for the elimination order probabilities $P_{A,B,C}(\sigma)$. Exact, as well as arbitrarily precise, computation of these probabilities is possible when $N := A + B + C$ is not too large. Linear interpolation can then give reasonable approximations for large N . One frequently used approximation, the independent chip model (ICM), is shown to be inadequate. A regression adjustment is proposed, which seems to give good approximations to the elimination order probabilities.

Key words and phrases: Gambler's ruin problem, tower problem, linear interpolation, independent chip model (ICM), Plackett–Luce model, linear regression.

REFERENCES

- AGUILAR, J. (2016). How to negotiate final table deals like a pro. Upswing Poker. Available at <https://upswingpoker.com/final-table-deal-making-tournaments/>.
- ALDOUS, D., LANOUE, D. and SALEZ, J. (2015). The compulsive gambler process. *Electron. J. Probab.* **20** no. 35, 18 pp. MR3335826 <https://doi.org/10.1214/EJP.v20-3582>
- BACHELIER, L. (1912). *Calcul des Probabilités*, Vol. 1. Gauthier-Villars, Paris. MR1191904
- BRUSS, F. T., LOUCHARD, G. and TURNER, J. W. (2003). On the N -tower problem and related problems. *Adv. in Appl. Probab.* **35** 278–294. MR1975514 <https://doi.org/10.1239/aap/1046366109>
- COVER, T. M. (1987). Gambler's ruin: A random walk on the simplex. In *Open Problems in Communication and Computation* (T. M. Cover and B. Gopinath, eds.) 155–155. Springer, New York.
- DAVID, G. (2015). Markov chain solution to the 3-tower problem. In *Information and Communication Technology* (I. Khalil, E. Neuhold, A. Tjoa, L. Xu and I. You, eds.). *ICT-EurAsia* 2015. *Lecture Notes in Computer Science* **9357** 121–128. Springer, Cham.
- DAVIS, T. A. (2006). *Direct Methods for Sparse Linear Systems. Fundamentals of Algorithms* **2**. SIAM, Philadelphia, PA. MR2270673 <https://doi.org/10.1137/1.9780898718881>
- DENISOV, D. and WACHTEL, V. (2015). Random walks in cones. *Ann. Probab.* **43** 992–1044. MR3342657 <https://doi.org/10.1214/13-AOP867>
- DIACONIS, P. (1988). *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. IMS, Hayward, CA. MR0964069
- DIACONIS, P. and FREEDMAN, D. (1979). On rounding percentages. *J. Amer. Statist. Assoc.* **74** 359–364. MR0548026
- DIACONIS, P., HOUSTON-EDWARDS, K. and SALOFF-COSTE, L. (2021). Gambler's ruin estimates on finite inner uniform domains. *Ann. Appl. Probab.* **31** 865–895. MR4254498 <https://doi.org/10.1214/20-aap1607>
- ENGEL, A. (1993). The computer solves the three tower problem. *Amer. Math. Monthly* **100** 62–64.
- ETHIER, S. N. (2010). *The Doctrine of Chances: Probabilistic Aspects of Gambling. Probability and Its Applications (New York)*. Springer, Berlin. MR2656351 <https://doi.org/10.1007/978-3-540-78783-9>
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. I*, 3rd ed. Wiley, New York. MR0228020
- FERGUSON, T. (1995). Gambler's ruin in three dimensions. Unpublished. Available at <https://www.math.ucla.edu/~tom/papers/unpublished/gamblersruin.pdf>.

- GANZFRIED, S. and SANDHOLM, T. (2008). Computing an approximate jam/fold equilibrium for 3-player no-limit Texas hold'em tournaments. In *AAMAS08: 7th International Conference on Autonomous Agents and Multi Agent Systems* (L. Padgham, D. Parkes, J. Müller and S. Parsons, eds.) 919–926. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.
- GILBERT, G. T. (2009). The independent chip model and risk aversion. Preprint. Available at [arXiv:0911.3100](https://arxiv.org/abs/0911.3100).
- GILLILAND, D., LEVENTAL, S. and XIAO, Y. (2007). A note on absorption probabilities in one-dimensional random walk via complex-valued martingales. *Statist. Probab. Lett.* **77** 1098–1105. [MR2395066 https://doi.org/10.1016/j.spl.2007.01.016](https://doi.org/10.1016/j.spl.2007.01.016)
- HAJEK, B. (1987). Gambler's ruin: A random walk on the simplex. In *Open Problems in Communication and Computation* (T. M. Cover and B. Gopinath, eds.) 204–207. Springer, New York.
- HARVILLE, D. A. (1973). Assigning probabilities to the outcomes of multi-entry competitions. *J. Amer. Statist. Assoc.* **68** 312–316.
- ICMIZER (2020). Poker ICM calculator for final table deals. Available at <https://www.icmpoker.com/icmcalculator/>.
- KEMENY, J. G. and SNELL, J. L. (1976). *Finite Markov Chains. Undergraduate Texts in Mathematics*. Springer, New York. [MR0410929](https://doi.org/10.1007/978-1-4939-9947-0)
- KIM, M. S. (2005). Gambler's ruin in many dimensions and optimal strategy in repeated multi-player games with application to poker. Masters thesis, UCLA.
- LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York. [MR0108411](https://doi.org/10.1002/9781118133233)
- LUCE, R. D. (1977). The choice axiom after twenty years. *J. Math. Psych.* **15** 215–233. [MR0462675 https://doi.org/10.1016/0022-2496\(77\)90032-3](https://doi.org/10.1016/0022-2496(77)90032-3)
- MALMUTH, M. (1987). *Gambling Theory and Other Topics*. Two Plus Two Publishing, Henderson, NV.
- MALMUTH, M. (2004). *Gambling Theory and Other Topics*, 6th ed. Two Plus Two Publishing, Henderson, NV.
- MARFIL, R. I. D. and DAVID, G. (2020). On the placing probabilities for the Four-Tower problem using recursions based on multigraphs. *Matimyas Mat.* **43** 19–32. [MR4159463](https://doi.org/10.1016/j.mat.2020.01.001)
- PLACKETT, R. L. (1975). The analysis of permutations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **24** 193–202. [MR0391338 https://doi.org/10.2307/2346567](https://doi.org/10.2307/2346567)
- RATZKIN, J. and TREIBERGS, A. (2009). A capture problem in Brownian motion and eigenvalues of spherical domains. *Trans. Amer. Math. Soc.* **361** 391–405. [MR2439411 https://doi.org/10.1090/S0002-9947-08-04505-4](https://doi.org/10.1090/S0002-9947-08-04505-4)
- ROSS, S. M. (2009). A simple solution to a multiple player gambler's ruin problem. *Amer. Math. Monthly* **116** 77–81. [MR2478757 https://doi.org/10.4169/193009709X469841](https://doi.org/10.4169/193009709X469841)
- SONG, S. and SONG, J. (2013). A note on the history of the gambler's ruin problem. *Commun. Stat. Appl. Methods* **20** 1–12.
- STERN, H. S. (2008). Estimating the probabilities of the outcomes of a horse race (alternatives to the Harville formulas). In *Efficiency of Racetrack Betting Markets* (D. B. Hausch, V. S. Y. Lo and W. T. Ziemba, eds.) 225–235. World Scientific, Singapore.
- STIRZAKER, D. (1994). Tower problems and martingales. *Math. Sci.* **19** 52–59. [MR1294785](https://doi.org/10.1080/00255718.1994.10556285)
- STIRZAKER, D. (2006). Three-handed gambler's ruin. *Adv. in Appl. Probab.* **38** 284–286. [MR2213975 https://doi.org/10.1239/aap/1143936151](https://doi.org/10.1239/aap/1143936151)
- SWAN, Y. C. and BRUSS, F. T. (2006). A matrix-analytic approach to the N -player ruin problem. *J. Appl. Probab.* **43** 755–766. [MR2274798 https://doi.org/10.1239/jap/1158784944](https://doi.org/10.1239/jap/1158784944)
- TURNER, H., VAN ETTEN, J., FIRTH, D. and KOSMIDIS, I. (2017). Introduction to PlackettLuce. Microsoft R Application Network. Available at <https://mran.microsoft.com/snapshot/2017-12-15/web/packages/PlackettLuce/vignettes/Overview.html>.
- WSOP (2019a). Hossein Ensan wins the 2019 WSOP Main Event (\$10,000,000). Available at <https://www.wsop.com/tournaments/updates/?aid=2&grid=1622&tid=17298&dayof=7661&rr=5&curpage=4>.
- WSOP (2019b). John Gorsuch completes epic comeback to win 2019 WSOP Millionaire Maker for \$1,344,930. Available at <https://www.wsop.com/tournaments/updates/?aid=2&grid=1622&tid=17287&dayof=7470&rr=5>.

Confidence Intervals for Seroprevalence

Thomas J. DiCiccio, David M. Ritzwoller, Joseph P. Romano and Azeem M. Shaikh

Abstract. This paper concerns the construction of confidence intervals in standard seroprevalence surveys. In particular, we discuss methods for constructing confidence intervals for the proportion of individuals in a population infected with a disease using a sample of antibody test results and measurements of the test's false positive and false negative rates. We begin by documenting erratic behavior in the coverage probabilities of standard Wald and percentile bootstrap intervals when applied to this problem. We then consider two alternative sets of intervals constructed with test inversion. The first set of intervals are approximate, using either asymptotic or bootstrap approximation to the finite-sample distribution of a chosen test statistic. We consider several choices of test statistic, including maximum likelihood estimators and generalized likelihood ratio statistics. We show with simulation that, at empirically relevant parameter values and sample sizes, the coverage probabilities for these intervals are close to their nominal level and are approximately equi-tailed. The second set of intervals are shown to contain the true parameter value with probability at least equal to the nominal level, but can be conservative in finite samples.

Key words and phrases: Confidence intervals, novel coronavirus, serology testing, seroprevalence, test inversion.

REFERENCES

- AGRESTI, A. and MIN, Y. (2002). Unconditional small-sample confidence intervals for the odds ratio. *Biostatistics* **3** 379–386.
- AINSWORTH, M., ANDERSSON, M., AUCKLAND, K., BAILLIE, J. K., BARNES, E., BEER, S., BEVERIDGE, A., BIBI, S., BLACKWELL, L. et al. (2020). Performance characteristics of five immunoassays for SARS-CoV-2: A head-to-head benchmark comparison. *Lancet Infect. Dis.* **20** 1390–1400.
- ALTER, G. and SEDER, R. (2020). The power of antibody-based surveillance. *N. Engl. J. Med.* **383** 1782–1784. <https://doi.org/10.1056/NEJMe2028079>
- BARNDORFF-NIELSEN, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. MR0855891 <https://doi.org/10.2307/2336207>
- BENDAVID, E., MULANEY, B., SOOD, N., SHAH, S., LING, E., BROMLEY-DULFANO, R., LAI, C., WEISSBERG, Z., SAAVEDRA, R. et al. (2020a). COVID-19 antibody seroprevalence in Santa Clara County, California. *MedRxiv* **April 11**.
- BENDAVID, E., MULANEY, B., SOOD, N., SHAH, S., LING, E., BROMLEY-DULFANO, R., LAI, C., WEISSBERG, Z., SAAVEDRA, R. et al. (2020b). COVID-19 antibody seroprevalence in Santa Clara County, California. *MedRxiv* **April 27**.
- BENDAVID, E., MULANEY, B., SOOD, N., SHAH, S., BROMLEY-DULFANO, R., LAI, C., WEISSBERG, Z., SAAVEDRA-WALKER, R., TEDROW, J. et al. (2021). COVID-19 antibody seroprevalence in Santa Clara County, California. *Int. J. Epidemiol.* **50** 410–419. <https://doi.org/10.1093/ije/dyab010>
- BENNETT, S. T. and STEYVERS, M. (2020). Estimating COVID-19 antibody seroprevalence in Santa Clara County, California. A reanalysis of Bendavid et al. *MedRxiv*.
- BERGER, R. L. and BOOS, D. D. (1994). P values maximized over a confidence set for the nuisance parameter. *J. Amer. Statist. Assoc.* **89** 1012–1016. MR1294746
- BRAZZALE, A. R., DAVISON, A. C. and REID, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **23**. Cambridge Univ. Press, Cambridge. MR2342742 <https://doi.org/10.1017/CBO9780511611131>
- BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. MR1861069 <https://doi.org/10.1214/ss/1009213286>
- CAI, B., IOANNIDIS, J., BENDAVID, E. and TIAN, L. (2020). Exact inference for disease prevalence based on a test with unknown specificity and sensitivity. Preprint. Available at [arXiv:2011.14423](https://arxiv.org/abs/2011.14423).
- CARPENTER, J. (1999). Test inversion bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 159–172. MR1664041

Thomas J. DiCiccio is an Associate Professor in the Department of Social Statistics, School of Industrial and Labor Relations at Cornell University, Ithaca, New York 14853, USA (e-mail: tjd9@cornell.edu). David M. Ritzwoller is a Ph.D student in Economics at the Stanford Graduate School of Business, Stanford, California 94305, USA (e-mail: ritzwooll@stanford.edu). Joseph P. Romano is a Professor in the Departments of Statistics and Economics at Stanford University, Stanford, California 94305, USA (e-mail: romano@stanford.edu). Azeem M. Shaikh is the Ralph and Mary Otis Isham Professor in the Kenneth C. Griffin Department of Economics at the University of Chicago, Chicago, Illinois 60637, USA (e-mail: amshaikh@uchicago.edu).

- <https://doi.org/10.1111/1467-9868.00169>
- CLOPPER, C. J. and PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** 404–413.
- COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39. [MR0893334](https://doi.org/10.2307/3313334)
- DATTA, G. S. and MUKERJEE, R. (2004). *Probability Matching Priors: Higher Order Asymptotics. Lecture Notes in Statistics* **178**. Springer, New York. [MR2053794](https://doi.org/10.1007/978-1-4612-2036-7) <https://doi.org/10.1007/978-1-4612-2036-7>
- DEEKS, J. J., DINNES, J., TAKWOINGI, Y., DAVENPORT, C., SPIJKER, R., TAYLOR-PHILLIPS, S., ADRIANO, A., BEESE, S., DRETZKE, J. et al. (2020). Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst. Rev.* **6** CD013652. <https://doi.org/10.1002/14651858.CD013652>
- DiCICCIO, T. J., MARTIN, M. A. and STERN, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *Canad. J. Statist.* **29** 67–76. [MR1834487](https://doi.org/10.2307/3316051) <https://doi.org/10.2307/3316051>
- DiCICCIO, T. J. and ROMANO, J. P. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *Int. Stat. Rev.* **58** 59–76.
- DiCICCIO, T. J. and ROMANO, J. P. (1995). On bootstrap procedures for second-order accurate confidence limits in parametric models. *Statist. Sinica* **5** 141–160. [MR1329292](https://doi.org/10.2307/3312929)
- EFRON, B. (1981). Nonparametric standard errors and confidence intervals. *Canad. J. Statist.* **9** 139–158. [MR0640014](https://doi.org/10.2307/3314608) <https://doi.org/10.2307/3314608>
- EFRON, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82** 171–185. [MR0883345](https://doi.org/10.2307/2286345)
- EISEN, M. B. and TIBSHIRANI, R. (2020). How to identify flawed research before it becomes dangerous. *N.Y. Times* **21**.
- FAGERLAND, M. W., LYDERSEN, S. and LAAKE, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Stat. Methods Med. Res.* **24** 224–254. [MR3336292](https://doi.org/10.1177/0962280211415469) <https://doi.org/10.1177/0962280211415469>
- FAUCI, A. S., LANE, H. C. and REDFIELD, R. R. (2020). Covid-19—Navigating the uncharted. *N. Engl. J. Med.* **382** 1268–1269. <https://doi.org/10.1056/NEJMe2002387>
- FITHIAN, W. (2020a). Statistical comment on the revision of Bendavid et al. Available at <https://www.stat.berkeley.edu/~wfithian/overdispersionSimple.html>.
- FITHIAN, W. (2020b). “I am grateful to moderator @jonc101x and speaker @jsross119 for allowing me to make a brief statement at the very interesting Stanford BMIR seminar yesterday (43:00 mark). A lightly edited text version follows:”. *Tweet @wfithian*, **22 May, 10:02 A.M.**
- FRASER, D. A. S. and REID, N. (1988). On conditional inference for a real parameter: A differential approach on the sample space. *Biometrika* **75** 251–264. [MR0946050](https://doi.org/10.1093/biomet/75.2.251) <https://doi.org/10.1093/biomet/75.2.251>
- FRYDENBERG, M. and JENSEN, J. L. (1989). Is the “improved likelihood ratio statistic” really improved in the discrete case? *Biometrika* **76** 655–661. [MR1041410](https://doi.org/10.1093/biomet/76.4.655) <https://doi.org/10.1093/biomet/76.4.655>
- GELMAN, A. (2020). Concerns with that Stanford study of coronavirus prevalence. In *Statistical Modeling, Causal Inference, and Social Science Blog*.
- GELMAN, A. and CARPENTER, B. (2020). Bayesian analysis of tests with unknown specificity and sensitivity. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 1269–1283. [MR4166866](https://doi.org/10.1093/biomet/69.3.647)
- HALL, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika* **69** 647–652. [MR0695211](https://doi.org/10.1093/biomet/69.3.647) <https://doi.org/10.1093/biomet/69.3.647>
- HALL, P. (2013). *The Bootstrap and Edgeworth Expansion. Springer Series in Statistics*. Springer, New York. [MR1145237](https://doi.org/10.1007/978-1-4612-4384-7) <https://doi.org/10.1007/978-1-4612-4384-7>
- HUI, S. L. and WALTER, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36** 167–171.
- JENSEN, J. L. (1986). Similar tests and the standardized log likelihood ratio statistic. *Biometrika* **73** 567–572. [MR0897847](https://doi.org/10.1093/biomet/73.3.567) <https://doi.org/10.1093/biomet/73.3.567>
- JENSEN, J. L. (1992). The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79** 693–703. [MR1209471](https://doi.org/10.1093/biomet/79.4.693) <https://doi.org/10.1093/biomet/79.4.693>
- JEWELL, N. P. (2004). *Statistics for Epidemiology*. CRC Press, Boca Raton, FL.
- KOLATA, G. (2020). Coronavirus infections may not be uncommon, tests suggest. *N.Y. Times*.
- KRAMMER, F. and SIMON, V. (2020). Serology assays to manage COVID-19. *Science* **368** 1060–1061. <https://doi.org/10.1126/science.abc1227>
- LAWLEY, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43** 295–303. [MR0082237](https://doi.org/10.1093/biomet/43.3-4.295) <https://doi.org/10.1093/biomet/43.3-4.295>
- LEE, S. M. S. and YOUNG, G. A. (2005). Parametric bootstrapping with nuisance parameters. *Statist. Probab. Lett.* **71** 143–153. [MR2126770](https://doi.org/10.1016/j.spl.2004.10.026) <https://doi.org/10.1016/j.spl.2004.10.026>
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. [MR2135927](https://doi.org/10.1007/978-1-4939-9826-3)
- MALLAPATY, S. (2020). Antibody tests suggest that coronavirus infections vastly exceed official counts. *Nature (London)*.
- PEELING, R. W., WEDDERBURN, C. J., GARCIA, P. J., BOERAS, D., FONGWEN, N., NKENGASONG, J., SALL, A., TANURI, A. and HEYMANN, D. L. (2020). Serology testing in the COVID-19 pandemic response. *Lancet Infect. Dis.* **20** e245–e249. [https://doi.org/10.1016/S1473-3099\(20\)30517-X](https://doi.org/10.1016/S1473-3099(20)30517-X)
- ROGAN, W. J. and GLADEN, B. (1978). Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* **107** 71–76.
- ROMANO, J. P., SHAIKH, A. and WOLF, M. (2011). Consonance and the closure method in multiple testing. *Int. J. Biostat.* **7** Art. 12, 27 pp. [MR2775079](https://doi.org/10.2202/1557-4679.1300) <https://doi.org/10.2202/1557-4679.1300>
- ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica* **82** 1979–2002. [MR3268401](https://doi.org/10.3982/ECTA11011) <https://doi.org/10.3982/ECTA11011>
- SEVERINI, T. A. (1999). On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. *Statist. Sinica* **9** 713–724. [MR1711651](https://doi.org/10.2307/3311651)
- SILVAPULLE, M. J. (1996). A test in the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **91** 1690–1693. [MR1439111](https://doi.org/10.2307/2291597) <https://doi.org/10.2307/2291597>
- TOULIS, P. (2021). Estimation of Covid-19 prevalence from serology tests: A partial identification approach. *J. Econometrics* **220** 193–213. [MR4185131](https://doi.org/10.1016/j.jeconom.2020.10.005) <https://doi.org/10.1016/j.jeconom.2020.10.005>
- WALTER, S. D. and IRWIG, L. M. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. *J. Clin. Epidemiol.* **41** 923–937.

Statistical Aspects of the Quantum Supremacy Demonstration

Yosef Rinott, Tomer Shoham and Gil Kalai

Abstract. In quantum computing, a demonstration of *quantum supremacy* (or quantum advantage) consists of presenting a task, possibly of no practical value, whose computation is feasible on a quantum device, but cannot be performed by classical computers in any feasible amount of time. The notable claim of quantum supremacy presented by Google’s team in 2019 consists of demonstrating the ability of a quantum circuit to generate, albeit with considerable noise, bitstrings from a distribution that is considered hard to simulate on classical computers. Very recently, in 2020, a quantum supremacy claim was presented by a group from the University of Science and Technology of China, using a different technology and generating a different distribution, but sharing some statistical principles with Google’s demonstration.

Verifying that the generated data is indeed from the claimed distribution and assessing the circuit’s noise level and its fidelity is a statistical undertaking. The objective of this paper is to explain the relations between quantum computing and some of the statistical aspects involved in demonstrating quantum supremacy in terms that are accessible to statisticians, computer scientists, and mathematicians. Starting with the statistical modeling and analysis in Google’s demonstration, which we explain, we study various estimators of the fidelity, and different approaches to testing the distributions generated by the quantum computer. We propose different noise models, and discuss their implications. A preliminary study of the Google data, focusing mostly on circuits of 12 and 14 qubits is given in different parts of the paper.

Key words and phrases: Google’s quantum computer, random distributions, estimation of sampling weights, size bias.

REFERENCES

- [1] AARONSON, S. and GUNN, S. (2020). On the classical hardness of spoofing linear cross-entropy benchmarking. *Theory Comput.* **16** 11. MR4194611 <https://doi.org/10.4086/toc.2020.v016a011>
- [2] ARRATIA, R., GOLDSTEIN, L. and KOCHMAN, F. (2019). Size bias for one and all. *Probab. Surv.* **16** 1–61. MR3896143 <https://doi.org/10.1214/13-ps221>
- [3] ARUTE et al. (2019). Quantum supremacy using a programmable superconducting processor. *Nature* **574** 505–510. <https://doi.org/10.5061/dryad.k6t1rj8>
- [4] CHAN, K. C. G. (2013). Survival analysis without survival data: Connecting length-biased and case-control data. *Biometrika* **100** 764–770. MR3094451 <https://doi.org/10.1093/biomet/ast008>
- [5] COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. MR0370837
- [6] DEVITT, S. J., MUNRO, W. J. and NEMOTO, K. (2013). Quantum error correction for beginners. *Rep. Progr. Phys.* **76** 076001. <https://doi.org/10.1088/0034-4885/76/7/076001>
- [7] GREENE, W. H. (2003). *Econometric Analysis*, 5th ed. Prentice Hall, Hoboken, NJ.
- [8] HENSHER, D. A., ROSE, J. M. and GREENE, W. H. (2015). *Applied Choice Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge.
- [9] HUANG, C. et al. (2020). Classical simulation of quantum supremacy circuits. Available at [arXiv:2005.06787](https://arxiv.org/abs/2005.06787).
- [10] KALAI, G. (2020). The argument against quantum computers, the quantum laws of nature, and Google’s supremacy claims. In *The Intercontinental Academia Laws: Rigidity and Dynamics* (M. J. Hannon and E. Z. Rabinovici, eds.). *Proceedings of the ICA Workshops 2018 & 2019*. World Scientific, Singapore. To appear.
- [11] KALAI, G. and KINDLER, G. (2014). Gaussian noise sen-

Yosef Rinott is Professor, Federmann Center for the Study of Rationality and Department of Statistics, The Hebrew University of Jerusalem, Israel (e-mail: yosef.rinott@mail.huji.ac.il). Tomer Shoham is a graduate student, Federmann Center for the Study of Rationality and Department of Statistics, The Hebrew University of Jerusalem, Israel (e-mail: tomer.shohamm@gmail.com). Gil Kalai is Professor, Einstein Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel and Efi Arazy School of Computer Science, IDC, Herzliya, Israel (e-mail: gil.kalai@gmail.com).

- sitivity and boson sampling. arXiv preprint. Available at [arXiv:1409.3093](https://arxiv.org/abs/1409.3093).
- [12] KINGMAN, J. F. C., (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B* **37** 1–15. [MR0368264](https://doi.org/10.2307/2343824)
- [13] KOTZ, S., BALAKRISHNAN, N. and JOHNSON, N. L. (2000). *Continuous Multivariate Distributions. Volume 1: Models and Applications*. Wiley, New York.
- [14] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. [MR1639875](https://doi.org/10.1007/978-1-4612-4378-6)
- [15] NATHAN, G. (2011). Superpopulation models in survey sampling. In *International Encyclopedia of Statistical Science* 1575–1577. Springer, Berlin.
- [16] NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge. [MR1796805](https://doi.org/10.1017/C9780521876223)
- [17] PAN, F. and ZHANG, P. (2021). Simulating the Sycamore quantum supremacy circuits. Available at [arXiv:2103.03074](https://arxiv.org/abs/2103.03074).
- [18] PEDNAULT, E., GUNNELS, J. A., NANNICINI, G., HORESH, L. and WISNIEFF, R. (2019). Leveraging secondary storage to simulate deep 54-qubit Sycamore circuits. Available at [arXiv:1910.09534](https://arxiv.org/abs/1910.09534).
- [19] PORTER, C. E. and THOMAS, R. G. (1956). Fluctuations of nuclear reaction widths. *Phys. Rev.* **104** 483–491.
- [20] RAUSAND, M. and HØYLAND, A. (2004). *System Reliability Theory: Models, Statistical Methods, and Applications*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. [MR2016162](https://doi.org/10.1002/9781118161622)
- [21] SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling. Springer Series in Statistics*. Springer, New York. [MR1140409](https://doi.org/10.1007/978-1-4612-4378-6) <https://doi.org/10.1007/978-1-4612-4378-6>
- [22] WANG, Y. and SONG, X. (2020). Quantum science and quantum technology. *Statist. Sci.* **35** 51–74. [MR4071358](https://doi.org/10.1214/19-STS745) <https://doi.org/10.1214/19-STS745>
- [23] ZHONG, H.-S., WANG, H., DENG, Y.-H., CHEN, M.-C., PENG, L.-C., LUO, Y.-H., QIN, J., WU, D., DING, X. et al. (2020). Quantum computational advantage using photons. *Science* **370** 1460–1463. <https://doi.org/10.1126/science.abe8770>

Choosing Among Notions of Multivariate Depth Statistics

Karl Mosler and Pavlo Mozharovskyi

Abstract. Classical multivariate statistics measures the outlyingness of a point by its Mahalanobis distance from the mean, which is based on the mean and the covariance matrix of the data. A multivariate depth function is a function which, given a point and a distribution in d -space, measures centrality by a number between 0 and 1, while satisfying certain postulates regarding invariance, monotonicity, convexity and continuity. Accordingly, numerous notions of multivariate depth have been proposed in the literature, some of which are also robust against extremely outlying data. The departure from classical Mahalanobis distance does not come without cost. There is a trade-off between invariance, robustness and computational feasibility. In the last few years, efficient exact algorithms as well as approximate ones have been constructed and made available in R-packages. Consequently, in practical applications the choice of a depth statistic is no more restricted to one or two notions due to computational limits; rather often more notions are feasible, among which the researcher has to decide. The article debates theoretical and practical aspects of this choice, including invariance and uniqueness, robustness and computational feasibility. Complexity and speed of exact algorithms are compared. The accuracy of approximate approaches like the random Tukey depth is discussed as well as the application to large and high-dimensional data. Extensions to local and functional depths and connections to regression depth are shortly addressed.

Key words and phrases: Depth statistics, computational complexity, robustness, approximation, random Tukey depth.

REFERENCES

- AGOSTINELLI, C. and ROMANAZZI, M. (2011). Local depth. *J. Statist. Plann. Inference* **141** 817–830. MR2732952 <https://doi.org/10.1016/j.jspi.2010.08.001>
- AZZALINI, A. and DALLA VALLE, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83** 715–726. MR1440039 <https://doi.org/10.1093/biomet/83.4.715>
- BARBER, C. B., DOBKIN, D. P. and HUHDANPAA, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Software* **22** 469–483. MR1428265 <https://doi.org/10.1145/235815.235821>
- BARNETT, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A* **139** 318–355. MR0445726 <https://doi.org/10.2307/2344839>
- BAZOVKIN, P. and MOSLER, K. (2012). An exact algorithm for weighted-mean trimmed regions in any dimension. *J. Stat. Softw.* **47** 1–29.
- BOČEK, P. and ŠIMAN, M. (2016). Directional quantile regression in Octave (and MATLAB). *Kybernetika (Prague)* **52** 28–51. MR3482609 <https://doi.org/10.14736/kyb-2016-1-0028>
- CASCOS, I. and LÓPEZ-DÍAZ, M. (2016). On the uniform consistency of the zonoid depth. *J. Multivariate Anal.* **143** 394–397. MR3431440 <https://doi.org/10.1016/j.jmva.2015.09.020>
- CASCOS, I. and MOLCHANOV, I. (2007). Multivariate risks and depth-trimmed regions. *Finance Stoch.* **11** 373–397. MR2322918 <https://doi.org/10.1007/s00780-007-0043-7>
- CHAKRABORTY, B. (2003). On multivariate quantile regression. *J. Statist. Plann. Inference* **110** 109–132. MR1944636 [https://doi.org/10.1016/S0378-3758\(01\)00277-4](https://doi.org/10.1016/S0378-3758(01)00277-4)
- CHEN, Y., DANG, X., PENG, H. and BART, H. L. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 288–305.
- CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 <https://doi.org/10.1214/16-AOS1450>
- CHOLAQUIDIS, A., FRAIMAN, R., GAMBOA, F. and MORENO, L. (2020). Weighted lens depth: Some applications to supervised classification. Available at arXiv:2011.11140.

- CHOWDHURY, J. and CHAUDHURI, P. (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli* **25** 395–423. MR3892324 <https://doi.org/10.3150/17-bej991>
- CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207 <https://doi.org/10.1016/j.csda.2008.04.021>
- DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313 <https://doi.org/10.1214/aos/1176348890>
- DÜMBGEN, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.* **14** 119–128. MR1173409 [https://doi.org/10.1016/0167-7152\(92\)90075-G](https://doi.org/10.1016/0167-7152(92)90075-G)
- DUTTA, S., GHOSH, A. K. and CHAUDHURI, P. (2011). Some intriguing properties of Tukey’s half-space depth. *Bernoulli* **17** 1420–1434. MR2854779 <https://doi.org/10.3150/10-BEJ322>
- DUTTA, S., SARKAR, S. and GHOSH, A. K. (2016). Multi-scale classification using localized spatial depth. *J. Mach. Learn. Res.* **17** Paper No. 218, 30. MR3595152
- DYCKERHOFF, R. (2002). Datentiefe: Begriff, Berechnung, Tests. Mimeo, Fakultät für Wirtschafts- und Sozialwissenschaften, Univ. zu Köln.
- DYCKERHOFF, R. (2004). Data depths satisfying the projection property. *Allg. Stat. Arch.* **88** 163–190. MR2074729 <https://doi.org/10.1007/s101820400167>
- DYCKERHOFF, R., KOSHEVOY, G. and MOSLER, K. (1996). Zonoid data depth: Theory and computation. In *COMPSTAT 1996. Proceedings in Computational Statistics* (A. Pratt, ed.) 235–240. Physica-Verlag, Heidelberg.
- DYCKERHOFF, R., LEY, C. and PAINDAVEINE, D. (2015). Depth-based runs tests for bivariate central symmetry. *Ann. Inst. Statist. Math.* **67** 917–941. MR3390173 <https://doi.org/10.1007/s10463-014-0480-y>
- DYCKERHOFF, R. and MOSLER, K. (2011). Weighted-mean trimming of multivariate data. *J. Multivariate Anal.* **102** 405–421. MR2755006 <https://doi.org/10.1016/j.jmva.2010.10.002>
- DYCKERHOFF, R. and MOZHAROVSKIY, P. (2016). Exact computation of the halfspace depth. *Comput. Statist. Data Anal.* **98** 19–30. MR3458519 <https://doi.org/10.1016/j.csda.2015.12.011>
- EDDY, W. F. (1981). Graphics for the multivariate two-sample problem: Comment. *J. Amer. Statist. Assoc.* **76** 287–289.
- EDELSBRUNNER, H. (1987). *Algorithms in Combinatorial Geometry*. *EATCS Monographs on Theoretical Computer Science* **10**. Springer, Berlin. MR0904271 <https://doi.org/10.1007/978-3-642-61568-9>
- ELMORE, R. T., HETTMANSPERGER, T. P. and XUAN, F. (2006). Spherical data depth and a multivariate median. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* (R. Y. Lui, R. Serfling and D. L. Souvaine, eds.) *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **72** 87–101. Amer. Math. Soc., Providence, RI. MR2343115 <https://doi.org/10.1090/dimacs/072/07>
- FEBRERO-BANDE, M. and OVIEDO DE LA FUENTE, M. (2012). Statistical computing in functional data analysis: The R package *fd*. *J. Stat. Softw.* **51** 1–28.
- FISCHER, D., MOSLER, K., MÖTTÖNEN, J., NORDHAUSEN, K., POKOTYLO, O. and VOGEL, D. (2020). Computing the Oja median in R: The package *OjaNP*. *J. Stat. Softw.* **92** 1–36.
- GENEST, M., MASSE, J.-C. and PLANTE, J.-F. (2019). *depth: Nonparametric Depth Functions for Multivariate Analysis*. R package version 2.1-1.1.
- GIJBELS, I. and NAGY, S. (2017). On a general definition of depth for functional data. *Statist. Sci.* **32** 630–639. MR3730526 <https://doi.org/10.1214/17-STS625>
- HABEL, K., GRASMAN, R., GRAMACY, R. B., MOZHAROVSKIY, P. and STERRATT, D. C. (2019). *geometry: Mesh Generation and Surface Tessellation*. R package version 0.4.5.
- HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 <https://doi.org/10.1214/09-AOS723>
- HALLIN, M., LU, Z., PAINDAVEINE, D. and ŠIMAN, M. (2015). Local bilinear multiple-output quantile/depth regression. *Bernoulli* **21** 1435–1466. MR3352050 <https://doi.org/10.3150/14-BEJ610>
- HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d : A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 <https://doi.org/10.1214/20-aos1996>
- HLUBINKA, D., KOTÍK, L. and VENCÁLEK, O. (2010). Weighted halfspace depth. *Kybernetika (Prague)* **46** 125–148. MR2666899
- HOBERG, R. (2000). Cluster analysis based on data depth. In *Data Analysis, Classification, and Related Methods* (H. A. L. Kiers, J. P. Rasson, P. J. F. Groenen and M. Schader, eds.) *Studies in Classification, Data Analysis, and Knowledge Organization* 17–22. Springer, Berlin.
- HU, Y., WANG, Y., WU, Y., LI, Q. and HOU, C. (2011). Generalized Mahalanobis depth in the reproducing kernel Hilbert space. *Statist. Papers* **52** 511–522. MR2821054 <https://doi.org/10.1007/s00362-009-0265-1>
- HUBERT, M., ROUSSEEUW, P. J. and SEGAERT, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.* **24** 177–202. MR3376852 <https://doi.org/10.1007/s10260-015-0297-8>
- KENT, J. T., ER, F. and CONSTABLE, P. D. L. (2015). Algorithms for the spatial median. In *Modern Nonparametric, Robust and Multivariate Methods* (K. Nordhausen and S. Taskinen, eds.) 205–224. Springer, Cham. MR3444328
- KLEINDESSNER, M. and VON LUXBURG, U. (2017). Lens depth function and k -relative neighborhood graph: Versatile tools for ordinal data analysis. *J. Mach. Learn. Res.* **18** Paper No. 58, 52. MR3687601
- KOSHEVOY, G. (1997). Integrable L -statistics, depths and ranks. Unpublished mimeo.
- KOSHEVOY, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.* **83** 360–364. MR1945958 <https://doi.org/10.1006/jmva.2001.2052>
- KOSHEVOY, G. A. (2003). Lift-zonoid and multivariate depths. In *Developments in Robust Statistics (Vorau, 2001)* (R. Dutter, P. Filzmoser, U. Gather and P. J. Rousseeuw, eds.) 194–202. Physica, Heidelberg. MR1977477
- KOSHEVOY, G. and MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998–2017. MR1474078 <https://doi.org/10.1214/aos/1069362382>
- KOSIOROWSKI, D. and ZAWADZKI, Z. (2019). *DepthProc: An R package for robust exploration of multidimensional economic phenomena*. Available at arXiv:1408.4542.
- KUELBS, J. and ZINN, J. (2013). Concerns with functional depth. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 831–855. MR3125749
- KUSTOSZ, C. P., MÜLLER, C. H. and WENDLER, M. (2016). Simplified simplicial depth for regression and autoregressive growth processes. *J. Statist. Plann. Inference* **173** 125–146. MR3465017 <https://doi.org/10.1016/j.jspi.2016.01.005>
- LANGE, T., MOSLER, K. and MOZHAROVSKIY, P. (2014a). Fast nonparametric classification based on data depth. *Statist. Papers* **55** 49–69. MR3152767 <https://doi.org/10.1007/s00362-012-0488-4>
- LANGE, T., MOSLER, K. and MOZHAROVSKIY, P. (2014b). $DD\alpha$ -classification of asymmetric and fat-tailed data. In *Data Analysis, Machine Learning and Knowledge Discovery* (M. Spiliopoulou, L. Schmidt-Thieme and R. Janning, eds.) 71–78. Springer, Cham.

- LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 <https://doi.org/10.1214/aos/1176347507>
- LIU, R. Y. (1992). Data depth and multivariate rank tests. In *L₁-Statistical Analysis and Related Methods (Neuchâtel, 1992)* 279–294. North-Holland, Amsterdam. MR1214839
- LIU, Z. and MODARRES, R. (2011). Lens data depth and median. *J. Nonparametr. Stat.* **23** 1063–1074. MR2854255 <https://doi.org/10.1080/10485252.2011.584621>
- LIU, X., MOSLER, K. and MOZHAROVSKIY, P. (2019). Fast computation of Tukey trimmed regions and median in dimension $p > 2$. *J. Comput. Graph. Statist.* **28** 682–697. MR4007750 <https://doi.org/10.1080/10618600.2018.1546595>
- LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. With discussion and a rejoinder by Liu and Singh. MR1724033 <https://doi.org/10.1214/aos/1018031260>
- LIU, X. and ZUO, Y. (2014). Computing projection depth and its associated estimators. *Stat. Comput.* **24** 51–63. MR3147697 <https://doi.org/10.1007/s11222-012-9352-6>
- LIU, X. and ZUO, Y. (2015). CompPD: A MATLAB package for computing projection depth. *J. Stat. Softw.* **65** 1–21.
- LOPUHÄÄ, H. P. and ROUSSEEUW, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248. MR1091847 <https://doi.org/10.1214/aos/1176347978>
- MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* **2** 49–55.
- MAHALANOBISH, O. and KARMAKAR, S. (2015). depth.plot: Multivariate Analogy of Quantiles. R package version 0.1.
- MILASEVIC, P. and DUCHARME, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15** 1332–1333. MR0902264 <https://doi.org/10.1214/aos/1176350511>
- MIZERA, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736. MR1969447 <https://doi.org/10.1214/aos/1043351254>
- MIZERA, I. and MÜLLER, C. H. (2004). Location-scale depth. *J. Amer. Statist. Assoc.* **99** 949–989. MR2109488 <https://doi.org/10.1198/016214504000001312>
- MIZERA, I. and VOLAUF, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: Diagnostics of depth-related methods. *J. Multivariate Anal.* **83** 365–388. MR1945959 <https://doi.org/10.1006/jmva.2001.2053>
- MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth. Lecture Notes in Statistics* **165**. Springer, Berlin. The lift zonoid approach. MR1913862 <https://doi.org/10.1007/978-1-4613-0045-8>
- MOSLER, K. and BAZOVKIN, P. (2014). Stochastic linear programming with a distortion risk constraint. *OR Spectrum* **36** 949–969. MR3258501 <https://doi.org/10.1007/s00291-014-0372-9>
- MOSLER, K., LANGE, T. and BAZOVKIN, P. (2009). Computing zonoid trimmed regions of dimension $d > 2$. *Comput. Statist. Data Anal.* **53** 2500–2510. MR2665902 <https://doi.org/10.1016/j.csda.2009.01.017>
- MOSLER, K. and POLYAKOVA, Y. (2018). General notions of depth for functional data. Available at [arXiv:1208.1981](https://arxiv.org/abs/1208.1981).
- MOZHAROVSKIY, P., MOSLER, K. and LANGE, T. (2015). Classifying real-world data with the $DD\alpha$ -procedure. *Adv. Data Anal. Classif.* **9** 287–314. MR3392037 <https://doi.org/10.1007/s11634-014-0180-8>
- NAGY, S. (2017). Monotonicity properties of spatial depth. *Statist. Probab. Lett.* **129** 373–378. MR3688558 <https://doi.org/10.1016/j.spl.2017.06.025>
- NAGY, S. (2020). The halfspace depth characterization problem. In *Nonparametric Statistics (M. La Rocca, B. Liseo and L. Salmaso, eds.)*. Springer Proc. Math. Stat. **339** 379–389. Springer, Cham. MR4178663 https://doi.org/10.1007/978-3-030-57306-5_34
- NAGY, S., DYCKERHOFF, R. and MOZHAROVSKIY, P. (2020). Uniform convergence rates for the approximated halfspace and projection depth. *Electron. J. Stat.* **14** 3939–3975. MR4165498 <https://doi.org/10.1214/20-EJS1759>
- NAGY, S., SCHÜTT, C. and WERNER, E. M. (2019). Halfspace depth and floating body. *Stat. Surv.* **13** 52–118. MR3973130 <https://doi.org/10.1214/19-ss123>
- NAGY, S., GIJBELS, I., OMEKKA, M. and HLUBINKA, D. (2016). Integrated depth for functional data: Statistical properties and consistency. *ESAIM Probab. Stat.* **20** 95–130. MR3528619 <https://doi.org/10.1051/ps/2016005>
- NIETO-REYES, A. and BATTEY, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.* **31** 61–79. Correction to this article: *Statistical Science* **32** (2017) p. 640. MR3458593 <https://doi.org/10.1214/15-STS532>
- NIINIMAA, A. and OJA, H. (1995). On the influence functions of certain bivariate medians. *J. Roy. Statist. Soc. Ser. B* **57** 565–574. MR1341324
- NIINIMAA, A., OJA, H. and TABLEMAN, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Lett.* **10** 325–328. MR1069910 [https://doi.org/10.1016/0167-7152\(90\)90050-H](https://doi.org/10.1016/0167-7152(90)90050-H)
- NORDHAUSEN, K. and TYLER, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika* **102** 573–588. MR3394276 <https://doi.org/10.1093/biomet/asv022>
- OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. MR0721446 [https://doi.org/10.1016/0167-7152\(83\)90054-8](https://doi.org/10.1016/0167-7152(83)90054-8)
- OSTRESH, L. M. JR. (1978). On the convergence of a class of iterative methods for solving the Weber location problem. *Oper. Res.* **26** 597–609. MR0496728 <https://doi.org/10.1287/opre.26.4.597>
- PAINDAVEINE, D. and ŠIMAN, M. (2012a). Computing multiple-output regression quantile regions. *Comput. Statist. Data Anal.* **56** 840–853. MR2888729 <https://doi.org/10.1016/j.csda.2010.11.014>
- PAINDAVEINE, D. and ŠIMAN, M. (2012b). Computing multiple-output regression quantile regions from projection quantiles. *Comput. Statist.* **27** 29–49. MR2877809 <https://doi.org/10.1007/s00180-011-0231-y>
- PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth: A focus on centrality. *J. Amer. Statist. Assoc.* **108** 1105–1119. MR3174687 <https://doi.org/10.1080/01621459.2013.813390>
- PAINDAVEINE, D. and VAN BEVER, G. (2018). Halfspace depths for scatter, concentration and shape matrices. *Ann. Statist.* **46** 3276–3307. MR3852652 <https://doi.org/10.1214/17-AOS1658>
- POKOTYLO, O., MOZHAROVSKIY, P. and DYCKERHOFF, R. (2019). Depth and depth-based classification with R-package ddalpha. *J. Stat. Softw.* **91** 1–46.
- POKOTYLO, O., MOZHAROVSKIY, P., DYCKERHOFF, R. and NAGY, S. (2020). ddalpha: Depth-Based Classification and Calculation of Data Depth. R package version 1.3.11.
- ROMANAZZI, M. (2001). Influence function of halfspace depth. *J. Multivariate Anal.* **77** 138–161. MR1838717 <https://doi.org/10.1006/jmva.2000.1929>
- ROUSSEEUW, P. J. and HUBERT, M. (1999). Depth in an arrangement of hyperplanes. *Discrete Comput. Geom.* **22** 167–176. MR1698539 <https://doi.org/10.1007/PL00009452>
- ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR0914792 <https://doi.org/10.1002/0471725382>
- ROUSSEEUW, P. J. and RUTS, I. (1998). Constructing the bivariate Tukey median. *Statist. Sinica* **8** 827–839. MR1651511

- ROUSSEEUW, P. J., RUTS, I. and TUKEY, J. W. (1999). The bagplot: A bivariate boxplot. *Amer. Statist.* **53** 382–387.
- ROUSSEEUW, P. J. and STRUYF, A. (1998). Computing location depth and regression depth in higher dimensions. *Stat. Comput.* **8** 193–203.
- RUTS, I. and ROUSSEEUW, P. J. (1996). Computing depth contours of bivariate point clouds. *Comput. Statist. Data Anal.* **23** 153–168.
- SEGAERT, P., HUBERT, M., ROUSSEEUW, P. and RAYMAEKERS, J. (2020). mrfDepth: Depth Measures in Multivariate, Regression and Functional Settings. R package version 1.0.13.
- SERFLING, R. (2002). A depth function and a scale curve based on spatial quantiles. In *Statistical Data Analysis Based on the L_1 -Norm and Related Methods (Neuchâtel, 2002)*. *Stat. Ind. Technol.* 25–38. Birkhäuser, Basel. MR2001302
- SERFLING, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. MR2738875 <https://doi.org/10.1080/10485250903431710>
- ŠIMAN, M. and BOČEK, P. (2019). modQR: Multiple-Output Directional Quantile Regression. R package version 0.1.2.
- TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2 523–531. MR0426989
- YANG, M. and MODARRES, R. (2018). β -skeleton depth functions and medians. *Comm. Statist. Theory Methods* **47** 5127–5143. MR3833887 <https://doi.org/10.1080/03610926.2017.1386320>
- ZAWADZKI, Z., KOSIOROWSKI, D., SŁOMCZYNSKI, K., BOCIAN, M. and WĘGRZYŃKIEWICZ, A. (2020). DepthProc: Statistical Depth Functions for Multivariate Analysis. R package version 2.1.3.
- ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822 <https://doi.org/10.1214/aos/1065705115>
- ZUO, Y. (2004). Robustness of weighted L^P -depth and L^P -median. *Allg. Stat. Arch.* **88** 215–234. MR2074731 <https://doi.org/10.1007/s101820400169>
- ZUO, Y. (2006). Multidimensional trimming based on projection depth. *Ann. Statist.* **34** 2211–2251. MR2291498 <https://doi.org/10.1214/009053606000000713>
- ZUO, Y. (2013). Multidimensional medians and uniqueness. *Comput. Statist. Data Anal.* **66** 82–88. MR3064025 <https://doi.org/10.1016/j.csda.2013.03.020>
- ZUO, Y. (2019). A new approach for the computation of halfspace depth in high dimensions. *Comm. Statist. Simulation Comput.* **48** 900–921. MR3938062 <https://doi.org/10.1080/03610918.2017.1402040>
- ZUO, Y. (2021). On general notions of depth for regression. *Statist. Sci.* **36** 142–157. MR4194208 <https://doi.org/10.1214/20-STS767>
- ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005 <https://doi.org/10.1214/aos/1016218226>

Power Calculations for Replication Studies

Charlotte Micheloud and Leonhard Held

Abstract. The reproducibility crisis has led to an increasing number of replication studies being conducted. Sample sizes for replication studies are often calculated using conditional power based on the effect estimate from the original study. However, this approach is not well suited as it ignores the uncertainty of the original result. Bayesian methods are used in clinical trials to incorporate prior information into power calculations. We propose to adapt this methodology to the replication framework and promote the use of predictive instead of conditional power in the design of replication studies. Moreover, we describe how extensions of the methodology to sequential clinical trials can be tailored to replication studies. Conditional and predictive power calculated at an interim analysis are compared and we argue that predictive power is a useful tool to decide whether to stop a replication study prematurely. A recent project on the replicability of social sciences is used to illustrate the properties of the different methods.

Key words and phrases: Replication studies, conditional power, predictive power, sequential design, interim analysis, futility stopping.

REFERENCES

- ACKERMAN, J. M., NOCERA, C. C. and BARGH, J. A. (2010). Incidental haptic sensations influence social judgments and decisions. *Science* **328** 1712–1715. <https://doi.org/10.1126/science.1189993>
- ANDERSON, S. F. and MAXWELL, S. E. (2017). Addressing the “replication crisis”: Using original studies to design replication studies with appropriate statistical power. *Multivar. Behav. Res.* **52** 305–324. <https://doi.org/10.1080/00273171.2017.1289361>
- BAUER, P. and KÖNIG, F. (2006). The reassessment of trial perspectives from interim data—a critical view. *Stat. Med.* **25** 23–36. <https://doi.org/10.1002/sim.2180>
- BEGLEY, C. G. and IOANNIDIS, J. P. A. (2015). Reproducibility in science. *Circ. Res.* **116** 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>
- BUTTON, K. S., IOANNIDIS, J. P., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. and MUNAFÒ, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14** 365. <https://doi.org/10.1038/nrn3475>
- CAMERER, C. F., DREBER, A., FORSELL, E., HO, T.-H., HUBER, J., JOHANNESSON, M., KIRCHLER, M., ALMENBERG, J., ALTMEJD, A. et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* **351** 1433–1436. <https://doi.org/10.1126/science.aaf0918>
- CAMERER, C. F., DREBER, A., HOLZMEISTER, F., HO, T.-H., HUBER, J., JOHANNESSON, M., KIRCHLER, M., NAVE, G., NOSEK, B. A. et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nat. Hum. Behav.* **2** 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- DALLOW, N. and FINA, P. (2011). The perils with the misuse of predictive power. *Pharm. Stat.* **10** 311–317. <https://doi.org/10.1002/pst.467>
- DEMETS, D. L. (2006). Futility approaches to interim monitoring by data monitoring committees. *Clin. Trials* **3** 522–529. <https://doi.org/10.1177/1740774506073115>
- DUNCAN, K., SADANAND, A. and DAVACHI, L. (2012). Memory’s penumbra: Episodic memory decisions induce lingering mnemonic biases. *Science* **337** 485–487. <https://doi.org/10.1126/science.1221936>
- FDA (1998). Providing clinical evidence of effectiveness for human drug and biological products. Available at www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-clinical-evidence-effectiveness-human-drug-and-biological-products.
- FISHER, L. D. (1999). One large, well-designed, multicenter study as an alternative to the usual FDA paradigm. *Drug Inf. J.* **33** 265–271. <https://doi.org/10.1177/009286159903300130>
- GIBSON, E. W. (2020). The role of p -values in judging the strength of evidence and realistic replication expectations. *Stat. Biopharm. Res.* **13** 6–18. <https://doi.org/10.1080/19466315.2020.1724560>
- GOODMAN, S. N. (1992). A comment on replication, P -values and evidence. *Stat. Med.* **11** 875–879. <https://doi.org/10.1002/sim.4780110705>
- GROUIN, J.-M., COSTE, M., BUNOUF, P. and LECOUTRE, B. (2007). Bayesian sample size determination in non-sequential clinical trials: Statistical aspects and some regulatory considerations. *Stat. Med.* **26** 4914–4924. <https://doi.org/10.1002/sim.2958>

Charlotte Micheloud is Ph.D. student, Department of Biostatistics, Institute of Epidemiology, Biostatistics and Prevention, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland (e-mail: charlotte.micheloud@uzh.ch). Leonhard Held is Professor, Department of Biostatistics, Institute of Epidemiology, Biostatistics and Prevention, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland (e-mail: leonhard.held@uzh.ch).

- HALPERIN, M., LAN, K. K. G., WARE, J. H., JOHNSON, N. J. and DEMETS, D. L. (1982). An aid to data monitoring in long-term clinical trials. *Control. Clin. Trials* **3** 311–323. [https://doi.org/10.1016/0197-2456\(82\)90022-8](https://doi.org/10.1016/0197-2456(82)90022-8)
- HELD, L. (2020). A new standard for the analysis and design of replication studies (with discussion). *J. R. Stat. Soc., A* **183** 431–469. <https://doi.org/10.1111/rssa.12493>
- HERSON, J. (1979). Predictive probability early termination plans for phase II clinical trials. *Biometrics* **35** 775–783.
- ICH E9 EXPERT WORKING GROUP (1999). Statistical principles for clinical trials: ICH harmonised tripartite guideline. *Stat. Med.* **18** 1905–1942.
- IOANNIDIS, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology* **19** 640–648. <https://doi.org/10.1097/EDE.0b013e31818131e7>
- JENNISON, C. and TURNBULL, B. W. (1999). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press/CRC, Boca Raton. <https://doi.org/10.1201/9780367805326>
- JIANG, W. and YU, W. (2016). Power estimation and sample size determination for replication studies of genome-wide association studies. *BMC Genomics* **17** 19–32. <https://doi.org/10.1186/s12864-017-3482-3>
- KIESER, M. and FRIEDE, T. (2003). Simple procedures for blinded sample size adjustment that do not affect the type I error rate. *Stat. Med.* **22** 3571–3581. <https://doi.org/10.1002/sim.1585>
- KUNZMANN, K., GRAYLING, M. J., LEE, K. M., ROBERTSON, D. S., RUFIBACH, K. and WASON, J. M. S. (2020). Conditional power and friends: The why and how of (un)planned, unblinded sample size recalculations in confirmatory trials. Technical Report. Available at <https://arxiv.org/abs/2010.06567>.
- LACHIN, J. M. (2005). A review of methods for futility stopping based on conditional power. *Stat. Med.* **24** 2747–2764. <https://doi.org/10.1002/sim.2151>
- LAKENS, D. (2014). Performing high-powered studies efficiently with sequential analyses. *Eur. J. Soc. Psychol.* **44** 701–710. <https://doi.org/10.1002/ejsp.2023>
- LY, A., ETZ, A., MARSMAN, M. and WAGENMAKERS, E.-J. (2018). Replication Bayes factors from evidence updating. *Behav. Res. Methods* **51** 2498–2508. <https://doi.org/10.3758/s13428-018-1092-x>
- MATTHEWS, J. N. (2006). *Introduction to Randomized Controlled Clinical Trials*. CRC Press/CRC, Boca Raton. <https://doi.org/10.1201/9781420011302>
- MICHELOUD, C. and HELD, L. (2022a). Supplement A to “Power Calculations for Replication Studies.” <https://doi.org/10.1214/21-ST828SUPP>
- MICHELOUD, C. and HELD, L. (2022b). Supplement B to “Power Calculations for Replication Studies.” <https://doi.org/10.1214/21-ST828SUPP>
- O’HAGAN, A. and STEVENS, J. W. (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Med. Decis. Mak.* **21** 219–230. <https://doi.org/10.1177/0272989X0102100307>
- O’HAGAN, A., STEVENS, J. W. and CAMPBELL, M. J. (2005). Assurance in clinical trial design. *Pharm. Stat.* **4** 187–201. <https://doi.org/10.1002/pst.175>
- OPEN SCIENCE COLLABORATION (2015). Estimating the reproducibility of psychological science. *Science* **349** aac4716. <https://doi.org/10.1126/science.aac4716>
- PATIL, P., PENG, R. D. and LEEK, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspect. Psychol. Sci.* **11** 539–544. <https://doi.org/10.1177/1745691616646366>
- PAWEL, S. and HELD, L. (2020). Probabilistic forecasting of replication studies. *PLoS ONE* **15** e0231416. <https://doi.org/10.1371/journal.pone.0231416>
- RAMIREZ, G. and BEILOCK, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science* **331** 211–213. <https://doi.org/10.1126/science.1199427>
- RUFIBACH, K., BURGER, H. U. and ABT, M. (2016). Bayesian predictive power: Choice of prior and some recommendations for its use as probability of success in drug development. *Pharm. Stat.* **15** 438–446. <https://doi.org/10.1002/pst.1764>
- SHAH, A. K., MULLAINATHAN, S. and SHAFIR, E. (2012). Some consequences of having too little. *Science* **338** 682–685. <https://doi.org/10.1126/science.1222426>
- SNAPINN, S., CHEN, M.-G., JIANG, Q. and KOUTSOUKOS, T. (2006). Assessment of futility in clinical trials. *Pharm. Stat.* **5** 273–281. <https://doi.org/10.1002/pst.216>
- SPIEGELHALTER, D. J., ABRAMS, K. R. and MYLES, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, New York. <https://doi.org/10.1002/0470092602>
- SPIEGELHALTER, D. J. and FREEDMAN, L. S. (1986). A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Stat. Med.* **5** 1–13. <https://doi.org/10.1002/sim.4780050103>
- SPIEGELHALTER, D. J., FREEDMAN, L. S. and BLACKBURN, P. R. (1986). Monitoring clinical trials: Conditional or predictive power? *Control. Clin. Trials* **7** 8–17. [https://doi.org/10.1016/0197-2456\(86\)90003-6](https://doi.org/10.1016/0197-2456(86)90003-6)
- WANG, Y., FU, H., KULKARNI, P. and KAISER, C. (2013). Evaluating and utilizing probability of study success in clinical development. *Clin. Trials* **10** 407–413. <https://doi.org/10.1177/1740774513478229>
- WASSMER, G. and BRANNATH, W. (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*. Springer, Berlin. <https://doi.org/10.1007/978-3-319-32562-0>

Intention-to-Treat Comparisons in Randomized Trials

Ross L. Prentice and Aaron K. Aragaki

Abstract. Intention-to-treat (ITT) comparisons have a central place in reporting on randomized controlled trials, though there are typically additional analyses of interest such as those making adjustments for nonadherence. In our ITT reporting of results from the Women’s Health Initiative (WHI) randomized trials, we have relied primarily on highly flexible hazard ratio (Cox) regression methods. However, these methods, especially the proportional hazards special case, have been criticized for being difficult to interpret and frequently oversimplified, and for not being consistent with modern causality theories using potential outcomes. Here we address these topics and extend our use of hazard rate methods for ITT comparisons in the WHI trials.

Key words and phrases: Causality, Cox model, failure time data, regression, restricted mean survival time.

REFERENCES

- AALEN, O., BORGAN, Ø. and GJESSING, H. (2010). *Survival and Event History Analysis: A Process Point of View*. Springer, New York.
- AALEN, O. O., COOK, R. J. and RØYSLAND, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal.* **21** 579–593. MR3397507 <https://doi.org/10.1007/s10985-015-9335-y>
- ANDERSEN, P. K. and GILL, R. D. (1982). Cox’s regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646
- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York. MR1198884 <https://doi.org/10.1007/978-1-4612-4348-9>
- ANDERSON, G., LIMACHER, M., ASSAF, A., BASSFORD, T., BERESFORD, S., BLACK, H., BONDS, D., BRUNNER, R., BRZYSKI, R. et al. (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The Women’s Health Initiative randomized controlled trial. *JAMA* **291** 1701–1712.
- BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452. MR0696057 <https://doi.org/10.1214/aos/1176346151>
- CHLEBOWSKI, R. T., KULLER, L. H., PRENTICE, R. L., STEFANICK, M. L., MANSON, J. E., GASS, M., ARAGAKI, A. K., OCKENE, J. K., LANE, D. S. et al. (2009). Breast cancer after use of estrogen plus progestin in postmenopausal women. *N. Engl. J. Med.* **360** 573–587.
- CHLEBOWSKI, R. T., ANDERSON, G. L., ARAGAKI, A. K., MANSON, J. E., STEFANICK, M. L., PAN, K., BARRINGTON, W., KULLER, L. H., SIMON, M. S. et al. (2020). Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up of the women’s health initiative randomized clinical trials. *JAMA* **324** 369–380. <https://doi.org/10.1001/jama.2020.9482>
- COOK, R. J. and LAWLESS, J. F. (2018). *Multistate Models for the Analysis of Life History Data*. Monographs on Statistics and Applied Probability **158**. CRC Press, Boca Raton, FL. MR3838371 <https://doi.org/10.1201/9781315119731>
- COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758
- COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. MR0400509 <https://doi.org/10.1093/biomet/62.2.269>
- COX, D. R. (1992). Causality: Some statistical aspects. *J. Roy. Statist. Soc. Ser. A* **155** 291–301. MR1157712 <https://doi.org/10.2307/2982962>
- FEIGL, P. and ZELEN, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* 826–838.
- FINE, J. P., JIANG, H. and CHAPPELL, R. (2001). On semi-competing risks data. *Biometrika* **88** 907–919. MR1872209 <https://doi.org/10.1093/biomet/88.4.907>
- GILL, R. D., VAN DER LAAN, M. J. and WELLNER, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 545–597. MR1338452
- HERNÁN, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21** 13–15.
- HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference: What If*. CRC Press/CRC, Boca Raton, FL.
- HSIA, J., LANGER, R., MANSON, J., KULLER, L., JOHNSON, K., HENDRIX, S., PETTINGER, M., HECKBERT, S., GREEP, N. et al. (2006). Women’s health initiative investigators. Conjugated equine estrogens and coronary heart disease: The Women’s Health Initiative. *Arch. Intern. Med.* **166** 357–365.

- KALBFLEISCH, J. D. and PRENTICE, R. L. (1981). Estimation of the average hazard ratio. *Biometrika* **68** 105–112. MR0614947 <https://doi.org/10.1093/biomet/68.1.105>
- KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1924807 <https://doi.org/10.1002/9781118032985>
- LAGAKOS, S. W. and SCHOENFELD, D. A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40** 1037–1048. MR0786178 <https://doi.org/10.2307/2531154>
- MANSON, J. E., HSIA, J., JOHNSON, K. C., ROSSOUW, J. E., ASSAF, A. R., LASSER, N. L., TREVISAN, M., BLACK, H. R., HECKBERT, S. R. et al. (2003). Estrogen plus progestin and the risk of coronary heart disease. *N. Engl. J. Med.* **349** 523–534.
- MANSON, J. E., CHLEBOWSKI, R. T., STEFANICK, M. L., ARAGAKI, A. K., ROSSOUW, J. E., PRENTICE, R. L., ANDERSON, G., HOWARD, B. V., THOMSON, C. A. et al. (2013). Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women’s Health Initiative randomized trials. *J. Am. Med. Assoc.* **310** 1353–1368.
- MANSON, J. E., ARAGAKI, A. K., BASSUK, S. S., CHLEBOWSKI, R. T., ANDERSON, G. L., ROSSOUW, J. E., HOWARD, B. V., THOMSON, C. A., STEFANICK, M. L. et al. (2019). Menopausal estrogen-alone therapy and health outcomes in women with and without bilateral oophorectomy: A randomized trial. *Ann. Intern. Med.* **171** 406–414. <https://doi.org/10.7326/M19-0274>
- MANTEL, N. and HAENZEL, W. (1959). Statistical analysis of data from retrospective studies. *J. Natl. Cancer Inst.* **22** 718–748.
- MARTINUSSEN, T., VANSTEELENDT, S. and ANDERSEN, P. K. (2020). Subtleties in the interpretation of hazard contrasts. *Life-time Data Anal.* **26** 833–855. MR4148449 <https://doi.org/10.1007/s10985-020-09501-5>
- MURPHY, S. A. and SEN, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Process. Appl.* **39** 153–180. MR1135092 [https://doi.org/10.1016/0304-4149\(91\)90039-F](https://doi.org/10.1016/0304-4149(91)90039-F)
- O’QUIGLEY, J. (2008). *Proportional Hazards Regression. Statistics for Biology and Health*. Springer, New York. MR2400249 <https://doi.org/10.1007/978-0-387-68639-4>
- PRENTICE, R. L. and ZHAO, S. (2019). *The Statistical Analysis of Multivariate Failure Time Data: A Marginal Modeling Approach. Monographs on Statistics and Applied Probability* **163**. CRC Press, Boca Raton, FL. MR3966434 <https://doi.org/10.1201/9780429162367>
- PRENTICE, R. L. and ZHAO, S. (2020). Regression models and multivariate life tables. *J. Amer. Statist. Assoc.* <https://doi.org/10.1080/01621459.2020.1713792>
- PRENTICE, R. L., LANGER, R., STEFANICK, M. L., HOWARD, B. V., PETTINGER, M., ANDERSON, G., BARAD, D., CURB, J. D., KOTCHEN, J. et al. (2005). Combined postmenopausal hormone therapy and cardiovascular disease: Toward resolving the discrepancy between observational studies and the women’s health initiative clinical trial. *Am. J. Epidemiol.* **162** 404–414.
- PRENTICE, R. L., LANGER, R. D., STEFANICK, M. L., HOWARD, B. V., PETTINGER, M., ANDERSON, G. L., BARAD, D., CURB, J. D., KOTCHEN, J. et al. (2006). Combined analysis of women’s health initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *Am. J. Epidemiol.* **163** 589–599.
- PRENTICE, R. L., MANSON, J. E., LANGER, R. D., ANDERSON, G. L., PETTINGER, M., JACKSON, R. D., JOHNSON, K. C., KULLER, L. H., LANE, D. S. et al. (2009). Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause. *Am. J. Epidemiol.* **170** 12–23. <https://doi.org/10.1093/aje/kwp115>
- PRENTICE, R. L., ARAGAKI, A. K., CHLEBOWSKI, R. T., ROSSOUW, J. E., ANDERSON, G. L., STEFANICK, M. L., WACTAWSKI-WENDE, J., KULLER, L. H., WALLACE, R. et al. (2020). Randomized trial evaluation of the benefits and risks of menopausal hormone therapy among women 50–59 years of age. *Am. J. Epidemiol.* **190** 365–375.
- PRENTICE, R. L., ARAGAKI, A. K., CHLEBOWSKI, R. T., ZHAO, S., ANDERSON, G. L., ROSSOUW, J. E., WALLACE, R., BANACK, H., SHADYAB, A. H. et al. (2020b). Dual outcome intention-to-treat analyses in the women’s health initiative randomized controlled hormone therapy trials. *Am. J. Epidemiol.* **189** 972–981.
- PRENTICE, R. L. and ARAGAKI, A. K. (2022). Supplement to “Intention-To-Treat Comparisons in Randomized Trials.” <https://doi.org/10.1214/21-STS830SUPP>
- RAVDIN, P. M., CRONIN, K. A., HOWLADER, N., BERG, C. D., CHLEBOWSKI, R. T., FEUER, E. J., EDWARDS, B. K. and BERRY, D. A. (2007). The decrease in breast-cancer incidence in 2003 in the United States. *N. Engl. J. Med.* **356** 1670–1674.
- ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. (Errata in Computers and Mathematics with Applications 1987; 14:917–921).
- 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>
- ROSSOUW, J., ANDERSON, G., PRENTICE, R., LACROIX, A., KOOPERBERG, C., STEFANICK, M., JACKSON, R., BERESFORD, S., HOWARD, B. et al. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women’s Health Initiative randomized controlled trial. *J. Am. Med. Assoc.* **288** 321–333.
- ROTH, J. A., ETZIONI, R., WATERS, T. M., PETTINGER, M., ROSSOUW, J. E., ANDERSON, G. L., CHLEBOWSKI, R. T., MANSON, J. E., HLATKY, M. et al. (2014). Economic return from the Women’s Health Initiative estrogen plus progestin clinical trial: A modeling study. *Ann. Intern. Med.* **160** 594–602. <https://doi.org/10.7326/M13-2348>
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.
- RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152
- SCHEMPER, M., WAKOUNIG, S. and HEINZE, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Stat. Med.* **28** 2473–2489. MR2751535 <https://doi.org/10.1002/sim.3623>
- SPIEKERMAN, C. F. and LIN, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.* **93** 1164–1175. MR1649210 <https://doi.org/10.2307/2669859>
- STRUTHERS, C. A. and KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369. MR0855896 <https://doi.org/10.1093/biomet/73.2.363>
- UNO, H., WITTES, J., FU, H., SOLOMON, S. D., CLAGGETT, B., TIAN, L., CAI, T., PFEFFER, M. A., EVANS, S. R. et al. (2015). Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Ann. Intern. Med.* **163** 127–134.
- WEI, L. J., LIN, D. Y. and WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84** 1065–1073. MR1134494

ZHAO, L., CLAGGETT, B., TIAN, L., UNO, H., PFEFFER, M. A.,
SOLOMON, S. D., TRIPPA, L. and WEI, L. J. (2016). On the re-

stricted mean survival time curve in survival analysis. *Biometrics*
72 215–221. MR3500590 <https://doi.org/10.1111/biom.12384>

Modeling the Occurrence of Events Subject to a Reporting Delay via an EM Algorithm

Roel Verbelen, Katrien Antonio, Gerda Claeskens and Jonas Crevecoeur

Abstract. A delay between the occurrence and the reporting of events often has practical implications such as for the amount of capital to hold for insurance companies, or for taking preventive actions in case of infectious diseases. The accurate estimation of the number of incurred but not (yet) reported events forms an essential part of properly dealing with this phenomenon. We review the current practice for analysing such data and we present a flexible regression framework to jointly estimate the occurrence and reporting of events. By linking this setting to an incomplete data problem, estimation is performed via an expectation-maximization algorithm. The resulting method is elegant, easy to understand and implement, and provides refined insights in the nowcasts. The proposed methodology is applied to a European general liability portfolio in insurance.

Key words and phrases: EM algorithm, nowcasting, Poisson regression model, reporting delay.

REFERENCES

- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)* (B. Petrov and F. Csáki, eds.) 267–281. MR0483125
- ANTONIO, K. and PLAT, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scand. Actuar. J.* **7** 649–669. MR3231366 <https://doi.org/10.1080/03461238.2012.755938>
- AVANZI, B., WONG, B. and YANG, X. (2016). A micro-level claim count model with overdispersion and reporting delays. *Insurance Math. Econom.* **71** 1–14. MR3578869 <https://doi.org/10.1016/j.insmatheco.2016.07.002>
- BACCHETTI, P., SEGAL, M. R. and JEWELL, N. P. (1993). Backcalculation of HIV infection rates. *Statist. Sci.* **8** 82–101.
- BADESCU, A. L., LIN, X. S. and TANG, D. (2016). A marked Cox model for the number of IBNR claims: Theory. *Insurance Math. Econom.* **69** 29–37. MR3515874 <https://doi.org/10.1016/j.insmatheco.2016.03.016>
- BADESCU, A. L., CHEN, T., LIN, X. S. and TANG, D. (2019). A marked Cox model for the number of IBNR claims: Estimation and application. *Astin Bull.* **49** 709–739. MR4010091 <https://doi.org/10.1017/asb.2019.15>
- BARRETO-SOUZA, W. and SIMAS, A. B. (2017). Improving estimation for beta regression models via EM-algorithm and related diagnostic tools. *J. Stat. Comput. Simul.* **87** 2847–2867. MR3679426 <https://doi.org/10.1080/00949655.2017.1350679>
- BASTOS, L. S., ECONOMOU, T., GOMES, M. F. C., VILLELA, D. A. M., COELHO, F. C., CRUZ, O. G., STONER, O., BAILEY, T. and CODEÇO, C. T. (2019). A modelling approach for correcting reporting delays in disease surveillance data. *Stat. Med.* **38** 4363–4377. MR4002436 <https://doi.org/10.1002/sim.8303>
- BATES, D. and MAECHLER, M. (2015). MatrixModels: Modelling with Sparse And Dense Matrices. R package version 0.4-1.
- BECKER, N. G. and CUI, J.-S. (1997). Estimating a delay distribution from incomplete data, with application to reporting lags for AIDS cases. *Stat. Med.* **16** 2339–2347.
- BELLOCCO, R. and MARSCHNER, I. C. (2000). Joint analysis of HIV and AIDS surveillance data in back-calculation. *Stat. Med.* **19** 297–311.
- BROOKMEYER, R. and GAIL, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Amer. Statist. Assoc.* **83** 301–308.
- CAVANAUGH, J. E. and SHUMWAY, R. H. (1998). An Akaike information criterion for model selection in the presence of incomplete data. *J. Statist. Plann. Inference* **67** 45–65. MR1624662 [https://doi.org/10.1016/S0378-3758\(97\)00115-8](https://doi.org/10.1016/S0378-3758(97)00115-8)
- CLAESKENS, G. and CONSENTINO, F. (2008). Variable selection with incomplete covariate data. *Biometrics* **64** 1062–1069. MR2522253

Roel Verbelen is Post-doctoral Researcher, Faculty of Economics and Business, KU Leuven, Belgium, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: roel.verbelen@kuleuven.be). Katrien Antonio is Professor, Faculty of Economics and Business, KU Leuven, Belgium, Faculty of Economics and Business, University of Amsterdam, The Netherlands, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: katrien.antonio@kuleuven.be). Gerda Claeskens is Professor, Faculty of Economics and Business, KU Leuven, Belgium, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium (e-mail: gerda.claeskens@kuleuven.be). Jonas Crevecoeur is Post-doctoral Researcher, Faculty of Economics and Business, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: jonas.crevecoeur@kuleuven.be).

- <https://doi.org/10.1111/j.1541-0420.2008.01003.x>
- CLAESKENS, G. and HJORT, N. L. (2008). *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics **27**. Cambridge Univ. Press, Cambridge. MR2431297 <https://doi.org/10.1017/CBO9780511790485>
- COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758
- CREVECOEUR, J., ANTONIO, K. and VERBELEN, R. (2019). Modeling the number of hidden events subject to observation delay. *European J. Oper. Res.* **277** 930–944. MR3958543 <https://doi.org/10.1016/j.ejor.2019.02.044>
- 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. With discussion. MR0501537
- DIRICK, L., CLAESKENS, G. and BAESSENS, B. (2015). An Akaike information criterion for multiple event mixture cure models. *European J. Oper. Res.* **241** 449–457. MR3276774 <https://doi.org/10.1016/j.ejor.2014.08.038>
- DONKER, T., VAN BOVEN, M., VAN BALLEGOOIJEN, W., VAN’T KLOOSTER, T., WIELDERS, C. and WALLINGA, J. (2011). Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. *Eur. J. Epidemiol.* **26** 195–201.
- ENGLAND, P. D. and VERRALL, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuar. J.* **8** 443–518.
- FARRINGTON, C. P., ANDREWS, N. J., BEALE, A. D. and CATCHPOLE, M. A. (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *J. Roy. Statist. Soc. Ser. A* **159** 547–563. MR1413665 <https://doi.org/10.2307/2983331>
- GREENE, S. K., MCGOUGH, S. F., CULP, G. M., GRAF, L. E., LIPSITCH, M., MENZIES, N. A. and KAHN, R. (2021). Nowcasting for real-time COVID-19 tracking in New York City: An evaluation using reportable disease data from early in the pandemic. *JMIR Public Health Surveill* **7** e25538.
- GÜNTHER, F., BENDER, A., KATZ, K., KÜCHENHOFF, H. and HÖHLE, M. (2021). Nowcasting the COVID-19 pandemic in Bavaria. *Biom. J.* **63** 490–502. MR4226590 <https://doi.org/10.1002/bimj.202000112>
- HAASTRUP, S. and ARJAS, E. (1996). Claims reserving in continuous time: A nonparametric Bayesian approach. *Astin Bull.* **26** 139–164.
- HACHEMEISTER, C. A. and STANARD, J. N. (1975). IBNR claims count estimation with static lag functions. In *Spring Meeting of the Casualty Actuarial Society*.
- HARRIS, J. E. (1990). Reporting delays and the incidence of AIDS. *J. Amer. Statist. Assoc.* **85** 915–924.
- HIABU, M., MAMMEN, E., MARTÍNEZ-MIRANDA, M. D. and NIELSEN, J. P. (2021). Smooth backfitting of proportional hazards with multiplicative components. *J. Amer. Statist. Assoc.* To appear. <https://doi.org/10.1080/01621459.2020.1753520>
- HÖHLE, M. and AN DER HEIDEN, M. (2014). Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics* **70** 993–1002. MR3295760 <https://doi.org/10.1111/biom.12194>
- JEWELL, W. S. (1989). Predicting IBNYR events and delays: I. Continuous time. *Astin Bull.* **19** 25–55.
- JUNG, R. C. and TREMAYNE, A. R. (2011). Useful models for time series of counts or simply wrong ones? *ASTA Adv. Stat. Anal.* **95** 59–91. MR2775582 <https://doi.org/10.1007/s10182-010-0139-9>
- KALBFLEISCH, J. D. and LAWLESS, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84** 360–372. MR1010328
- KALBFLEISCH, J. D. and LAWLESS, J. F. (1991). Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statist. Sinica* **1** 19–32.
- KALBFLEISCH, J., LAWLESS, J. and ROBINSON, J. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics* **33** 273–285.
- LAGAKOS, S. W., BARRAJ, L. M. and DE GRUTTOLA, V. (1988). Nonparametric analysis of truncated survival data, with applications to AIDS. *Biometrika* **75** 515–523. MR0967591 <https://doi.org/10.1093/biomet/75.3.515>
- LAWLESS, J. F. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events. *Canad. J. Statist.* **22** 15–31. MR1271442 <https://doi.org/10.2307/3315820>
- LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233. MR0676213
- MACK, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bull.* **21** 93–109.
- MACK, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bull.* **23** 213–225.
- MARTÍNEZ MIRANDA, M. D., NIELSEN, J. P. and SPERLICH, S. VERRALL, R. (2013). Continuous chain ladder: Reformulating and generalizing a classical insurance problem. *Expert Syst. Appl.* **40** 5588–5603.
- MCGOUGH, S. F., JOHANSSON, M. A., LIPSITCH, M. and MENZIES, N. A. (2020). Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Comput. Biol.* **16** 1–20.
- MENG, X.-L. and RUBIN, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *J. Amer. Statist. Assoc.* **86** 899–909.
- NORBERG, R. (1993). Prediction of outstanding liabilities in non-life insurance. *Astin Bull.* **23** 95–115.
- NORBERG, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *Astin Bull.* **29** 5–27.
- NOUFAILY, A., GHEBREMICHAEL-WELDESELASSIE, Y., ENKI, D. G., ANDREWS, N., CHARLETT, A. and FARRINGTON, P. (2015). Modelling reporting delays for outbreak detection in infectious disease data. *J. Roy. Statist. Soc. Ser. A* **178** 205–222. MR3291768 <https://doi.org/10.1111/rssa.12055>
- NOUFAILY, A., FARRINGTON, P., GARTHWAITE, P., ENKI, D. G., ANDREWS, N. and CHARLETT, A. (2016). Detection of infectious disease outbreaks from laboratory data with reporting delays. *J. Amer. Statist. Assoc.* **111** 488–499. MR3538681 <https://doi.org/10.1080/01621459.2015.1119047>
- OAKES, D. (1999). Direct calculation of the information matrix via the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 479–482. MR1680298 <https://doi.org/10.1111/1467-9868.00188>
- PAGANO, M., TU, X. M., GRUTTOLA, V. D. and MAWHINNEY, S. (1994). Regression analysis of censored and truncated data: Estimating reporting- delay distributions and aids incidence from surveillance data. *Biometrics* **50** 1203–1214.
- RENSHAW, A. E. and VERRALL, R. J. (1998). A stochastic model underlying the chain-ladder technique. *British Actuar. J.* **4** 903–923.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014
- SELLERO, C. S., FERNÁNDEZ, E. V., MANTEIGA, W., OTERO, X. L., HERVADA, X., FERNÁNDEZ, E. and TABOADA, X. A. (1996). Reporting delay: A review with simulation study and application to Spanish data. *Stat. Med.* **15** 305–321.
- TABNAK, F., MÜLLER, H., WANG, J., CHIOU, J.-M. and SUN, R. (2000). A change-point model for reporting delays under change of AIDS case definition. *Eur. J. Epidemiol.* **16** 1135–1141.
- TAYLOR, G. (2000). *Loss Reserving: An Actuarial Perspective*. Kluwer Academic, Norwell.
- VAN DE KASSTEELE, J., EILERS, P. and WALLINGA, J. (2019). Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained P-spline smoothing. *Epidemiology* **30** 737–745.

- VERBELEN, R., ANTONIO, K., CLAESKENS, G. and CREVE-COEUR, J. (2022). Supplement to “Modeling the occurrence of events subject to a reporting delay via an EM algorithm.” <https://doi.org/10.1214/21-ST831SUPP>
- VERRALL, R. J. and WÜTHRICH, M. V. (2016). Understanding reporting delay in general insurance. *Risks* **4** 25.
- WAHL, F. (2019). Explicit moments for a class of micro-models in non-life insurance. *Insurance Math. Econom.* **89** 140–156. MR4020515 <https://doi.org/10.1016/j.insmatheco.2019.10.001>
- WU, S. (2013). A review on coarse warranty data and analysis. *Reliab. Eng. Syst. Saf.* **114** 1–11.
- WÜTHRICH, M. V. and MERZ, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. *Wiley Finance* **435**. Wiley.
- WÜTHRICH, M. V. and MERZ, M. (2015). *Stochastic Claims Reserving Manual: Advances in Dynamic Modeling*. Swiss Finance Institute Research Paper No. 15-34.
- ZHU, H., LEE, S.-Y., WEI, B.-C. and ZHOU, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88** 727–737. MR1859405 <https://doi.org/10.1093/biomet/88.3.727>

A Unifying Framework of High-Dimensional Sparse Estimation with Difference-of-Convex (DC) Regularizations

Shanshan Cao, Xiaoming Huo and Jong-Shi Pang

Abstract. Under the linear regression framework, we study the variable selection problem when the underlying model is assumed to have a small number of nonzero coefficients. Nonconvex penalties in specific forms are well studied in the literature for sparse estimation. Recent work pointed out that nearly all existing nonconvex penalties can be represented as difference-of-convex (DC) functions, which are the difference of two convex functions, while itself may not be convex. There is a large existing literature on optimization problems when their objectives and/or constraints involve DC functions. Efficient numerical solutions have been proposed. Under the DC framework, directional-stationary (d-stationary) solutions are considered, and they are usually not unique. In this paper, we show that under some mild conditions, a certain subset of d-stationary solutions in an optimization problem (with a DC objective) has some ideal statistical properties: namely, asymptotic estimation consistency, asymptotic model selection consistency, asymptotic efficiency. Our assumptions are either weaker than or comparable with those conditions that have been adopted in other existing works. This work shows that DC is a nice framework to offer a unified approach to these existing works where nonconvex penalties are involved. Our work bridges the communities of optimization and statistics.

Key words and phrases: (Generalized) linear regression, high-dimensional sparse estimation, nonconvex regularization, difference of convex (DC) functions, DC algorithms, asymptotic optimality, model selection consistency.

REFERENCES

- [1] AHN, M., PANG, J.-S. and XIN, J. (2017). Difference-of-convex learning: Directional stationarity, optimality, and sparsity. *SIAM J. Optim.* **27** 1637–1665. MR3683697 <https://doi.org/10.1137/16M1084754>
- [2] ALEKSANDROV, A. (1950). Surfaces represented as a difference of two convex functions, Russian Acad. Sci. In *Dokl. Math* **1**.
- [3] AN, L. T. H. and TAO, P. D. (2005). The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Ann. Oper. Res.* **133** 23–46. MR2119311 <https://doi.org/10.1007/s10479-004-5022-1>
- [4] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 <https://doi.org/10.1214/08-AOS620>
- [5] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer, Heidelberg. MR2807761 <https://doi.org/10.1007/978-3-642-20192-9>
- [6] CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. MR2382644 <https://doi.org/10.1214/009053606000001523>
- [7] CAO, S., XIAOMING, H. and PANG, J.-S. (2022). Supplement to “A unifying framework of high-dimensional sparse estimation with Difference-of-Convex (DC) regularizations.” <https://doi.org/10.1214/21-STS832SUPP>
- [8] CHEN, S. and DONOHO, D. L. (1995). Examples of basis pursuit. In *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation* 564–574. International Society for Optics and Photonics.
- [9] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 <https://doi.org/10.1214/009053604000000067>
- [10] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1861973 <https://doi.org/10.1198/016214501000000763>

- Statist. Assoc.* **96** 1348–1360. MR1946581 <https://doi.org/10.1198/016214501753382273>
- [11] FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194 <https://doi.org/10.1214/009053604000000256>
- [12] FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. MR3210988 <https://doi.org/10.1214/13-AOS1198>
- [13] HARTMAN, P. (1959). On functions representable as a difference of convex functions. *Pacific J. Math.* **9** 707–713. MR0110773
- [14] HIRIART-URRUTY, J.-B. (1985). Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. In *Convexity and Duality in Optimization (Groningen, 1984). Lecture Notes in Econom. and Math. Systems* **256** 37–70. Springer, Berlin. MR0873269 https://doi.org/10.1007/978-3-642-45610-7_3
- [15] HORST, R. and THOAI, N. V. (1999). DC programming: Overview. *J. Optim. Theory Appl.* **103** 1–43. MR1715016 <https://doi.org/10.1023/A:1021765131316>
- [16] HUANG, J., MA, S. and ZHANG, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statist. Sinica* **18** 1603–1618. MR2469326
- [17] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 <https://doi.org/10.1109/TIT.2014.2343629>
- [18] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
- [19] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875
- [20] LOH, P.-L. and WAINWRIGHT, M. J. (2013). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems* 476–484.
- [21] LV, J. and FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37** 3498–3528. MR2549567 <https://doi.org/10.1214/09-AOS683>
- [22] MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. MR2894769 <https://doi.org/10.1198/jasa.2011.tm09738>
- [23] NOUIEHED, M., PANG, J.-S. and RAZAVIYAYN, M. (2017). On the pervasiveness of difference-convexity in optimization and statistics. arXiv preprint. Available at [arXiv:1704.03535](https://arxiv.org/abs/1704.03535).
- [24] PANG, J.-S., RAZAVIYAYN, M. and ALVARADO, A. (2017). Computing B-stationary points of nonsmooth DC programs. *Math. Oper. Res.* **42** 95–118. MR3622251 <https://doi.org/10.1287/moor.2016.0795>
- [25] ROCKAFELLAR, R. T. (2015). *Convex Analysis*. Princeton University Press, Princeton, NJ.
- [26] SRIPERUMBUDUR, B. K. and LANCKRIET, G. R. G. (2012). A proof of convergence of the concave-convex procedure using Zangwill’s theory. *Neural Comput.* **24** 1391–1407. MR2962072 https://doi.org/10.1162/NECO_a_00283
- [27] TAO, P. D. and AN, L. T. H. (1997). Convex analysis approach to d.c. programming: Theory, algorithms and applications. *Acta Math. Vietnam.* **22** 289–355. MR1479751
- [28] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
- [29] TUY, H. (1987). Global minimization of a difference of two convex functions. *Nonlinear Analysis and Optimization* 150–182.
- [30] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 <https://doi.org/10.1214/14-AOS1221>
- [31] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 <https://doi.org/10.1109/TIT.2009.2016018>
- [32] WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. MR3127873 <https://doi.org/10.1214/13-AOS1159>
- [33] WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201. MR3269977 <https://doi.org/10.1214/14-AOS1238>
- [34] YUILLE, A. L. and RANGARAJAN, A. (2003). The concave-convex procedure. *Neural Comput.* **15** 915–936.
- [35] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 <https://doi.org/10.1214/09-AOS729>
- [36] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 <https://doi.org/10.1111/rssb.12026>
- [37] ZHANG, S. and XIN, J. (2018). Minimization of transformed L_1 penalty: Theory, difference of convex function algorithm, and robust application in compressed sensing. *Math. Program.* **169** 307–336. MR3785680 <https://doi.org/10.1007/s10107-018-1236-x>
- [38] ZHANG, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.* **11** 1081–1107. MR2629825
- [39] ZHANG, T. (2013). Multi-stage convex relaxation for feature selection. *Bernoulli* **19** 2277–2293. MR3160554 <https://doi.org/10.3150/12-BEJ452>
- [40] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449
- [41] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 <https://doi.org/10.1198/016214506000000735>
- [42] ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 <https://doi.org/10.1214/009053607000000802>

Challenges in Markov Chain Monte Carlo for Bayesian Neural Networks

Theodore Papamarkou, Jacob Hinkle, M. Todd Young and David Womble

Abstract. Markov chain Monte Carlo (MCMC) methods have not been broadly adopted in Bayesian neural networks (BNNs). This paper initially reviews the main challenges in sampling from the parameter posterior of a neural network via MCMC. Such challenges culminate to lack of convergence to the parameter posterior. Nevertheless, this paper shows that a non-converged Markov chain, generated via MCMC sampling from the parameter space of a neural network, can yield via Bayesian marginalization a valuable posterior predictive distribution of the output of the neural network. Classification examples based on multilayer perceptrons showcase highly accurate posterior predictive distributions. The postulate of limited scope for MCMC developments in BNNs is partially valid; an asymptotically exact parameter posterior seems less plausible, yet an accurate posterior predictive distribution is a tenable research avenue.

Key words and phrases: Bayesian inference, Bayesian neural networks, convergence diagnostics, Markov chain Monte Carlo, posterior predictive distribution.

REFERENCES

- ANDRIEU, C., DE FREITAS, J. F. G. and DOUCET, A. (1999). Sequential Bayesian estimation and model selection applied to neural networks.
- ANDRIEU, C., DE FREITAS, N. and DOUCET, A. (2000). Reversible jump MCMC simulated annealing for neural networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence* 11–18.
- BADRINARAYANAN, V., MISHRA, B. and CIPOLLA, R. (2015). Symmetry-invariant optimization in deep networks. ArXiv.
- BENNETT, J. E., RACINE-POON, A. and WAKEFIELD, J. C. (1996). MCMC for nonlinear hierarchical models. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson and D. Spiegelhalter, eds.) 339–358. Chapman and Hall/CRC.
- BERNARDO, J.-M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **41** 113–147. With discussion. MR0547240
- BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 <https://doi.org/10.1080/01621459.2017.1285773>
- BLIER, L. and OLLIVIER, Y. (2018). The description length of deep learning models. In *Advances in Neural Information Processing Systems* **31**.
- BREA, J., SIMSEK, B., ILLING, B. and GERSTNER, W. (2019). Weight-space symmetry in deep networks gives rise to permutation saddles, connected by equal-loss valleys across the loss landscape. ArXiv.
- BROOKS, S. P. and GELMAN, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Statist.* **7** 434–455. MR1665662 <https://doi.org/10.2307/1390675>
- CANNON, A., COBB, G., HARTLAUB, B., LEGLER, J., LOCK, R., MOORE, T., ROSSMAN, A. and WITMER, J. (2019). Stat2Data: Datasets for Stat2. R package version 2.0.0.
- CHEN, T., FOX, E. and GUESTRIN, C. (2014). Stochastic gradient Hamiltonian Monte Carlo. In *Proceedings of the 31st International Conference on Machine Learning* **32** 1683–1691.
- CHEN, A. M., LU, H. and HECHT-NIELSEN, R. (1993). On the geometry of feedforward neural network error surfaces. *Neural Comput.* **5** 910–927.
- CHEN, W. Y., BARP, A., BRIOL, F.-X., GORHAM, J., GIROLAMI, M., MACKEY, L. and OATES, C. (2019). Stein point Markov chain Monte Carlo. In *Proceedings of the 36th International Conference on Machine Learning* **97** 1011–1021.
- CHOLLET, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1251–1258.

Theodore Papamarkou is a Reader in the Mathematics of Data Science at the Department of Mathematics, The University of Manchester, Manchester, UK (e-mail: theodore.papamarkou@manchester.ac.uk). Jacob Hinkle is a Research Scientist at the Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA (e-mail: hinklejd@ornl.gov). M. Todd Young is a Machine Learning Engineer at the Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA (e-mail: youngmt1@ornl.gov). David Womble is the Artificial Intelligence (AI) Program Director of the Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA (e-mail: womblede@ornl.gov).

- CHWIALKOWSKI, K., STRATHMANN, H. and GRETTON, A. (2016). A kernel test of goodness of fit. In *Proceedings of the 33rd International Conference on Machine Learning* **48** 2606–2615.
- COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91** 883–904. MR1395755 <https://doi.org/10.2307/2291683>
- CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2** 303–314. MR1015670 <https://doi.org/10.1007/BF02551274>
- DAI, N. and JONES, G. L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *J. Multivariate Anal.* **159** 184–199. MR3668555 <https://doi.org/10.1016/j.jmva.2017.05.009>
- DANIELS, M. J. and KASS, R. E. (1998). A note on first-stage approximation in two-stage hierarchical models. *Sankhya, Ser. B* **60** 19–30. MR1717073
- DE FREITAS, N. (1999). Bayesian methods for neural networks. PhD thesis, Univ. Cambridge.
- DE FREITAS, N., ANDRIEU, C., HØJEN-SØRENSEN, P., NIRANJAN, M. and GEE, A. (2001). Sequential Monte Carlo methods for neural networks. In *Sequential Monte Carlo Methods in Practice. Stat. Eng. Inf. Sci.* 359–379. Springer, New York. MR1847800 https://doi.org/10.1007/978-1-4757-3437-9_17
- DE SA, C., CHEN, V. and WONG, W. (2018). Minibatch Gibbs sampling on large graphical models. In *Proceedings of the 35th International Conference on Machine Learning* **80** 1165–1173.
- DUPUY, C. and BACH, F. (2017). Online but accurate inference for latent variable models with local Gibbs sampling. *J. Mach. Learn. Res.* **18** Paper No. 126, 45. MR3763760
- ENSIGN, D., NEVILLE, S., PAUL, A. and VENKATASUBRAMANIAN, S. (2017). The complexity of explaining neural networks through (group) invariants. In *International Conference on Algorithmic Learning Theory. Proc. Mach. Learn. Res. (PMLR)* **76** 19. MR3776700
- ESMAEILI, B., WU, H., JAIN, S., BOZKURT, A., SIDDHARTH, N., PAIGE, B., BROOKS, D. H., DY, J. and VAN DE MEENT, J.-W. (2019). Structured disentangled representations. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* **89** 2525–2534.
- FREEMAN, I., ROESE-KOERNER, L. and KUMMERT, A. (2018). Effnet: An efficient structure for convolutional neural networks. In *25th IEEE International Conference on Image Processing* 6–10.
- FRIEL, N. and PETTITT, A. N. (2008). Marginal likelihood estimation via power posteriors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 589–607. MR2420416 <https://doi.org/10.1111/j.1467-9868.2007.00650.x>
- GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.
- GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR2027492
- GILKS, W. R. and ROBERTS, G. O. (1996). Strategies for improving MCMC. In *Markov Chain Monte Carlo in Practice* (W. R. Gilks, S. Richardson and D. Spiegelhalter, eds.) 89–114. Chapman and Hall/CRC.
- GIORDANO, R. J., BRODERICK, T. and JORDAN, M. I. (2015). Linear response methods for accurate covariance estimates from mean field variational Bayes. In *Advances in Neural Information Processing Systems* **28** 1441–1449.
- GONG, L. and FLEGAL, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *J. Comput. Graph. Statist.* **25** 684–700. MR3533633 <https://doi.org/10.1080/10618600.2015.1044092>
- GONG, W., LI, Y. and HERNÁNDEZ-LOBATO, J. M. (2019). Meta-learning for stochastic gradient MCMC. In *International Conference on Learning Representations*.
- GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773
- GRAF, S. and LUSCHGY, H. (2007). *Foundations of Quantization for Probability Distributions*. Springer. MR1764176 <https://doi.org/10.1007/BFb0103945>
- GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716
- GU, S. S., GHAHRAMANI, Z. and TURNER, R. E. (2015). Neural adaptive sequential Monte Carlo. In *Advances in Neural Information Processing Systems* **28** 2629–2637.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed. Springer. MR2722294 <https://doi.org/10.1007/978-0-387-84858-7>
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 <https://doi.org/10.1093/biomet/57.1.97>
- HECHT-NIELSEN, R. (1990). On the algebraic structure of feedforward network weight spaces. In *Advanced Neural Computers* 129–135.
- HORNIK, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4** 251–257.
- HORST, A. M., HILL, A. P. and GORMAN, K. B. (2020). palmerpenquins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0.
- HOWARD, A. G., ZHU, M., CHEN, B., KALENICHENKO, D., WANG, W., WEYAND, T., ANDRETTA, M. and ADAM, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. ArXiv.
- HU, S. X., ZAGORUYKO, S. and KOMODAKIS, N. (2019). Exploring weight symmetry in deep neural networks. *Comput. Vis. Image Underst.* **187** 102786.
- HUANG, C.-W., SANKARAN, K., DHEKANE, E., LACOSTE, A. and COURVILLE, A. (2019). Hierarchical importance weighted autoencoders. In *Proceedings of the 36th International Conference on Machine Learning* **97** 2869–2878.
- IANDOLA, F. N., HAN, S., MOSKEWICZ, M. W., ASHRAF, K., DALLY, W. J. and KEUTZER, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. ArXiv.
- IZMAILOV, P., MADDOX, W. J., KIRICHENKO, P., GARIPPOV, T., VETROV, D. and WILSON, A. G. (2020). Subspace inference for Bayesian deep learning. In *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference* **115** 1169–1179.
- JARRETT, K., KAVUKCUOGLU, K., RANZATO, M. and LECUN, Y. (2009). What is the best multi-stage architecture for object recognition? In *IEEE 12th International Conference on Computer Vision* 2146–2153.
- JAYNES, E. T. (1968). Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* **4** 227–241.
- JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford. MR0187257
- JOHNDROW, J. E., PILLAI, N. S. and SMITH, A. (2020). No free lunch for approximate MCMC. ArXiv.
- KASS, R. E., CARLIN, B. P., GELMAN, A. and NEAL, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *Amer. Statist.* **52** 93–100. MR1628427 <https://doi.org/10.2307/2685466>

- KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 25 1097–1105.
- LEE, H. K. H. (2000). Consistency of posterior distributions for neural networks. *Neural Netw.* **13** 629–642.
- LEE, H. K. H. (2003). A noninformative prior for neural networks. *Mach. Learn.* **50** 197–212.
- LEE, H. K. H. (2004). Priors for neural networks. In *Classification, Clustering, and Data Mining Applications. Stud. Classification Data Anal. Knowledge Organ.* 141–150. Springer, Berlin. [MR2113604](https://doi.org/10.1007/978-3-540-24848-2_11)
- LEE, H. K. (2005). Neural networks and default priors. In *Proceedings of the American Statistical Association, Section on Bayesian Statistical Science.*
- LEE, H. K. H. (2007). Default priors for neural network classification. *J. Classification* **24** 53–70. [MR2321027](https://doi.org/10.1007/s00357-007-0001-2) <https://doi.org/10.1007/s00357-007-0001-2>
- LIU, M. (2020). *Nonconvex Min-Max Optimization in Deep Learning: Algorithms and Applications*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—The University of Iowa. [MR4158092](https://doi.org/10.1007/978-1-4939-9848-2)
- LU, Z., PU, H., WANG, F., HU, Z. and WANG, L. (2017). The expressive power of neural networks: A view from the width. In *Advances in Neural Information Processing Systems* 30 6231–6239.
- MA, Y.-A., FOTI, N. J. and FOX, E. B. (2017). Stochastic gradient MCMC methods for hidden Markov models. In *Proceedings of the 34th International Conference on Machine Learning* **70** 2265–2274.
- MACKAY, D. J. (1995). Developments in probabilistic modelling with neural networks—ensemble learning. In *Neural Networks: Artificial Intelligence and Industrial Applications* 191–198.
- MADDISON, C. J., HUANG, A., SUTSKEVER, I. and SILVER, D. (2015). Move evaluation in Go using deep convolutional neural networks. In *International Conference on Learning Representations*.
- MANDT, S., HOFFMAN, M. D. and BLEI, D. M. (2017). Stochastic gradient descent as approximate Bayesian inference. *J. Mach. Learn. Res.* **18** Paper No. 134, 35. [MR3763768](https://doi.org/10.1162/jmlr.2017.18.1.134)
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- MINSKY, M. L. and PAPERT, S. A. (1988). *Perceptrons: Expanded Edition*. MIT press.
- MOORE, D. A. (2016). Symmetrized variational inference. In *NIPS Workshop on Advances in Approximate Bayesian Inference*.
- NAIR, V. and HINTON, G. E. (2009). 3D object recognition with deep belief nets. In *Advances in Neural Information Processing Systems* 22 1339–1347.
- NALISNICK, E. T. (2018). On priors for Bayesian neural networks. Ph.D. thesis, UC Irvine.
- 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](https://doi.org/10.1002/9781118131218.ch11)
- NEMETH, C. and SHERLOCK, C. (2018). Merging MCMC subposteriors through Gaussian-process approximations. *Bayesian Anal.* **13** 507–530. [MR3780433](https://doi.org/10.1214/17-BA1063) <https://doi.org/10.1214/17-BA1063>
- NWANKPA, C., IJOMAH, W., GACHAGAN, A. and MARSHALL, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. ArXiv.
- ONG, V. M.-H., NOTT, D. J. and SMITH, M. S. (2018). Gaussian variational approximation with a factor covariance structure. *J. Comput. Graph. Statist.* **27** 465–478. [MR3863750](https://doi.org/10.1080/10618600.2017.1390472) <https://doi.org/10.1080/10618600.2017.1390472>
- PEARCE, T., ZAKI, M., BRINTRUP, A. and NEELY, A. (2019). Expressive priors in Bayesian neural networks: Kernel combinations and periodic functions. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*.
- POLSON, N. G. and SOKOLOV, V. (2017). Deep learning: A Bayesian perspective. *Bayesian Anal.* **12** 1275–1304. [MR3724986](https://doi.org/10.1214/17-BA1082) <https://doi.org/10.1214/17-BA1082>
- POURZANJANI, A. A., JIANG, R. M. and PETZOLD, L. R. (2017). Improving the identifiability of neural networks for Bayesian inference. In *NIPS Workshop on Bayesian Deep Learning*.
- QUIROZ, M., KOHN, R., VILLANI, M. and TRAN, M.-N. (2019). Speeding up MCMC by efficient data subsampling. *J. Amer. Statist. Assoc.* **114** 831–843. [MR3963184](https://doi.org/10.1080/01621459.2018.1448827) <https://doi.org/10.1080/01621459.2018.1448827>
- RANGANATH, R., TRAN, D. and BLEI, D. (2016). Hierarchical variational models. In *Proceedings of the 33rd International Conference on Machine Learning* **48** 324–333.
- ROBERT, C. P., ELVIRA, V., TAWN, N. and WU, C. (2018). Accelerating MCMC algorithms. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1435, 14. [MR3850448](https://doi.org/10.1002/wics.1435) <https://doi.org/10.1002/wics.1435>
- ROSENBLATT, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65** 386.
- RUDOLF, D. and SCHWEIZER, N. (2018). Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli* **24** 2610–2639. [MR3779696](https://doi.org/10.3150/17-BEJ938) <https://doi.org/10.3150/17-BEJ938>
- SARGENT, D. J., HODGES, J. S. and CARLIN, B. P. (2000). Structured Markov chain Monte Carlo. *J. Comput. Graph. Statist.* **9** 217–234. [MR1823802](https://doi.org/10.2307/1390651) <https://doi.org/10.2307/1390651>
- SEITA, D., PAN, X., CHEN, H. and CANNY, J. (2018). An efficient minibatch acceptance test for Metropolis–Hastings. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* 5359–5363.
- SEN, D., PAPAMARKOU, T. and DUNSON, D. (2020). Bayesian neural networks and dimensionality reduction. ArXiv.
- SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T. G. and SØRBYE, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statist. Sci.* **32** 1–28. [MR3634300](https://doi.org/10.1214/16-STSS76) <https://doi.org/10.1214/16-STSS76>
- SMITH, J. W., EVERHART, J., DICKSON, W., KNOWLER, W. and JOHANNES, R. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* 261.
- STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. [MR1796293](https://doi.org/10.1111/1467-9868.00265) <https://doi.org/10.1111/1467-9868.00265>
- TITSIAS, M. K. and RUIZ, F. (2019). Unbiased implicit variational inference. In *Proceedings of Machine Learning Research* **89** 167–176.
- TITTERINGTON, D. M. (2004). Bayesian methods for neural networks and related models. *Statist. Sci.* **19** 128–139. [MR2082152](https://doi.org/10.1214/08834230400000099) <https://doi.org/10.1214/08834230400000099>
- TRUONG, T.-D., NGUYEN, V.-T. and TRAN, M.-T. (2018). Lightweight deep convolutional network for tiny object recognition. In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods* 675–682.
- VATS, D. and FLEGAL, J. M. (2018). Lugsail lag windows and their application to MCMC. ArXiv.
- VATS, D., FLEGAL, J. M. and JONES, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. [MR3949306](https://doi.org/10.1093/biomet/asz002) <https://doi.org/10.1093/biomet/asz002>
- VATS, D. and KNUDSON, C. (2018). Revisiting the Gelman–Rubin diagnostic. ArXiv.

- VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. and BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718. [MR4298989 https://doi.org/10.1214/20-ba1221](https://doi.org/10.1214/20-ba1221)
- VLADIMIROVA, M., VERBEEK, J., MESEJO, P. and ARBEL, J. (2019). Understanding priors in Bayesian neural networks at the unit level. In *Proceedings of the 36th International Conference on Machine Learning* **97** 6458–6467.
- WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* 681–688.
- WILLIAMS, P. M. (1995). Bayesian regularization and pruning using a Laplace prior. *Neural Comput.* **7** 117–143.
- WILLIAMS, C. K. I. (2000). An MCMC approach to hierarchical mixture modelling. In *Advances in Neural Information Processing Systems* 12 680–686.
- WILSON, A. G. and IZMAILOV, P. (2020). Bayesian deep learning and a probabilistic perspective of generalization. arXiv.
- ZHANG, G., SUN, S., DUVENAUD, D. and GROSSE, R. (2018a). Noisy natural gradient as variational inference. In *Proceedings of the 35th International Conference on Machine Learning* **80** 5852–5861.
- ZHANG, X., ZHOU, X., LIN, M. and SUN, J. (2018b). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 6848–6856.

A Conversation with Stephen Portnoy

Xuming He and Xiaofeng Shao

Abstract. Steve Portnoy was born in Kankakee, Illinois in 1942. He did his undergraduate studies in mathematics at Massachusetts Institute of Technology, and then earned a master’s degree and a Ph.D. degree from the statistics department at Stanford University in 1966 and 1969, respectively.

Steve Portnoy has had a distinguished career and is widely recognized as a preeminent mathematical statistician. He has made pioneering and influential contributions in several areas in statistics, including asymptotic theory, robust statistics, quantile regression, and statistics in biology. He has published more than 100 research articles. He is a former co-editor of *Journal of the American Statistical Association (Theory and Methods)*, an elected fellow of American Statistical Association (ASA), Institute of Mathematical Statistics (IMS) and American Association for the Advancement of Science (AAAS).

Steve’s professional positions have included being on the faculty of the Department of Statistics at Harvard University and the University of Illinois at Urbana-Champaign for more than 30 years. He was a founding member of the Department of Statistics at the University of Illinois in 1985 and served as the division chair for the Statistics Program in the Mathematics department before the Statistics department was established.

Key words and phrases: Asymptotic approximation, high-dimensional models, quantile regression, robust statistics.

REFERENCES

- [1] PORTNOY, S. (1991). Asymptotic behavior of the number of regression quantile breakpoints. *SIAM J. Sci. Statist. Comput.* **12** 867–883. MR1102413 <https://doi.org/10.1137/0912047>
- [2] PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. MR2041488 <https://doi.org/10.1198/016214503000000954>
- [3] PORTNOY, S. (2012). Nearly root- n approximation for regression quantile processes. *Ann. Statist.* **40** 1714–1736. MR3015041 <https://doi.org/10.1214/12-AOS1021>
- [4] PORTNOY, S. and KOENKER, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300. With comments by Ronald A. Thisted and M. R. Osborne and a rejoinder by the authors. MR1619189 <https://doi.org/10.1214/ss/1030037960>
- [5] ZABELL, S. L. (2008). On Student’s 1908 article “The probable error of a mean.” *J. Amer. Statist. Assoc.* **103** 1–20. With comments and a rejoinder by the author. MR2394634 <https://doi.org/10.1198/016214508000000030>

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

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

President-Elect: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

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

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

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

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

IMS EDITORS

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

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

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

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

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

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



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