STATISTICAL SCIENCE

Volume 38, Number 1 February 2023

***Statistical Science* [ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 38, Number 1, February 2023. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

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

Statistical Science Volume 38, Number 1 (1–183) February 2023

Volume 38 **Number 1** **February 2023**

Statistical Science
2023, Vol. 38, No. 1, 1–12
https://doi.org/10.1214/21-STS843

The Costs and Benefits of Uniformly Valid Causal Inference with High-Dimensional Nuisance Parameters

Niloofar Moosavi, Jenny Häggström and Xavier de Luna

Abstract. Important advances have recently been achieved in developing procedures yielding uniformly valid inference for a low dimensional causal parameter when high-dimensional nuisance models must be estimated. In this paper, we review the literature on uniformly valid causal inference and discuss the costs and benefits of using uniformly valid inference procedures. Naive estimation strategies based on regularization, machine learning, or a preliminary model selection stage for the nuisance models have finite sample distributions which are badly approximated by their asymptotic distributions. To solve this serious problem, estimators which converge uniformly in distribution over a class of data generating mechanisms have been proposed in the literature. In order to obtain uniformly valid results in high-dimensional situations, sparsity conditions for the nuisance models need typically to be made, although a double robustness property holds, whereby if one of the nuisance model is more sparse, the other nuisance model is allowed to be less sparse. While uniformly valid inference is a highly desirable property, uniformly valid procedures pay a high price in terms of inflated variability. Our discussion of this dilemma is illustrated by the study of a double-selection outcome regression estimator, which we show is uniformly asymptotically unbiased, but is less variable than uniformly valid estimators in the numerical experiments conducted.

Key words and phrases: Double robustness, machine learning, post-model selection inference, regularization, superefficiency.

REFERENCES

[1] ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

[2] AVAGYAN, V. and VANSTEELANDT, S. (2021). High-dimensional inference for the average treatment effect under model misspecification using penalized bias-reduced double-robust estimation. *Biostatistics & Epidemiology* 1–18. https://doi.org/10.1080/24709360.2021.1898730

[3] BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

[4] BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429. MR3001131 https://doi.org/10.3982/ECTA9626

[5] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

[6] BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

[7] BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. MR0663424

[8] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. Reprint of the 1993 original. MR1623559

[9] BRADIC, J., WAGER, S. and ZHU, Y. (2019). Sparsity double robust inference of average treatment effects. ArXiv preprint. Available at arXiv:1905.00744.

Niloofar Moosavi is Ph.D. Student, Department of Statistics, USBE, Umeå University, 901 87, Umeå, Sweden (e-mail: niloofar.moosavi@umu.com). Jenny Häggström is Associate Professor, Department of Statistics, USBE, Umeå University, 901 87, Umeå, Sweden (e-mail: jenny.haggstrom@umu.se). Xavier de Luna is Professor, Department of Statistics, USBE, Umeå University, 901 87, Umeå, Sweden (e-mail: xavier.de.luna@umu.se).

[10] Cai, W. and van der Laan, M. (2019). Nonparametric bootstrap inference for the targeted highly adaptive LASSO estimator. ArXiv preprint. Available at arXiv:1905.10299.

[11] Cattaneo, M. D., Jansson, M. and Ma, X. (2019). Two-step estimation and inference with possibly many included covariates. *Rev. Econ. Stud.* **86** 1095–1122. MR3945564 https://doi.org/10.1093/restud/rdy053

[12] Cattaneo, M. D., Jansson, M. and Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *J. Amer. Statist. Assoc.* **113** 1350–1361. MR3862362 https://doi.org/10.1080/01621459.2017.1328360

[13] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[14] Chernozhukov, V., Hansen, C. and Spindler, M. (2016). hdm: High-dimensional metrics. ArXiv preprint. Available at arXiv:1608.00354.

[15] Chernozhukov, V., Newey, W. and Singh, R. (2020). De-Biased Machine Learning of Global and Local Parameters Using Regularized Riesz Representers.

[16] Cui, Y. and Tchetgen Tchetgen, E. (2019). Selective machine learning of doubly robust functionals. ArXiv preprint. Available at arXiv:1911.02029.

[17] D'Amour, A., Ding, P., Feller, A., Lei, L. and Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. *J. Econometrics* **221** 644–654. MR4215042 https://doi.org/10.1016/j.jeconom.2019.10.014

[18] de Luna, X. and Johansson, P. (2014). Testing for the unconfoundedness assumption using an instrumental assumption. *J. Causal Inference* **2** 187–199. MR4289420 https://doi.org/10.1515/jci-2013-0011

[19] de Luna, X., Waernbaum, I. and Richardson, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98** 861–875. MR2860329 https://doi.org/10.1093/biomet/asr041

[20] Díaz, I. (2020). Machine learning in the estimation of causal effects: Targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics* **21** 353–358. MR4132549 https://doi.org/10.1093/biostatistics/kxz042

[21] Díaz, I., Luedtke, A. R. and van der Laan, M. J. (2018). Sensitivity analysis. In *Targeted Learning in Data Science*. *Springer Ser. Statist.* 511–522. Springer, Cham. MR3820742

[22] Dukes, O. and Vansteelandt, S. (2021). Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika* **108** 321–334. MR4259134 https://doi.org/10.1093/biomet/asaa071

[23] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

[24] Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 https://doi.org/10.1016/j.jeconom.2015.06.017

[25] Farrell, M. H., Liang, T. and Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica* **89** 181–213. MR4220387 https://doi.org/10.3982/ecta16901

[26] Genbäck, M. and de Luna, X. (2019). Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. *Biometrics* **75** 506–515. MR3999174 https://doi.org/10.1111/biom.13001

[27] Gruber, S. and van der Laan, M. J. (2012). tmle: An R package for targeted maximum likelihood estimation. *J. Stat. Softw.* **51** 1–35. https://doi.org/doi:10.18637/jss.v051.i13

[28] Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242 https://doi.org/10.2307/2998560

[29] Hahn, J. (2004). Functional restriction and efficiency in causal inference. *Rev. Econ. Stat.* **86** 73–76.

[30] Imbens, G. W., Newey, W. K. and Ridder, G. (2005). Mean-square-error calculations for average treatment effects. IEPR Working Paper 05.34.

[31] Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

[32] Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and Their Applications in Public Health Research* (H. He, P. Wu and D.-G. Chen, eds.) *ICSA Book Ser. Stat.* 141–167. Springer, Cham. MR3617956

[33] Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

[34] Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21** 21–59. MR2153856 https://doi.org/10.1017/S0266466605050036

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

[36] Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics*: *The Harald Cramér Volume* (U. Grenander, ed.) 416–444. Wiley, New York. MR0112201

[37] Neyman, J. (1979). $C(\alpha)$ tests and their use. *Sankhyā Ser. A* **41** 1–21. MR0615037

[38] Ning, Y., Peng, S. and Imai, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika* **107** 533–554. MR4138975 https://doi.org/10.1093/biomet/asaa020

[39] R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[40] Robins, J., Li, L., Tchetgen Tchetgen, E., van der Vaart, A. et al. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A. Freedman* 335–421. IMS.

[41] Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[42] Rotnitzky, A. and Smucler, E. (2020). Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J. Mach. Learn. Res.* **21** Paper No. 188. MR4209474

[43] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

[44] Rubin, D. B. (1990). Formal modes of statistical inference for causal effects. *J. Statist. Plann. Inference* **25** 279–292.

[45] Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478 https://doi.org/10.2307/2669923

[46] SCHNITZER, M. E., LOK, J. J. and GRUBER, S. (2016). Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *Int. J. Biostat.* **12** 97–115. MR3505689 https://doi.org/10.1515/ijb-2015-0017

[47] SEMENOVA, V. and CHERNOZHUKOV, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econom. J.* **24** 264–289. MR4281225 https://doi.org/10.1093/ectj/utaa027

[48] SHORTREED, S. M. and ERTEFAIE, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics* **73** 1111–1122. MR3744525 https://doi.org/10.1111/biom.12679

[49] SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust \ell_1 regularized estimation of causal contrasts. ArXiv preprint. Available at arXiv:1904.03737.

[50] SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. D'abrowska and T. P. Speed. MR1092986

[51] TAN, Z. (2007). Comment: Understanding OR, PS and DR [MR2420458]. *Statist. Sci.* **22** 560–568. MR2420461 https://doi.org/10.1214/07-STS227A

[52] TAN, Z. (2020). Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *Ann. Statist.* **48** 811–837. MR4102677 https://doi.org/10.1214/19-AOS1824

[53] TANG, D., KONG, D., PAN, W. and WANG, L. (2020). Outcome model free causal inference with ultra-high dimensional covariates. ArXiv preprint. Available at arXiv:2007.14190.

[54] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

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

[56] 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

[57] VAN DER LAAN, M. J. (2014). Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int. J. Biostat.* **10** 29–57. MR3208072 https://doi.org/10.1515/ijb-2012-0038

[58] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

[59] VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11. MR2306500 https://doi.org/10.2202/1557-4679.1043

[60] VAN DER VAART, A. W. (1997). Superefficiency. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 397–410. Springer, New York. MR1462961

[61] 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

[62] ZHENG, W. and VAN DER LAAN, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning. Springer Ser. Statist.* 459–474. Springer, New York. MR2867139 https://doi.org/10.1007/978-1-4419-9782-1_27

Statistical Science
2023, Vol. 38, No. 1, 13–29
https://doi.org/10.1214/21-STS846

Additive Bayesian Variable Selection under Censoring and Misspecification

David Rossell and Francisco Javier Rubio

Abstract. We discuss the role of misspecification and censoring on Bayesian model selection in the contexts of right-censored survival and concave log-likelihood regression. Misspecification includes wrongly assuming the censoring mechanism to be noninformative. Emphasis is placed on additive accelerated failure time, Cox proportional hazards and probit models. We offer a theoretical treatment that includes local and nonlocal priors, and a general nonlinear effect decomposition to improve power-sparsity trade-offs. We discuss a fundamental question: what solution can one hope to obtain when (inevitably) models are misspecified, and how to interpret it? Asymptotically, covariates that do not have predictive power for neither the outcome nor (for survival data) censoring times, in the sense of reducing a likelihood-associated loss, are discarded. Misspecification and censoring have an asymptotically negligible effect on false positives, but their impact on power is exponential. We show that it can be advantageous to consider simple models that are computationally practical yet attain good power to detect potentially complex effects, including the use of finite-dimensional basis to detect truly nonparametric effects. We also discuss algorithms to capitalize on sufficient statistics and fast likelihood approximations for Gaussian-based survival and binary models.

Key words and phrases: Additive regression, generalized additive model, misspecification, model selection, survival.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series, No.* 55. U.S. Government Printing Office, Washington, D.C. MR0167642

[2] BOCHKINA, N. A. and GREEN, P. J. (2014). The Bernstein-von Mises theorem and nonregular models. *Ann. Statist.* **42** 1850–1878. MR3262470 https://doi.org/10.1214/14-AOS1239

[3] BURRIDGE, J. (1981). A note of maximum likelihood estimation for regression models using grouped data. *J. Roy. Statist. Soc. Ser. B* **43** 41–45. MR0610375

[4] CALON, A., ESPINET, E., PALOMO-PONCE, S., TAURIELLO, D. V. F., IGLESIAS, M., CÉSPEDES, M. V., SEVILLANO, M., NADAL, C., JUNG, P. et al. (2012). Dependency of colorectal cancer on a TGF-beta-driven programme in stromal cells for metastasis initiation. *Cancer Cell* **22** 571–584.

[5] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[6] CHEN, Y. Q. and JEWELL, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika* **88** 687–702. MR1859402 https://doi.org/10.1093/biomet/88.3.687

[7] CONSONNI, G., FOUSKAKIS, D., LISEO, B. and NTZOUFRAS, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Anal.* **13** 627–679. MR3807861 https://doi.org/10.1214/18-BA1103

[8] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

[9] DIRIENZO, A. G. and LAGAKOS, S. W. (2001). Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 745–757. MR1872064 https://doi.org/10.1111/1467-9868.00310

[10] DUNSON, D. B. and HERRING, A. H. (2005). Bayesian model selection and averaging in additive and proportional hazards models. *Lifetime Data Anal.* **11** 213–232. MR2158783 https://doi.org/10.1007/s10985-004-0384-x

[11] FARAGGI, D. and SIMON, R. (1998). Bayesian variable selection method for censored survival data. *Biometrics* **54** 1475–1485. MR1671590 https://doi.org/10.2307/2533672

[12] GASULL, A. and UTZET, F. (2014). Approximating Mills

David Rossell is Associate Professor, Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain (e-mail: david.rossell@upf.edu). Francisco Javier Rubio is Assistant Professor, Department of Statistical Science, University College London, London, United Kingdom (e-mail: f.j.rubio@ucl.ac.uk).

ratio. *J. Math. Anal. Appl.* **420** 1832–1853. MR3240110 https://doi.org/10.1016/j.jmaa.2014.05.034

[13] GRIFFIN, J. E., ŁATUSZYŃSKI, K. G. and STEEL, M. F. J. (2021). In search of lost mixing time: Adaptive Markov chain Monte Carlo schemes for Bayesian variable selection with very large p. *Biometrika* **108** 53–69. MR4226189 https://doi.org/10.1093/biomet/asaa055

[14] HAHN, P. R. and CARVALHO, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *J. Amer. Statist. Assoc.* **110** 435–448. MR3338514 https://doi.org/10.1080/01621459.2014.993077

[15] HARRELL JR., F. E., LEE, K. L. and MARK, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15** 361–387.

[16] HATTORI, S. (2012). Testing the no-treatment effect based on a possibly misspecified accelerated failure time model. *Statist. Probab. Lett.* **82** 371–377. MR2875225 https://doi.org/10.1016/j.spl.2011.10.016

[17] HJORT, N. L. (1992). On inference in parametric survival data models. *Int. Stat. Rev.* **60** 355–387.

[18] HJORT, N. L. and POLLARD, D. (2011). Asymptotics for minimisers of convex processes. Available at arXiv:1107.3806.

[19] HOUGAARD, P. (1995). Frailty models for survival data. *Lifetime Data Anal.* **1** 255–273.

[20] HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. MR2247210 https://doi.org/10.1111/j.1541-0420.2006.00562.x

[21] HUTTON, J. L. and MONAGHAN, P. F. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: Asymptotic results. *Lifetime Data Anal.* **8** 375–393. MR1942343 https://doi.org/10.1023/A:1020570922072

[22] IBRAHIM, J. G., CHEN, M.-H. and MACEACHERN, S. N. (1999). Bayesian variable selection for proportional hazards models. *Canad. J. Statist.* **27** 701–717. MR1767142 https://doi.org/10.2307/3316126

[23] JOHNSON, V. E. and ROSSELL, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 143–170. MR2830762 https://doi.org/10.1111/j.1467-9868.2009.00730.x

[24] JOHNSON, V. E. and ROSSELL, D. (2012). Bayesian model selection in high-dimensional settings. *J. Amer. Statist. Assoc.* **107** 649–660. MR2980074 https://doi.org/10.1080/01621459.2012.682536

[25] KEIDING, N., ANDERSEN, P. K. and KLEIN, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat. Med.* **16** 215–224.

[26] KHAN, M. H. R. and SHAW, J. E. H. (2019). Variable selection for accelerated lifetime models with synthesized estimation techniques. *Stat. Methods Med. Res.* **28** 937–952. MR3922901 https://doi.org/10.1177/0962280217739522

[27] LAIRD, N. and OLIVIER, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *J. Amer. Statist. Assoc.* **76** 231–240. MR0624329

[28] IBRAHIM, J. G. and CHEN, M. H. (2014). Bayesian model selection in survival analysis. In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society.

[29] LEE, C.-I. C. (1992). On Laplace continued fraction for the normal integral. *Ann. Inst. Statist. Math.* **44** 107–120. MR1165575 https://doi.org/10.1007/BF00048673

[30] LIN, D. Y. and WEI, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84** 1074–1078. MR1134495

[31] LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403 https://doi.org/10.1214/16-AOS1471

[32] NARISETTY, N. N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. MR3210987 https://doi.org/10.1214/14-AOS1207

[33] NIKOOIENEJAD, A., WANG, W. and JOHNSON, V. E. (2020). Bayesian variable selection for survival data using inverse moment priors. *Ann. Appl. Stat.* **14** 809–828. MR4117831 https://doi.org/10.1214/20-AOAS1325

[34] PANOV, M. and SPOKOINY, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

[35] POLSON, N. G. and SUN, L. (2019). Bayesian l_0-regularized least squares. *Appl. Stoch. Models Bus. Ind.* **35** 717–731. MR3974246 https://doi.org/10.1002/asmb.2381

[36] ROSEN, J. B. (1971). Minimum error bounds for multidimensional spline approximation. *J. Comput. System Sci.* **5** 430–452. MR0283462 https://doi.org/10.1016/S0022-0000(71)80026-0

[37] ROSSELL, D. (2021). A framework for posterior consistency in model selection. *Bayesian Anal.* **in press**.

[38] ROSSELL, D., ABRIL, O. and BHATTACHARYA, A. (2021). Approximate Laplace approximations for scalable model selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 853–879. MR4320004

[39] ROSSELL, D. and RUBIO, F. J. (2018). Tractable Bayesian variable selection: Beyond normality. *J. Amer. Statist. Assoc.* **113** 1742–1758. MR3902243 https://doi.org/10.1080/01621459.2017.1371025

[40] ROSSELL, D. and RUBIO, F. J. (2023). Supplement to "Additive Bayesian variable selection under censoring and misspecification." https://doi.org/10.1214/21-STS846SUPPA, https://doi.org/10.1214/21-STS846SUPPB

[41] ROSSELL, D. and TELESCA, D. (2017). Nonlocal priors for high-dimensional estimation. *J. Amer. Statist. Assoc.* **112** 254–265. MR3646569 https://doi.org/10.1080/01621459.2015.1130634

[42] ROSSELL, D., TELESCA, D. and JOHNSON, V. E. (2013). High-dimensional Bayesian classifiers using non-local priors. In *Statistical Models for Data Analysis XV* 305–314. Springer, Berlin.

[43] SANDERSON, C. and CURTIN, R. (2016). Armadillo: A template-based C ++ library for linear algebra. *J. Open Sour. Softw.* **1** 26.

[44] SCHEIPL, F., FAHRMEIR, L. and KNEIB, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *J. Amer. Statist. Assoc.* **107** 1518–1532. MR3036413 https://doi.org/10.1080/01621459.2012.737742

[45] SCOTT, J. G. and BERGER, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Statist.* **38** 2587–2619. MR2722450 https://doi.org/10.1214/10-AOS792

[46] SHA, N., TADESSE, M. G. and VANNUCCI, M. (2006). Bayesian variable selection for the analysis of microarray data with censored outcomes. *Bioinformatics* **22** 2262–2268.

[47] SILVAPULLE, M. J. and BURRIDGE, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *J. Roy. Statist. Soc. Ser. B* **48** 100–106. MR0848055

[48] SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39** 1–13. https://doi.org/10.18637/jss.v039.i05

[49] SOLOMON, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika* **71** 291–298. MR0767157 https://doi.org/10.1093/biomet/71.2.291

[50] Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I. et al. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* **54** 1–30.

[51] 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

[52] Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395.

[53] Tong, X., Zhu, L., Leng, C., Leisenring, W. and Robison, L. L. (2013). A general semiparametric hazards regression model: Efficient estimation and structure selection. *Stat. Med.* **32** 4980–4994. MR3127189 https://doi.org/10.1002/sim.5885

[54] Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108. MR0600535

[55] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, New York.

[56] Yang, Y. and Pati, D. (2017). Bayesian model selection consistency and oracle inequality with intractable marginal likelihood. Available at arXiv:1701.00311.

[57] Yang, Y., Wainwright, M. J. and Jordan, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. MR3576552 https://doi.org/10.1214/15-AOS1417

[58] Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21** 76–99. MR1212167 https://doi.org/10.1214/aos/1176349016

[59] Zanella, G. and Roberts, G. (2019). Scalable importance tempering and Bayesian variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 489–517. MR3961496

[60] Zhang, Z., Sinha, S., Maiti, T. and Shipp, E. (2018). Bayesian variable selection in the accelerated failure time model with an application to the surveillance, epidemiology, and end results breast cancer data. *Stat. Methods Med. Res.* **27** 971–990. MR3770130 https://doi.org/10.1177/0962280215626947

Statistical Science
2023, Vol. 38, No. 1, 30–51
https://doi.org/10.1214/21-STS847

On Some Connections Between Esscher's Tilting, Saddlepoint Approximations, and Optimal Transportation: A Statistical Perspective

Davide La Vecchia, Elvezio Ronchetti and Andrej Ilievski

Abstract. We showcase some unexplored connections between saddlepoint approximations, measure transportation, and some key topics in information theory. To bridge these different areas, we review selectively the fundamental results available in the literature and we draw the connections between them. First, for a generic random variable we explain how the Esscher's tilting (which is a result rooted in information theory and lies at the heart of saddlepoint approximations) is connected to the solution of the dual Kantorovich problem (which lies at the heart of measure transportation theory) via the Legendre transform of the cumulant generating function. Then, we turn to statistics: we illustrate the connections when the random variable we work with is the sample mean or a statistic with known (either exact or approximate) cumulant generating function. The unveiled connections offer the possibility to look at the saddlepoint approximations from different angles, putting under the spotlight the links to convex analysis (via the notion of duality) or differential geometry (via the notion of geodesic). We feel these possibilities can trigger a knowledge transfer between statistics and other disciplines, like mathematics and machine learning. A discussion on some topics for future research concludes the paper.

Key words and phrases: Change of variable, Kullback–Leibler divergence, geodesic, optimal transportation map, Wasserstein distance.

REFERENCES

AEBERHARD, W. H., CANTONI, E. and HERITIER, S. (2017). Saddlepoint tests for accurate and robust inference on overdispersed count data. *Comput. Statist. Data Anal.* **107** 162–175. MR3575066 https://doi.org/10.1016/j.csda.2016.10.009

AMARI, S. (1989). The geometry of asymptotic inference: Comment. *Statist. Sci.* **4** 220–222.

AMARI, S. (2016). *Information Geometry and Its Applications. Applied Mathematical Sciences* **194**. Springer, Tokyo. MR3495836 https://doi.org/10.1007/978-4-431-55978-8

AMARI, S., KARAKIDA, R. and OIZUMI, M. (2018). Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem. *Inf. Geom.* **1** 13–37. MR3974671 https://doi.org/10.1007/s41884-018-0002-8

ARCONES, M. A. (2006). Large deviations for M-estimators. *Ann. Inst. Statist. Math.* **58** 21–52. MR2281205 https://doi.org/10.1007/s10463-005-0017-5

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No.* 4. SIAM, Philadelphia, PA. MR0315820

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications. *J. Roy. Statist. Soc. Ser. B* **41** 279–312. MR0557595

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics. Monographs on Statistics and Applied Probability*. CRC Press, London. MR1010226 https://doi.org/10.1007/978-1-4899-3424-6

BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions. Classics in Applied Mathematics*

Davide La Vecchia is Associate Professor, Research Center for Statistics, Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland (e-mail: davide.lavecchia@unige.ch). Elvezio Ronchetti is Professor Emeritus, Research Center for Statistics, Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland (e-mail: elvezio.ronchetti@unige.ch). Andrej Ilisevski is master student in Statistics at ETH Zürich, Zürich, Switzerland (e-mail: andrej.ilievski.98@gmail.com).

64. SIAM, Philadelphia, PA. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

BOBKOV, S. and LEDOUX, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

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

CARTER, K. M., RAICH, R. and HERO, A. (2007). Learning on statistical manifolds for clustering and visualization. In 45*th Allerton Conference on Communication*, *Control*, *and Computing*, Monticello, IL.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. MR0057518 https://doi.org/10.1214/aoms/1177729330

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

COSTA, S. I. R., SANTOS, S. A. and STRAPASSON, J. E. (2015). Fisher information distance: A geometrical reading. *Discrete Appl. Math.* **197** 59–69. MR3398961 https://doi.org/10.1016/j.dam.2014.10.004

CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

DAVISON, A. C. and HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75** 417–431. MR0967581 https://doi.org/10.1093/biomet/75.3.417

DE PHILIPPIS, G. and FIGALLI, A. (2014). The Monge–Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc.* (*N.S.*) **51** 527–580. MR3237759 https://doi.org/10.1090/S0273-0979-2014-01459-4

DEB, N. and SEN, B. (2021). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* 1–16.

DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671. MR4147635 https://doi.org/10.1016/j.jmva.2020.104671

DEL BARRIO, E., CUESTA-ALBERTOS, J. A., MATRÁN, C. and RODRÍGUEZ-RODRÍGUEZ, J. M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.* **27** 1230–1239. MR1740113 https://doi.org/10.1214/aos/1017938923

EASTON, G. S. and RONCHETTI, E. (1986). General saddlepoint approximations with applications to L statistics. *J. Amer. Statist. Assoc.* **81** 420–430. MR0845879

ESSCHER, F. (1932). On the probability function in the collective theory of risk. *Skand. Aktuarietidskr.* **15** 175–195.

FASIOLO, M., WOOD, S. N., HARTIG, F. and BRAVINGTON, M. V. (2018). An extended empirical saddlepoint approximation for intractable likelihoods. *Electron. J. Stat.* **12** 1544–1578. MR3806432 https://doi.org/10.1214/18-ejs1433

FIELD, C. (1982). Small sample asymptotic expansions for multivariate M-estimates. *Ann. Statist.* **10** 672–689. MR0663425

FIELD, C. and RONCHETTI, E. (1990). *Small Sample Asymptotics*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **13**. IMS, Hayward, CA. MR1088480

GALICHON, A. (2016). *Optimal Transport Methods in Economics*. Princeton Univ. Press, Princeton, NJ. MR3586373 https://doi.org/10.1515/9781400883592

GALICHON, A. (2017). A survey of some recent applications of optimal transport methods to econometrics. *Econom. J.* **20** C1–C11. MR3685644 https://doi.org/10.1111/ectj.12083

GATTO, R. (2017). Multivariate saddlepoint tests on the mean direction of the von Mises–Fisher distribution. *Metrika* **80** 733–747. MR3716614 https://doi.org/10.1007/s00184-017-0625-0

GHOSAL, P. and SEN, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. Available at https://arxiv.org/abs/1905.05340.

GOUTIS, C. and CASELLA, G. (1999). Explaining the saddlepoint approximation. *Amer. Statist.* **53** 216–224. MR1711555 https://doi.org/10.2307/2686100

GRENANDER, U. and MILLER, M. I. (2007). *Pattern Theory*: *From Representation to Inference*. Oxford Univ. Press, Oxford. MR2285439

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7

HALLIN, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: A measure transportation approach. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

HALLIN, M., HLUBINKA, D. and HUDECOVÁ, Š. (2020). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. ArXiv preprint. Available at arXiv:2007.15496.

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020a). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2020.1832501

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020b). Rank-based testing for semiparametric VAR models: A measure transportation approach. *Bernoulli* 1–51. Available at arXiv:2011.06062.

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

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HOLCBLAT, B. and SOWELL, F. (2019). The empirical saddlepoint estimator. Available at https://arxiv.org/abs/1905.06977.

HUBER, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. (2013). *An Introduction to Statistical Learning*: *With Applications in R*. *Springer Texts in Statistics* **103**. Springer, New York. MR3100153 https://doi.org/10.1007/978-1-4614-7138-7

JENSEN, J. L. (1995). *Saddlepoint Approximations*. *Oxford Statistical Science Series* **16**. The Clarendon Press, Oxford University Press, New York. MR1354837

JIANG, C., LA VECCHIA, D., RONCHETTI, E. and SCAILLET, O. (2021). Saddlepoint approximations for spatial panel data models. *J. Amer. Statist. Assoc.* 1–28. https://doi.org/10.1080/01621459.2021.1981913

KANTOROVITCH, L. (1942). On the translocation of masses. *C. R.* (*Dokl.*) *Acad. Sci. URSS* **37** 199–201. MR0009619

KASS, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4** 188–234. MR1015274

KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657 https://doi.org/10.1017/CBO9780511754098

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

KOLASSA, J. E. (2006). *Series Approximation Methods in Statistics. Lecture Notes in Statistics* **88**. Springer, New York.

KOLASSA, J. and LI, J. (2010). Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli* **16** 1191–1207. MR2759175 https://doi.org/10.3150/09-BEJ237

KOLOURI, S., PARK, S. R., THORPE, M., SLEPCEV, D. and ROHDE, G. K. (2017). Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Process. Mag.* **34** 43–59.

KREMER, E. (1982). A characterization of the Esscher-transformation. *Astin Bull.* **13** 57–59.

KULLBACK, S. (1997). *Information Theory and Statistics*. Dover, Mineola, NY. MR1461541

LA VECCHIA, D. (2016). Stable asymptotics for M-estimators. *Int. Stat. Rev.* **84** 267–290. MR3537156 https://doi.org/10.1111/insr.12102

LA VECCHIA, D. and RONCHETTI, E. (2019). Saddlepoint approximations for short and long memory time series: A frequency domain approach. *J. Econometrics* **213** 578–592. MR4023923 https://doi.org/10.1016/j.jeconom.2018.10.009

LA VECCHIA, D., RONCHETTI, E. and TROJANI, F. (2012). Higher-order infinitesimal robustness. *J. Amer. Statist. Assoc.* **107** 1546–1557. MR3036415 https://doi.org/10.1080/01621459.2012.738580

LÉONARD, C. (2007). A large deviation approach to optimal transport. Working paper.

LÔ, S. N. and RONCHETTI, E. (2009). Robust and accurate inference for generalized linear models. *J. Multivariate Anal.* **100** 2126–2136. MR2543091 https://doi.org/10.1016/j.jmva.2009.06.012

MCCANN, R. J. and GUILLEN, N. (2013). Five lectures on optimal transportation: Geometry, regularity and applications. In *Analysis and Geometry of Metric Measure Spaces. CRM Proc. Lecture Notes* **56** 145–180. Amer. Math. Soc., Providence, RI. MR3060502 https://doi.org/10.1090/crmp/056/06

MCCULLAGH, P. (2018). *Tensor Methods in Statistics*. Dover Publications, New York.

MONGE, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de L'Académie Royale des Sciences de Paris*.

MONTI, A. C. and RONCHETTI, E. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika* **80** 329–338. MR1243508 https://doi.org/10.1093/biomet/80.2.329

MURPHY, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge.

PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

PANARETOS, V. M. and ZEMEL, Y. (2020). *An Invitation to Statistics in Wasserstein Space. SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham. MR4350694 https://doi.org/10.1007/978-3-030-38438-8

PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.

PITIÉ, F., KOKARAM, A. C. and DAHYOT, R. (2007). Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.* **107** 123–137.

R CORE TEAM (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

REID, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3** 213–238. MR0968390

REID, N. and FRASER, D. A. S. (1989). The geometry of asymptotic inference: Comment. *Statist. Sci.* **4** 231–233.

RIO, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 https://doi.org/10.1214/08-AIHP187

ROBINSON, J., RONCHETTI, E. and YOUNG, G. A. (2003). Saddlepoint approximations and tests based on multivariate M-estimates. *Ann. Statist.* **31** 1154–1169. MR2001646 https://doi.org/10.1214/aos/1059655909

ROCKAFELLAR, R. T. (2015). *Convex Analysis*. Princeton Univ. Press, Princeton, NJ.

RONCHETTI, E. and SABOLOVA, R. (2016). Saddlepoint tests for quantile regression. *Canad. J. Statist.* **44** 271–299. MR3536198 https://doi.org/10.1002/cjs.11288

RONCHETTI, E. and WELSH, A. H. (1994). Empirical saddlepoint approximations for multivariate M-estimators. *J. Roy. Statist. Soc. Ser. B* **56** 313–326. MR1281936

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics, Vol.* 162. Wiley, New York.

SMALL, C. G. (2010). *Expansions and Asymptotics for Statistics*. CRC Press/CRC, Boca Raton.

TOMA, A. and BRONIATOWSKI, M. (2011). Dual divergence estimators and tests: Robustness results. *J. Multivariate Anal.* **102** 20–36. MR2729417 https://doi.org/10.1016/j.jmva.2010.07.010

TOMA, A. and LEONI-AUBIN, S. (2010). Robust tests based on dual divergence estimators and saddlepoint approximations. *J. Multivariate Anal.* **101** 1143–1155. MR2595297 https://doi.org/10.1016/j.jmva.2009.11.001

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

VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

YOUNG, G. A. (2009). Routes to higher-order accuracy in parametric inference. *Aust. N. Z. J. Stat.* **51** 115–126. MR2531982 https://doi.org/10.1111/j.1467-842X.2009.00548.x

Statistical Science
2023, Vol. 38, No. 1, 52–67
https://doi.org/10.1214/21-STS848

Bayesian Adaptive Randomization with Compound Utility Functions

Alessandra Giovagnoli and Isabella Verdinelli

Abstract. Bayesian adaptive designs formalize the use of previous knowledge at the planning stage of an experiment, permitting recursive updating of the prior information. They often make use of utility functions, while also allowing for randomization. We review frequentist and Bayesian adaptive design methods and show that some of the frequentist adaptive design methodology can also be employed in a Bayesian context. We use compound utility functions for the Bayesian designs, that are a trade-off between an optimal design information criterion, that represents the acquisition of scientific knowledge, and some ethical or utilitarian gain. We focus on binary response models on two groups with independent Beta prior distributions on the success probabilities. The treatment allocation is shown to converge to the allocation that produces the maximum utility. Special cases are the Bayesian Randomized (simply) Adaptive Compound (BRAC) design, an extension of the frequentist Sequential Maximum Likelihood (SML) design and the Bayesian Randomized (doubly) Adaptive Compound Efficient (BRACE) design, a generalization of the Efficient Randomized Adaptive DEsign (ERADE). Numerical simulation studies compare BRAC with BRACE when D-optimality is the information criterion chosen. In analogy with the frequentist theory, the BRACE-D design appears more efficient than the BRAC-D design.

Key words and phrases: Bayesian designs, binary response model, doubly adaptive designs, optimal design criteria, optimal target, utility functions.

REFERENCES

[1] ASHBY, D. (2006). Bayesian statistics in medicine: A 25 year review. *Stat. Med.* **25** 3589–3631. MR2252415 https://doi.org/10.1002/sim.2672

[2] ATKINSON, A. and BISWAS, A. (2014). *Randomised Response-Adaptive Designs in Clinical Trials*. CRC Press, Boca Raton.

[3] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2010). Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika* **97** 935–946. MR2746162 https://doi.org/10.1093/biomet/asq055

[4] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2015). *Adaptive Designs for Sequential Treatment Allocation*. *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. MR3616325

[5] BALDI ANTOGNINI, A., VAGHEGGINI, A. and ZAGORAIOU, M. (2018). Is the classical Wald test always suitable under response-adaptive randomization? *Stat. Methods Med. Res.* **27** 2294–2311. MR3825908 https://doi.org/10.1177/0962280216680241

[6] BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2015). On the almost sure convergence of adaptive allocation procedures. *Bernoulli* **21** 881–908. MR3338650 https://doi.org/10.3150/13-BEJ591

[7] BALDI-ANTOGNINI, A., GIOVAGNOLI, A. and ZAGORAIOU, M. (2012). Some recent developments in the design of adaptive clinical trials. *Statistica* **72** 375–393.

[8] BERRY, D. (2011). Adaptive clinical trials: The promise and the caution. *J. Clin. Oncol.* **29** 606–609.

[9] BERRY, D. A. (1985). Interim analyses in clinical trials: Classical vs. Bayesian approaches. *Stat. Med.* **4** 521–526.

[10] BERRY, S. M., CARLIN, B. P., LEE, J. J. and MÜLLER, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. *Chapman & Hall/CRC Biostatistics Series* **38**. CRC Press, Boca Raton, FL. MR2723582

[11] BURTON, P. R., GURRIN, L. C. and HUSSEY, M. H. (1997). Interpreting the clinical trials of extracorporeal membrane oxygenation in the treatment of persistent pulmonary hypertension of the newborn. *Seminars in Neonatology* **2** 69–79.

[12] CHALONER, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12** 283–300. MR0733514 https://doi.org/10.1214/aos/1176346407

Alessandra Giovagnoli is retired Professor, Department of Statistical Sciences, Alma Mater Studiorum, Università di Bologna, Bologna, Italy (e-mail: alessandra.giovagnoli@unibo.it). Isabella Verdinelli is Professor in Residence, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (e-mail: isabella@stat.cmu.edu).

[13] CHALONER, K. and LARNTZ, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* **21** 191–208. MR0985457 https://doi.org/10.1016/0378-3758(89)90004-9

[14] CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

[15] CHENG, Q. and YANG, M. (2019). On multiple-objective optimal designs. *J. Statist. Plann. Inference* **200** 87–101. MR3907271 https://doi.org/10.1016/j.jspi.2018.09.007

[16] CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Stat.* **30** 755–770. MR0108874 https://doi.org/10.1214/aoms/1177706205

[17] CHEVRET, S. (2012). Bayesian adaptive clinical trials: A dream for statisticians only? *Stat. Med.* **31** 1002–1013. MR2925674 https://doi.org/10.1002/sim.4363

[18] CLYDE, M. and CHALONER, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *J. Amer. Statist. Assoc.* **91** 1236–1244. MR1424621 https://doi.org/10.2307/2291742

[19] COOK, J. D. and NADARAJAH, S. (2006). Stochastic inequality probabilities for adaptively randomized clinical trials. *Biom. J.* **48** 356–365. MR2240085 https://doi.org/10.1002/bimj.200510220

[20] DOOB, J. L. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et Ses Applications*. *Colloques Internationaux du Centre National de la Recherche Scientifique*, *No.* 13 23–27. Centre National de la Recherche Scientifique, Paris. MR0033460

[21] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660 https://doi.org/10.1093/biomet/58.3.403

[22] EISELE, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *J. Statist. Plann. Inference* **38** 249–261. MR1256599 https://doi.org/10.1016/0378-3758(94)90038-8

[23] ETZIONI, R. and KADANE, J. B. (1993). Optimal experimental design for another's analysis. *J. Amer. Statist. Assoc.* **88** 1404–1411. MR1245377

[24] GIOVAGNOLI, A. (2017). Bayesian optimal experimental designs for binary responses in an adaptive framework. *Appl. Stoch. Models Bus. Ind.* **33** 260–268. MR3666929 https://doi.org/10.1002/asmb.2207

[25] GIOVAGNOLI, A. (2021). The Bayesian design of adaptive clinical trials. *Int. J. Environ. Res. Public Health* **18** 530.

[26] HEY, S. P. and KIMMELMAN, J. (2015). Are outcome-adaptive allocation trials ethical? *Clin. Trials* **12** 102–106.

[27] HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2245329 https://doi.org/10.1002/047005588X

[28] HU, F. and ZHANG, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann. Statist.* **32** 268–301. MR2051008 https://doi.org/10.1214/aos/1079120137

[29] HU, F., ZHANG, L.-X. and HE, X. (2009). Efficient randomized-adaptive designs. *Ann. Statist.* **37** 2543–2560. MR2543702 https://doi.org/10.1214/08-AOS655

[30] JENNISON, C. and TURNBULL, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Anal.* **20** 209–234. MR1868774 https://doi.org/10.1081/SQA-100107646

[31] KADANE, J. B. and SEIDENFELD, T. (1990). Randomization in a Bayesian perspective. *J. Statist. Plann. Inference* **25** 329–345. MR1064430 https://doi.org/10.1016/0378-3758(90)90080-E

[32] KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319. MR0113263

[33] KORN, E. L. and FREIDLIN, B. B. (2011). Outcome-adaptive randomization: Is it useful? *J. Clin. Oncol.* **29** 771–776.

[34] KORN, E. L. and FREIDLIN, B. B. (2017). Adaptive clinical trials: Advantages and disadvantages of various adaptive design elements. *J. Natl. Cancer Inst.* **109**.

[35] LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

[36] LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379–398. MR0440812 https://doi.org/10.1080/02331937408842203

[37] LINDLEY, D. V. (1971). *Bayesian Statistics*, *a Review*. *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics*, *No.* 2. SIAM, Philadelphia, PA. MR0329081

[38] MELFI, V. F. and PAGE, C. (2000). Estimation after adaptive allocation. *J. Statist. Plann. Inference* **87** 353–363. MR1771124 https://doi.org/10.1016/S0378-3758(99)00198-6

[39] POCOCK, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64** 191–199.

[40] PROSCHAN, M. and EVANS, S. (2020). Resist the temptation of response-adaptive randomization. *Clin. Infect. Dis.* **71** 3002–3004.

[41] ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci.* **11** 137–149.

[42] ROSENBERGER, W. F. and LACHIN, J. M. (2016). *Randomization in Clinical Trials*: *Theory and Practice*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3443072 https://doi.org/10.1002/9781118742112

[43] ROSNER, G. (2020). Bayesan adaptive design in drug development. In *Bayesian Methods in Pharmaceutical Research* (E. Lesaffre, G. Baio and B. Boulanger, eds.) CRC Press, Boca Raton.

[44] SLIVKINS, A. (2021). Introduction to multi-armed bandits. Available at arXiv:1904.07272v6 [cs.LG].

[45] SMITH, A. F. M. and VERDINELLI, I. (1980). A note on Bayes designs for inference using a hierarchical linear model. *Biometrika* **67** 613–619. MR0601099 https://doi.org/10.1093/biomet/67.3.613

[46] SPIEGELHALTER, D. J. (1986). Probabilistic prediction in patient management and clinical trials. *Stat. Med.* **5** 421–433.

[47] STALLARD, N., TODD, S., RYAN, E. G. and GATES, S. (2020). Comparison of Bayesian and frequentist group-sequential clinical trial designs. *BMC Med. Res. Methodol.* **20** 1–14.

[48] SVERDLOV, O., ed. (2016). *Modern Adaptive Randomized Clinical Trials*: *Statistical and Practical Aspects*. *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. MR3642968

[49] SVERDLOV, O. and ROSENBERGER, W. F. (2013). On recent advances in optimal allocation designs in clinical trials. *J. Stat. Theory Pract.* **7** 753–773. MR3196633 https://doi.org/10.1080/15598608.2013.783726

[50] SVERDLOV, O., RYEZNIK, Y. and WONG, W.-K. (2014). Efficient and ethical response-adaptive randomization designs for multi-arm clinical trials with Weibull time-to-event outcomes. *J. Biopharm. Statist.* **24** 732–754. MR3210428 https://doi.org/10.1080/10543406.2014.903261

[51] THALL, P., FOX, P. and WATHEN, J. (2015). Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann. Oncol.* **26** 1621–1628. https://doi.org/10.1093/annonc/mdv238

[52] THALL, P. and WATHEN, J. (2007). Practical Bayesian adaptive randomisation in clinical trials. *Eur. J. Cancer* **43** 859–866.

[53] THOMPSON, W. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.

[54] Ventz, S., Parmigiani, G. and Trippa, L. (2017). Combining Bayesian experimental designs and frequentist data analyses: Motivations and examples. *Appl. Stoch. Models Bus. Ind.* **33** 302–313. MR3666933 https://doi.org/10.1002/asmb.2249

[55] Verdinelli, I. and Kadane, J. B. (1992). Bayesian designs for maximizing information and outcome. *J. Amer. Statist. Assoc.* **87** 510–515. MR1173814

[56] Villar, S. S., Robertson, D. S. and Rosenberger, W. F. (2021). The temptation of overgeneralizing response-adaptive randomization. *Clin. Infect. Dis.* **73** e842. https://doi.org/10.1093/cid/ciaa1027

[57] Ware, J. H. (1989). Investigating therapies of potentially great benefit: ECMO. *Statist. Sci.* **4** 298–340. MR1041761

[58] Wason, J. M. S. and Trippa, L. (2014). A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Stat. Med.* **33** 2206–2221. MR3256687 https://doi.org/10.1002/sim.6086

[59] Wathen, J. K. and Thall, P. F. (2017). A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clin. Trials* **14** 432–440. https://doi.org/10.1177/1740774517692302

[60] Wei, L. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.

[61] Williamson, S. F. and Villar, S. S. (2020). A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. *Biometrics* **76** 197–209. https://doi.org/10.1111/biom.13119

[62] Wong, K. (1999). Recent advances in multiple-objective design strategies. *Stat. Neerl.* **53** 257–276.

[63] Xiao, Y., Liu, Z. and Hu, F. (2017). Bayesian doubly adaptive randomization in clinical trials. *Sci. China Math.* **60** 2503–2514. MR3736374 https://doi.org/10.1007/s11425-016-0056-1

[64] Yuan, Y. and Yin, G. (2011). On the usefulness of outcome-adaptive randomization. *J. Clin. Oncol.* **29** 390–392.

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

[66] Zhu, H. and Hu, F. (2010). Sequential monitoring of response-adaptive randomized clinical trials. *Ann. Statist.* **38** 2218–2241. MR2676888 https://doi.org/10.1214/10-AOS796

[67] Zhu, X., Singla, A., Zilles, S. and Rafferty, A. N. (2018). An overview of machine teaching. ArXiv preprint. Available at arXiv:1801.05927.

Statistical Science
2023, Vol. 38, No. 1, 68–91
https://doi.org/10.1214/22-STS850

Double-Estimation-Friendly Inference for High-Dimensional Misspecified Models

Rajen D. Shah and Peter Bühlmann

Abstract. All models may be wrong—but that is not necessarily a problem for inference. Consider the standard t-test for the significance of a variable X for predicting response Y while controlling for p other covariates Z in a random design linear model. This yields correct asymptotic type I error control for the null hypothesis that X is conditionally independent of Y given Z under an *arbitrary* regression model of Y on (X, Z), provided that a linear regression model for X on Z holds. An analogous robustness to misspecification, which we term the "double-estimation-friendly" (DEF) property, also holds for Wald tests in generalised linear models, with some small modifications.

In this expository paper, we explore this phenomenon, and propose methodology for high-dimensional regression settings that respects the DEF property. We advocate specifying (sparse) generalised linear regression models for both Y and the covariate of interest X; our framework gives valid inference for the conditional independence null if either of these hold. In the special case where both specifications are linear, our proposal amounts to a small modification of the popular debiased Lasso test. We also investigate constructing confidence intervals for the regression coefficient of X via inverting our tests; these have coverage guarantees even in partially linear models where the contribution of Z to Y can be arbitrary. Numerical experiments demonstrate the effectiveness of the methodology.

Key words and phrases: Conditional independence, high-dimensional inference, Debiased Lasso, generalised linear models, double robustness.

REFERENCES

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324 https://doi.org/10.1093/biomet/asr043

BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799. MR0431440

BRADIC, J., WAGER, S. and ZHU, Y. (2019). Sparsity double robust inference of average treatment effects. Preprint. Available at arXiv:1905.00744.

BRILLINGER, D. R. (1983). A generalized linear model with "Gaussian" regressor variables. In *A Festschrift for Erich L. Lehmann*. *Wadsworth Statist./Probab. Ser.* 97–114. Wadsworth, Belmont, CA. MR0689741

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

BÜHLMANN, P. and VAN DE GEER, S. (2015). High-dimensional inference in misspecified linear models. *Electron. J. Stat.* **9** 1449–1473. MR3367666 https://doi.org/10.1214/15-EJS1041

BUJA, A., BROWN, L., BERK, R., GEORGE, E., PITKIN, E., TRASKIN, M., ZHANG, K. and ZHAO, L. (2019a). Models as approximations I: Consequences illustrated with linear regression. *Statist. Sci.* **34** 523–544. MR4048582 https://doi.org/10.1214/18-STS693

BUJA, A., BROWN, L., KUCHIBHOTLA, A. K., BERK, R., GEORGE, E. and ZHAO, L. (2019b). Models as approximations II: A model-free theory of parametric regression. *Statist. Sci.* **34** 545–565. MR4048583 https://doi.org/10.1214/18-STS694

CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled vari-

Rajen D. Shah is Professor of Statistics, Statistical Laboratory, University of Cambridge, Cambridge, United Kingdom (e-mail: r.shah@statslab.cam.ac.uk). Peter Bühlmann is Professor of Statistics, Seminar for Statistics, ETH Zurich, Zurich, Switzerland (e-mail: buehlmann@stat.math.ethz.ch).

able selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

CAO, W., TSIATIS, A. A. and DAVIDIAN, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96** 723–734. MR2538768 https://doi.org/10.1093/biomet/asp033

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

BÜHLMANN, P., KALISCH, M. and MEIER, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annu. Rev. Stat. Appl.* **1** 255–278.

DEZEURE, R., BÜHLMANN, P., MEIER, L. and MEINSHAUSEN, N. (2015). High-dimensional inference: Confidence intervals, *p*-values and R-software `hdi`. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

DUAN, N. and LI, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530. MR1105834 https://doi.org/10.1214/aos/1176348109

DUKES, O., AVAGYAN, V. and VANSTEELANDT, S. (2020). Doubly robust tests of exposure effects under high-dimensional confounding. *Biometrics* **76** 1190–1200. MR4186835 https://doi.org/10.1111/biom.13231

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

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley*, *Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

JANKOVÁ, J., SHAH, R. D., BÜHLMANN, P. and SAMWORTH, R. J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 773–795. MR4112784 https://doi.org/10.1111/rssb.12371

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Stat.* **40** 633–643. MR0238419 https://doi.org/10.1214/aoms/1177697731

KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227

LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052. MR1015136 https://doi.org/10.1214/aos/1176347254

MACKINNON, J. G. and WHITE, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econometrics* **29** 305–325.

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

ROBINS, J. M., MARK, S. D. and NEWEY, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48** 479–495. MR1173493 https://doi.org/10.2307/2532304

ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

ROTNITZKY, A., LEI, Q., SUED, M. and ROBINS, J. M. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika* **99** 439–456. MR2931264 https://doi.org/10.1093/biomet/ass013

SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478 https://doi.org/10.2307/2669923

SHAH, R. D. and BÜHLMANN, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 113–135. MR3744714 https://doi.org/10.1111/rssb.12234

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. Preprint. Available at arXiv:1904.03737.

SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166 https://doi.org/10.1093/biomet/ass043

SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled lasso. *J. Mach. Learn. Res.* **14** 3385–3418. MR3144466

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

VAN DE GEER, S. (2016). *Estimation and Testing Under Sparsity*. *Lecture Notes in Math.* **2159**. Springer, Cham. MR3526202 https://doi.org/10.1007/978-3-319-32774-7

VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316 https://doi.org/10.1214/09-EJS506

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

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

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

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

ZHU, Y. and BRADIC, J. (2018a). Significance testing in non-sparse high-dimensional linear models. *Electron. J. Stat.* **12** 3312–3364. MR3861831 https://doi.org/10.1214/18-EJS1443

ZHU, Y. and BRADIC, J. (2018b). Linear hypothesis testing in dense high-dimensional linear models. *J. Amer. Statist. Assoc.* **113** 1583–1600. MR3902231 https://doi.org/10.1080/01621459.2017.1356319

Statistical Science
2023, Vol. 38, No. 1, 92–107
https://doi.org/10.1214/22-STS851

Randomization-Based Test for Censored Outcomes: A New Look at the Logrank Test

Xinran Li and Dylan S. Small

Abstract. Two-sample tests with censored outcomes are a classical topic in statistics with wide use even in cutting edge applications. There are at least two modes of inference used to justify two-sample tests. One is usual superpopulation inference assuming that units are independent and identically distributed (i.i.d.) samples from some superpopulation; the other is finite population inference that relies on the random assignments of units into different groups. When randomization is actually implemented, the latter has the advantage of avoiding distributional assumptions on the outcomes. In this paper, we focus on finite population inference for censored outcomes, which has been less explored in the literature. Moreover, we allow the censoring time to depend on treatment assignment, under which exact permutation inference is unachievable. We find that, surprisingly, the usual logrank test can also be justified by randomization. Specifically, under a Bernoulli randomized experiment with noninformative i.i.d. censoring, the logrank test is asymptotically valid for testing Fisher's null hypothesis of no treatment effect on any unit. The asymptotic validity of the logrank test does not require any distributional assumption on the potential event times. We further extend the theory to the stratified logrank test, which is useful for randomized block designs and when censoring mechanisms vary across strata. In sum, the developed theory for the logrank test from finite population inference supplements its classical theory from usual superpopulation inference, and helps provide a broader justification for the logrank test.

Key words and phrases: Potential outcome, potential censoring time, design-based inference, noninformative censoring, stratified logrank test.

REFERENCES

[1] AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726. MR0491547

[2] ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica* **88** 265–296. MR4084709 https://doi.org/10.3982/ecta12675

[3] ATHEY, S. and IMBENS, G. W. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments* (A. Banerjee and E. Duflo, eds.) **1** 73–140, 3. North-Holland, Amsterdam.

[4] BASSE, G., DING, Y. and TOULIS, P. (2022). Minimax designs for causal effects in temporal experiments with treatment habituation. *Biometrika*. To appear.

[5] BOJINOV, I. and SHEPHARD, N. (2019). Time series experiments and causal estimands: Exact randomization tests and trading. *J. Amer. Statist. Assoc.* **114** 1665–1682. MR4047291 https://doi.org/10.1080/01621459.2018.1527225

[6] BOJINOV, I., SIMCHI-LEVI, D. and ZHAO, J. (2020). Design and analysis of switchback experiments. Available at SSRN 3684168.

[7] BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2140250

[8] BROWN, M. (1984). On the choice of variance for the log rank test. *Biometrika* **71** 65–74. MR0738327 https://doi.org/10.1093/biomet/71.1.65

[9] CAI, J., SEN, P. K. and ZHOU, H. (1999). A random effects model for multivariate failure time data from multicenter clinical trials. *Biometrics* **55** 182–189.

[10] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

[11] DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 727–753. MR3382595 https://doi.org/10.1111/rssb.12085

Xinran Li is Assistant Professor, Department of Statistics, University of Illinois, Champaign, Illinois 61820, USA (e-mail: xinranli@illinois.edu). Dylan S. Small is Professor, Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu).

[12] DING, P. (2017). A paradox from randomization-based causal inference. *Statist. Sci.* **32** 331–345. MR3695995 https://doi.org/10.1214/16-STS571

[13] FERNÁNDEZ, T., GRETTON, A., RINDT, D. and SEJDINOVIC, D. (2021). A kernel log-rank test of independence for right-censored data. *J. Amer. Statist. Assoc.* To appear.

[14] FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, Edinburgh, London.

[15] FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1100924

[16] FREEDMAN, D. A. (2008). Randomization does not justify logistic regression. *Statist. Sci.* **23** 237–249. MR2516822 https://doi.org/10.1214/08-STS262

[17] FREEDMAN, D. A. (2008). On regression adjustments in experiments with several treatments. *Ann. Appl. Stat.* **2** 176–196. MR2415599 https://doi.org/10.1214/07-AOAS143

[18] FREEDMAN, D. A. (2008). On regression adjustments to experimental data. *Adv. in Appl. Math.* **40** 180–193. MR2388610 https://doi.org/10.1016/j.aam.2006.12.003

[19] GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52** 203–223. MR0207130 https://doi.org/10.1093/biomet/52.1-2.203

[20] GILL, R. D. (1980). *Censoring and Stochastic Integrals. Mathematical Centre Tracts* **124**. Mathematisch Centrum, Amsterdam. MR0596815

[21] GUO, K. and BASSE, G. (2021). The generalized Oaxaca–Blinder estimator. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2021.1941053

[22] HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 361–374. MR0125612

[23] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application. Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

[24] HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference: What If.* CRC Press/CRC, Boca Raton.

[25] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

[26] KOU, S. G. and YING, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statist. Sinica* **6** 809–829. MR1422405

[27] LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks. Holden-Day Series in Probability and Statistics*. Holden-Day, Inc., San Francisco, CA. MR0395032

[28] LEUNG, K. M., ELASHOFF, R. M. and AFIFI, A. A. (1997). Censoring issues in survival analysis. *Annu. Rev. Public Health* **18** 83–104. https://doi.org/10.1146/annurev.publhealth.18.1.83

[29] LI, X. and SMALL, D. S. (2023). Supplement to "Randomization-based test for censored outcomes: A new look at the logrank test." https://doi.org/10.1214/22-STS851SUPP

[30] LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

[31] LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

[32] LITTLE, R. J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *J. Amer. Statist. Assoc.* **99** 546–556. MR2109316 https://doi.org/10.1198/016214504000000467

[33] MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50** 163–170.

[34] MANTEL, N. and HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22** 719–748.

[35] NEYMAN, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9 (translated). Reprinted ed. *Statist. Sci.* **5** 465–472. MR1092986

[36] PAPADOGEORGOU, G., IMAI, K., LYALL, J. and LI, F. (2022). Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in Iraq. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear.

[37] PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179. MR0497517 https://doi.org/10.1093/biomet/65.1.167

[38] ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

[39] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[40] ROSENBERGER, W. F. and LACHIN, J. M. (2016). *Randomization in Clinical Trials: Theory and Practice*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3443072 https://doi.org/10.1002/9781118742112

[41] RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

[42] THE SOLVD INVESTIGATORS (1990). Studies of left ventricular dysfunction (SOLVD)—Rationale, design and methods: Two trials that evaluate the effect of enalapril in patients with reduced ejection fraction. *Am. J. Cardiol.* **66** 315–322.

[43] THE SOLVD INVESTIGATORS (1991). Effect of Enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N. Engl. J. Med.* **325** 293–302.

[44] TSIATIS, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA* **72** 20–22. MR0356425 https://doi.org/10.1073/pnas.72.1.20

[45] VATUTIN, V. A. and MIKHAILOV, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory Probab. Appl.* **27** 734–743.

[46] ZHANG, Y. and ROSENBERGER, W. F. (2005). On asymptotic normality of the randomization-based logrank test. *J. Nonparametr. Stat.* **17** 833–839. MR2180368 https://doi.org/10.1080/10485250500270826

Statistical Science
2023, Vol. 38, No. 1, 108–119
https://doi.org/10.1214/22-STS852

The Poisson Binomial Distribution—Old & New

Wenpin Tang and Fengmin Tang

Abstract. This is an expository article on the Poisson binomial distribution. We review lesser known results and recent progress on this topic, including geometry of polynomials and distribution learning. We also provide examples to illustrate the use of the Poisson binomial machinery. Some open questions of approximating rational fractions of the Poisson binomial are presented.

Key words and phrases: Distribution learning, geometry of polynomials, Poisson binomial distribution, Poisson/normal approximation, stochastic ordering, strong Rayleigh property.

REFERENCES

[1] AFFANDI, R. H., FOX, E., ADAMS, R. and TASKAR, B. (2014). Learning the parameters of determinantal point process kernels. In *International Conference on Machine Learning* 1224–1232.

[2] AISSEN, M., EDREI, A., SCHOENBERG, I. J. and WHITNEY, A. (1951). On the generating functions of totally positive sequences. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **37** 303–307. MR0041897 https://doi.org/10.1073/pnas.37.5.303

[3] AISSEN, M., SCHOENBERG, I. J. and WHITNEY, A. M. (1952). On the generating functions of totally positive sequences. I. *J*. *Anal*. *Math*. **2** 93–103. MR0053174 https://doi.org/10.1007/BF02786970

[4] ANARI, N., GHARAN, S. O. and REZAEI, A. (2016). Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In *Conference on Learning Theory* 103–115.

[5] ANDERSON, T. W. and SAMUELS, S. M. (1967). Some inequalities among binomial and Poisson probabilities. In *Proc*. *Fifth Berkeley Sympos*. *Math*. *Statist*. *and Probability* (*Berkeley*, *Calif*., 1965/66) 1–12. Univ. California Press, Berkeley, CA. MR0214176

[6] ANDO, T. (1987). Totally positive matrices. *Linear Algebra Appl*. **90** 165–219. MR0884118 https://doi.org/10.1016/0024-3795(87)90313-2

[7] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math*. *J*. (2) **19** 357–367. MR0221571 https://doi.org/10.2748/tmj/1178243286

[8] BARBOUR, A. D. and HALL, P. (1984). On the rate of Poisson convergence. *Math*. *Proc*. *Cambridge Philos*. *Soc*. **95** 473–480. MR0755837 https://doi.org/10.1017/S0305004100061806

[9] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. *Oxford Studies in Probability* **2**. The Clarendon Press, Oxford University Press, New York. MR1163825

[10] BEREND, D. and TASSA, T. (2010). Improved bounds on Bell numbers and on moments of sums of random variables. *Probab*. *Math*. *Statist*. **30** 185–205. MR2792580

[11] BICKEL, P. J. and VAN ZWET, W. R. (1980). On a theorem of Hoeffding. In *Asymptotic Theory of Statistical Tests and Estimation* (*Proc*. *Adv*. *Internat*. *Sympos*., *Univ*. *North Carolina*, *Chapel Hill*, *N*.*C*., 1979) 307–324. Academic Press, New York-Toronto, Ont. MR0571346

[12] BIEHLER, R. (1879). Sur une classe d'équations algébriques dont toutes les racines sont réelles. *J*. *Reine Angew*. *Math*. **87** 350–352. MR1579799 https://doi.org/10.1515/crll.1879.87.350

[13] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[14] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann*. *Statist*. **25** 970–981. MR1447736 https://doi.org/10.1214/aos/1069362733

[15] BISCARRI, W., ZHAO, S. D. and BRUNNER, R. J. (2018). A simple and fast method for computing the Poisson binomial distribution function. *Comput*. *Statist*. *Data Anal*. **122** 92–100. MR3765817 https://doi.org/10.1016/j.csda.2018.01.007

[16] BOBKOV, S. G. (2018). Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probab*. *Theory Related Fields* **170** 229–262. MR3748324 https://doi.org/10.1007/s00440-017-0756-2

[17] BOLAND, P. J. (2007). The probability distribution for the number of successes in independent trials. *Comm*. *Statist*. *Theory Methods* **36** 1327–1331. MR2396538 https://doi.org/10.1080/03610920601077014

[18] BOLAND, P. J. and PROSCHAN, F. (1983). The reliability of k out of n systems. *Ann*. *Probab*. **11** 760–764. MR0704562

[19] BOLAND, P. J., SINGH, H. and CUKIC, B. (2002). Stochastic orders in partition and random testing of software. *J*. *Appl*. *Probab*. **39** 555–565. MR1928890 https://doi.org/10.1239/jap/1034082127

[20] BOLAND, P. J., SINGH, H. and CUKIC, B. (2004). The stochastic precedence ordering with applications in sampling and testing. *J*. *Appl*. *Probab*. **41** 73–82. MR2036272 https://doi.org/10.1239/jap/1077134668

Wenpin Tang is Assistant Professor, Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027, USA (e-mail: wt2319@columbia.edu). Fengmin Tang is Graduate student, Institute for Computational & Mathematical Engineering, Stanford University, Stanford, California 94305, USA (e-mail: tfmin127@stanford.edu).

[21] Borcea, J. and Brändén, P. (2008). Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products. *Duke Math. J.* **143** 205–223. MR2420507 https://doi.org/10.1215/00127094-2008-018

[22] Borcea, J. and Brändén, P. (2009). Pólya–Schur master theorems for circular domains and their boundaries. *Ann. of Math.* (2) **170** 465–492. MR2521123 https://doi.org/10.4007/annals.2009.170.465

[23] Borcea, J., Brändén, P. and Liggett, T. M. (2009). Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* **22** 521–567. MR2476782 https://doi.org/10.1090/S0894-0347-08-00618-8

[24] Brenti, F. (1989). Unimodal, log-concave and Pólya frequency sequences in combinatorics. *Mem. Amer. Math. Soc.* **81** viii+106. MR0963833 https://doi.org/10.1090/memo/0413

[25] Broderick, T., Pitman, J. and Jordan, M. I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836. MR3150470 https://doi.org/10.1214/13-BA823

[26] Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106. MR2121796 https://doi.org/10.1214/154957805100000096

[27] Chen, L. H. Y. and Shao, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Related Fields* **120** 236–254. MR1841329 https://doi.org/10.1007/PL00008782

[28] Chen, L. H. Y. and Shao, Q.-M. (2005). Stein's method for normal approximation. In *An Introduction to Stein's Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 1–59. Singapore Univ. Press, Singapore. MR2235448 https://doi.org/10.1142/9789812567680_0001

[29] Chen, S. X. and Liu, J. S. (1997). Statistical applications of the Poisson-binomial and conditional Bernoulli distributions. *Statist. Sinica* **7** 875–892. MR1488647

[30] Chen, X.-H., Dempster, A. P. and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika* **81** 457–469. MR1311090 https://doi.org/10.1093/biomet/81.3.457

[31] Choi, K. P. and Xia, A. (2002). Approximating the number of successes in independent trials: Binomial versus Poisson. *Ann. Appl. Probab.* **12** 1139–1148. MR1936586 https://doi.org/10.1214/aoap/1037125856

[32] Darroch, J. N. (1964). On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* **35** 1317–1321. MR0164359 https://doi.org/10.1214/aoms/1177703287

[33] Daskalakis, C., Diakonikolas, I., O'Donnell, R., Servedio, R. A. and Tan, L.-Y. (2013). Learning sums of independent integer random variables. In 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science—FOCS* 2013 217–226. IEEE Computer Soc., Los Alamitos, CA. MR3246223 https://doi.org/10.1109/FOCS.2013.31

[34] Daskalakis, C., Diakonikolas, I. and Servedio, R. A. (2015). Learning Poisson binomial distributions. *Algorithmica* **72** 316–357. MR3332935 https://doi.org/10.1007/s00453-015-9971-3

[35] Daskalakis, C. and Papadimitriou, C. (2015). Sparse covers for sums of indicators. *Probab. Theory Related Fields* **162** 679–705. MR3383340 https://doi.org/10.1007/s00440-014-0582-8

[36] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation. Springer Series in Statistics*. Springer, New York. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

[37] Diakonikolas, I., Kane, D. M. and Stewart, A. (2016). Properly learning Poisson binomial distributions in almost polynomial time. In *Conference on Learning Theory* 850–878.

[38] Diakonikolas, I., Kane, D. M. and Stewart, A. (2016). The Fourier transform of Poisson multinomial distributions and its algorithmic applications. In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1060–1073. ACM, New York. MR3536636 https://doi.org/10.1145/2897518.2897552

[39] Diakonikolas, I., Kane, D. M. and Stewart, A. (2016). Optimal learning via the Fourier transform for sums of independent integer random variables. In *Conference on Learning Theory* 831–849.

[40] Duffie, D., Saita, L. and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *J. Financ. Econ.* **83** 635–665.

[41] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16. MR1093412 https://doi.org/10.1016/0167-7152(91)90170-V

[42] Fallat, S., Johnson, C. R. and Sokal, A. D. (2017). Total positivity of sums, Hadamard products and Hadamard powers: Results and counterexamples. *Linear Algebra Appl.* **520** 242–259. MR3611466 https://doi.org/10.1016/j.laa.2017.01.013

[43] Fernández, M. and Williams, S. (2010). Closed-form expression for the Poisson-binomial probability density function. *IEEE Trans. Aerosp. Electron. Syst.* **46** 803–817.

[44] Fernandez, M. F. and Aridgides, T. (2003). Measures for evaluating sea mine identification processing performance and the enhancements provided by fusing multisensor/multiprocess data via an M-out-of-N voting scheme. In *Detection and Remediation Technologies for Mines and Minelike Targets VIII* **5089** 425–436.

[45] Gail, M. H., Lubin, J. H. and Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* **68** 703–707. MR0637792 https://doi.org/10.1093/biomet/68.3.703

[46] Gasca, M. and Peña, J. M. (1992). Total positivity and Neville elimination. *Linear Algebra Appl.* **165** 25–44. MR1149743 https://doi.org/10.1016/0024-3795(92)90226-Z

[47] Ghosh, S., Liggett, T. M. and Pemantle, R. (2017). Multivariate CLT follows from strong Rayleigh property. In 2017 *Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics* (*ANALCO*) 139–147. SIAM, Philadelphia, PA. MR3630942 https://doi.org/10.1137/1.9781611974775.14

[48] Ghosh, S. and Rigollet, P. (2020). Gaussian determinantal processes: A new model for directionality in data. *Proc. Natl. Acad. Sci. USA* **117** 13207–13213. MR4240249 https://doi.org/10.1073/pnas.1917151117

[49] Gleser, L. J. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* **3** 182–188. MR0365651 https://doi.org/10.1214/aop/1176996461

[50] Goldstein, L. (2010). Bounds on the constant in the mean central limit theorem. *Ann. Probab.* **38** 1672–1689. MR2663641 https://doi.org/10.1214/10-AOP527

[51] Handelman, D. (2013). Arguments of zeros of highly log concave polynomials. *Rocky Mountain J. Math.* **43** 149–177. MR3065459 https://doi.org/10.1216/RMJ-2013-43-1-149

[52] Harper, L. H. (1967). Stirling behavior is asymptotically normal. *Ann. Math. Stat.* **38** 410–414. MR0211432 https://doi.org/10.1214/aoms/1177698956

[53] Hermite, C. (1856). Extrait d'une lettre de Mr. Ch. Hermite de Paris à Mr. Borchardt de Berlin sur le nombre des

racines d'une équation algébrique comprises entre des limites données. *J. Reine Angew. Math.* **52** 39–51. MR1578969 https://doi.org/10.1515/crll.1856.52.39

[54] HILLION, E. and JOHNSON, O. (2017). A proof of the Shepp–Olkin entropy concavity conjecture. *Bernoulli* **23** 3638–3649. MR3654818 https://doi.org/10.3150/16-BEJ860

[55] HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* **27** 713–721. MR0080391 https://doi.org/10.1214/aoms/1177728178

[56] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[57] HOLTZ, O. and TYAGLOV, M. (2012). Structured matrices, continued fractions, and root localization of polynomials. *SIAM Rev.* **54** 421–509. MR2966723 https://doi.org/10.1137/090781127

[58] HONG, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Comput. Statist. Data Anal.* **59** 41–51. MR3000040 https://doi.org/10.1016/j.csda.2012.10.006

[59] HONG, Y., MEEKER, W. Q. and MCCALLEY, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Ann. Appl. Stat.* **3** 857–879. MR2750685 https://doi.org/10.1214/00-AOAS231

[60] HUTCHINSON, J. I. (1923). On a remarkable class of entire functions. *Trans. Amer. Math. Soc.* **25** 325–332. MR1501248 https://doi.org/10.2307/1989293

[61] JANSON, S. (1994). Coupling and Poisson approximation. *Acta Appl. Math.* **34** 7–15. MR1273843 https://doi.org/10.1007/BF00994254

[62] JOGDEO, K. and SAMUELS, S. M. (1968). Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. *Ann. Math. Stat.* **39** 1191–1195.

[63] KARR, A. F. (1993). *Probability*. *Springer Texts in Statistics*. Springer, New York. MR1231974 https://doi.org/10.1007/978-1-4612-0891-4

[64] KATKOVA, O. M. and VISHNYAKOVA, A. M. (2008). A sufficient condition for a polynomial to be stable. *J. Math. Anal. Appl.* **347** 81–89. MR2433826 https://doi.org/10.1016/j.jmaa.2008.05.079

[65] KOSTOV, V. P. and SHAPIRO, B. (2013). Hardy–Petrovitch–Hutchinson's problem and partial theta function. *Duke Math. J.* **162** 825–861. MR3047467 https://doi.org/10.1215/00127094-2087264

[66] KOU, S. G. and YING, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statist. Sinica* **6** 809–829. MR1422405

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

[68] KURTZ, D. C. (1992). A sufficient condition for all the roots of a polynomial to be real. *Amer. Math. Monthly* **99** 259–263. MR1216215 https://doi.org/10.2307/2325063

[69] LAVANCIER, F., MØLLER, J. and RUBAK, E. (2015). Determinantal point process models and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 853–877. MR3382600 https://doi.org/10.1111/rssb.12096

[70] LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197. MR0142174

[71] LIGGETT, T. M. (2021). Approximating multiples of strong Rayleigh random variables. Available at https://celebratio.org/Liggett_T/article/858/.

[72] MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2759813 https://doi.org/10.1007/978-0-387-68276-1

[73] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems and the geometry of polynomials. Available at arXiv:1908.09020.

[74] MICHELEN, M. and SAHASRABUDHE, J. (2019). Central limit theorems from the roots of probability generating functions. *Adv. Math.* **358** 106840. MR4021875 https://doi.org/10.1016/j.aim.2019.106840

[75] NEAMMANEE, K. (2005). On the constant in the nonuniform version of the Berry–Esseen theorem. *Int. J. Math. Math. Sci.* **12** 1951–1967. MR2176447 https://doi.org/10.1155/IJMMS.2005.1951

[76] NEDELMAN, J. and WALLENIUS, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality. *Amer. Statist.* **40** 286–289. MR0866910 https://doi.org/10.2307/2684605

[77] NOVAK, S. Y. (2019). Poisson approximation. *Probab. Surv.* **16** 228–276. MR3992498 https://doi.org/10.1214/18-PS318

[78] OVEIS GHARAN, S., SABERI, A. and SINGH, M. (2011). A randomized rounding approach to the traveling salesman problem. In 2011 *IEEE* 52*nd Annual Symposium on Foundations of Computer Science—FOCS* 2011 550–559. IEEE Computer Soc., Los Alamitos, CA. MR2932731 https://doi.org/10.1109/FOCS.2011.80

[79] PADITZ, L. (1989). On the analytical structure of the constant in the nonuniform version of the Esseen inequality. *Statistics* **20** 453–464. MR1012316 https://doi.org/10.1080/02331888908802196

[80] PEKÖZ, E. A., RÖLLIN, A., ČEKANAVIČIUS, V. and SHWARTZ, M. (2009). A three-parameter binomial approximation. *J. Appl. Probab.* **46** 1073–1085. MR2582707 https://doi.org/10.1239/jap/1261670689

[81] PEMANTLE, R. (2000). Towards a theory of negative dependence. *J. Math. Phys.* **41** 1371–1390. MR1757964 https://doi.org/10.1063/1.533200

[82] PETROV, V. V. (1965). A bound for the deviation of the distribution of a sum of independent random variables from the normal law. *Dokl. Akad. Nauk SSSR* **160** 1013–1015. MR0178497

[83] PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 82. Springer, New York. Translated from the Russian by A. A. Brown. MR0388499

[84] PITMAN, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A* **77** 279–303. MR1429082 https://doi.org/10.1006/jcta.1997.2747

[85] PLATONOV, M. L. (1979). *Combinatorial Numbers of a Class of Mappings and Their Applications*. "Nauka", Moscow. MR0555244

[86] PLEDGER, G. and PROSCHAN, F. (1971). Comparisons of order statistics and of spacings from heterogeneous distributions. In *Optimizing Methods in Statistics* (*Proc. Sympos.*, *Ohio State Univ.*, *Columbus*, *Ohio*, 1971) 89–113. MR0341738

[87] POISSON, S. D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile*. Bachelier.

[88] RIO, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 https://doi.org/10.1214/08-AIHP187

[89] RÖLLIN, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. MR2358635 https://doi.org/10.1214/105051607000000258

[90] ROOS, B. (1999). Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. *Bernoulli* **5** 1021–1034. MR1735783 https://doi.org/10.2307/3318558

[91] ROOS, B. (2000). Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Teor. Veroyatn. Primen.* **45** 328–344. MR1967760 https://doi.org/10.1137/S0040585X9797821X

[92] ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

[93] ROSENMAN, E. (2019). Some new results for Poisson binomial models. Available at arXiv:1907.09053.

[94] SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Stat.* **36** 1272–1278. MR0179825 https://doi.org/10.1214/aoms/1177699998

[95] SCHUMACHER, N. (1999). Binomial option pricing with nonidentically distributed returns and its implications. *Math. Comput. Modelling* **29** 121–143. MR1704770 https://doi.org/10.1016/S0895-7177(99)00097-7

[96] SHEPP, L. A. and OLKIN, I. (1981). Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to Probability* 201–206. Academic Press, New York. MR0618689

[97] SHIGANOV, I. S. (1982). Refinement of the upper bound of a constant in the remainder term of the central limit theorem. In *Stability Problems for Stochastic Models* (*Moscow*, 1982) 109–115. Vsesoyuz. Nauchno-Issled. Inst. Sistem. Issled., Moscow. MR0734011

[98] SKIPPER, M. (2012). A Pólya approximation to the Poisson-binomial law. *J. Appl. Probab.* **49** 745–757. MR3012097 https://doi.org/10.1239/jap/1346955331

[99] SOSHNIKOV, A. (2002). Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** 171–187. MR1894104 https://doi.org/10.1214/aop/1020107764

[100] STANLEY, R. P. (1989). Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In *Graph Theory and Its Applications*: *East and West* (*Jinan*, 1986). *Ann. New York Acad. Sci.* **576** 500–535. New York Acad. Sci., New York. MR1110850 https://doi.org/10.1111/j.1749-6632.1989.tb16434.x

[101] STEIN, C. (1990). Application of Newton's identities to a generalized birthday problem and to the Poisson binomial distribution. Technical Report 354, Department of Statistics, Stanford University.

[102] TANG, W. and TANG, F. (2023). Supplement to "The Poisson binomial distribution—Old & New." https://doi.org/10.1214/22-STS852SUPP

[103] TEJADA, A. and ARNOLD, J. (2011). The role of Poisson's binomial distribution in the analysis of TEM images. *Ultramicroscopy* **111** 1553–1556.

[104] THONGTHA, P. and NEAMMANEE, K. (2007). Refinement on the constants in the non-uniform version of the Berry–Esseen theorem. *Thai J. Math.* **5** 1–13. MR2407441

[105] VAN BEEK, P. (1972). An application of Fourier methods to the problem of sharpening the Berry–Esseen inequality. *Z. Wahrsch. Verw. Gebiete* **23** 187–196. MR0329000 https://doi.org/10.1007/BF00536558

[106] VATUTIN, V. A. and MIKHAĬLOV, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for the distribution of particles by groups. *Theory Probab. Appl.* **27** 734–743. MR0681461

[107] WANG, Y. H. (1993). On the number of successes in independent trials. *Statist. Sinica* **3** 295–312. MR1243388

[108] XU, M. and BALAKRISHNAN, N. (2011). On the convolution of heterogeneous Bernoulli random variables. *J. Appl. Probab.* **48** 877–884. MR2884823 https://doi.org/10.1017/s0021900200008391

[109] ZHANG, C.-H. (1999). Sub-Bernoulli functions, moment inequalities and strong laws for nonnegative and symmetrized U-statistics. *Ann. Probab.* **27** 432–453. MR1681165 https://doi.org/10.1214/aop/1022677268

Statistical Science
2023, Vol. 38, No. 1, 120–139
https://doi.org/10.1214/22-STS863

Stein's Method Meets Computational Statistics: A Review of Some Recent Developments

Andreas Anastasiou, Alessandro Barp, François-Xavier Briol, Bruno Ebner, Robert E. Gaunt, Fatemeh Ghaderinezhad, Jackson Gorham, Arthur Gretton, Christophe Ley, Qiang Liu, Lester Mackey, Chris J. Oates, Gesine Reinert and Yvik Swan

Abstract. Stein's method compares probability distributions through the study of a class of linear operators called Stein operators. While mainly studied in probability and used to underpin theoretical statistics, Stein's method has led to significant advances in computational statistics in recent years. The goal of this survey is to bring together some of these recent developments, and in doing so, to stimulate further research into the successful field of Stein's method and statistics. The topics we discuss include tools to benchmark and compare sampling methods such as approximate Markov chain Monte Carlo, deterministic alternatives to sampling methods, control variate techniques, parameter estimation and goodness-of-fit testing.

Key words and phrases: Stein's method, sample quality, approximate Markov chain Monte Carlo, variational inference, control variates, goodness-of-fit testing, maximum likelihood estimator, likelihood ratio, prior sensitivity.

REFERENCES

[1] AHN, S., KORATTIKARA, A. and WELLING, M. (2012). Bayesian posterior sampling via stochastic gradient Fisher scoring. In *International Conference on Machine Learning* (*ICML*) 1591–1598.

[2] ALLISON, J. S., BETSCH, S., EBNER, B. and VISAGIE, I. J. H. (2022). On testing the adequacy of the inverse Gaussian distribution. *Mathematics* **10** 350.

[3] ANASTASIOU, A. (2017). Bounds for the normal approximation of the maximum likelihood estimator from m-dependent random variables. *Statist*. *Probab. Lett*. **129** 171–181. MR3688530 https://doi.org/10.1016/j.spl.2017.04.022

[4] ANASTASIOU, A. and GAUNT, R. E. (2021). Wasserstein distance error bounds for the multivariate normal approximation of the maximum likelihood estimator. *Electron*. *J. Stat*. **15** 5758–5810. MR4355697 https://doi.org/10.1214/21-ejs1920

[5] ANASTASIOU, A. and LEY, C. (2017). Bounds for the asymptotic normality of the maximum likelihood estimator using the

Andreas Anastasiou is a Lecturer, Department of Mathematics and Statistics, University of Cyprus, P.O. Box: 20537, 1678 Nicosia, Cyprus (e-mail: anastasiou.andreas@ucy.ac.cy). Alessandro Barp is a Research Associate, University of Cambridge, Engineering Dept, Trumpington St, Cambridge CB2 1PZ, UK (e-mail: ab2286@cam.ac.uk). François-Xavier Briol is a Lecturer, University College London, 1-19 Torrington Place, London WC1E 7HB, UK (e-mail: f.briol@ucl.ac.uk). Bruno Ebner is a Lecturer, Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, 76128 Karlsruhe, Germany (e-mail: Bruno.Ebner@kit.edu). Robert E. Gaunt is a Lecturer, The University of Manchester, Oxford Road, Manchester M13 9PL, UK (e-mail: robert.gaunt@manchester.ac.uk). Fatemeh Ghaderinezhad is a Data Analyst, amfori, The Gradient Building, Avenue de Tervueren 270, 1150 Brussels, Belgium (e-mail: Fatemeh.Ghaderinezhad@amfori.org). Jackson Gorham is a Data Scientist, Whisper.ai, Inc., USA (e-mail: jacksongorham@gmail.com). Arthur Gretton is Professor with the Gatsby Computational Neuroscience Unit, University College London, Sainsbury Wellcome Centre, 25 Howland Street, London W1T 4JG, UK (e-mail: arthur.gretton@gmail.com). Christophe Ley is Associate Professor, University of Luxembourg, Maison du Nombre, 6 Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg (e-mail: christophe.ley@uni.lu). Qiang Liu is Assistant Professor, The University of Texas at Austin, Austin, Texas 78712, USA (e-mail: lqiang@utexas.edu). Lester Mackey is Principal Researcher, Microsoft Research New England, 1 Memorial Drive, Cambridge, Massachusetts 02142, USA (e-mail: lmackey@microsoft.com). Chris J. Oates is Professor, Newcastle University, UK (e-mail: Chris.Oates@newcastle.ac.uk). Gesine Reinert is Professor, University of Oxford, Department of Statistics, 24-29 St Giles', Oxford OX1 3LB, UK (e-mail: gesine.reinert@keble.ox.ac.uk). Yvik Swan is Professor, Université Libre de Bruxelles, Department of Mathematics—CP 210, Boulevard du Triomphe, 1050 Brussels, Belgium (e-mail: Yvik.Swan@ulb.be).

delta method. *ALEA Lat. Am. J. Probab. Math. Stat.* **14** 153–171. MR3622464

[6] ANASTASIOU, A. and REINERT, G. (2017). Bounds for the normal approximation of the maximum likelihood estimator. *Bernoulli* **23** 191–218. MR3556771 https://doi.org/10.3150/15-BEJ741

[7] ANASTASIOU, A. and REINERT, G. (2020). Bounds for the asymptotic distribution of the likelihood ratio. *Ann. Appl. Probab.* **30** 608–643. MR4108117 https://doi.org/10.1214/19-AAP1510

[8] ANDRADÓTTIR, S., HEYMAN, D. P. and OTT, T. J. (1993). Variance reduction through smoothing and control variates for Markov chain simulations. *ACM Trans. Model. Comput. Simul.* **3** 167–189.

[9] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437 https://doi.org/10.2307/1990404

[10] ARRAS, B. and HOUDRÉ, C. (2019). *On Stein's Method for Infinitely Divisible Laws with Finite First Moment. SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham. MR3931309

[11] ASSARAF, R. and CAFFAREL, M. (1999). Zero-variance principle for Monte Carlo algorithms. *Phys. Rev. Lett.* **83** 4682.

[12] BANERJEE, T., LIU, Q., MUKHERJEE, G. and SUN, W. (2021). A general framework for empirical Bayes estimation in discrete linear exponential family. *J. Mach. Learn. Res.* **22** 67. MR4253760

[13] BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184.

[14] BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659 https://doi.org/10.1007/BF01197887

[15] BARBOUR, A. D. and CHEN, L. H. Y. (2014). Stein's (magic) method. ArXiv preprint. Available at arXiv:1411.1179.

[16] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. The Clarendon Press, New York. MR1163825

[17] BARBOUR, A. D. and XIA, A. (1999). Poisson perturbations. *ESAIM Probab. Stat.* **3** 131–150. MR1716120 https://doi.org/10.1051/ps:1999106

[18] BARINGHAUS, L. and HENZE, N. (1991). A class of consistent tests for exponentiality based on the empirical Laplace transform. *Ann. Inst. Statist. Math.* **43** 551–564. MR1143640 https://doi.org/10.1007/BF00053372

[19] BARINGHAUS, L. and HENZE, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statist. Probab. Lett.* **13** 269–274. MR1160747 https://doi.org/10.1016/0167-7152(92)90033-2

[20] BARP, A. A. (2020). The Bracket Geometry of Statistics Ph.D. thesis Imperial College London.

[21] BARP, A. A., BRIOL, F. X., DUNCAN, A. B., GIROLAMI, M. and MACKEY, L. (2019). Minimum Stein discrepancy estimators. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 12964–12976.

[22] BARP, A. A., OATES, C., PORCU, E. and GIROLAMI, M. (2018). A Riemannian-Stein kernel method. ArXiv preprint. Available at arXiv:1810.04946.

[23] BELOMESTNY, D., IOSIPOI, L., MOULINES, E., NAUMOV, A. and SAMSONOV, S. (2020). Variance reduction for Markov chains with application to MCMC. *Stat. Comput.* **30** 973–997. MR4108687 https://doi.org/10.1007/s11222-020-09931-z

[24] BELOMESTNY, D., IOSIPOI, L. and ZHIVOTOVSKIY, N. (2017). Variance reduction via empirical variance minimization: Convergence and complexity. ArXiv preprint. Available at arXiv:1712.04667.

[25] BELOMESTNY, D., MOULINES, E., SHAGADATOV, N. and URUSOV, M. (2019). Variance reduction for MCMC methods via martingale representations. ArXiv preprint. Available at arXiv:1903.07373.

[26] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

[27] BETSCH, S. and EBNER, B. (2019). A new characterization of the Gamma distribution and associated goodness-of-fit tests. *Metrika* **82** 779–806. MR4008662 https://doi.org/10.1007/s00184-019-00708-7

[28] BETSCH, S. and EBNER, B. (2020). Testing normality via a distributional fixed point property in the Stein characterization. *TEST* **29** 105–138. MR4063385 https://doi.org/10.1007/s11749-019-00630-0

[29] BETSCH, S. and EBNER, B. (2021). Fixed point characterizations of continuous univariate probability distributions and their applications. *Ann. Inst. Statist. Math.* **73** 31–59. MR4205241 https://doi.org/10.1007/s10463-019-00735-1

[30] BETSCH, S., EBNER, B. and KLAR, B. (2021). Minimum L^q-distance estimators for non-normalized parametric models. *Canad. J. Statist.* **49** 514–548. MR4267931 https://doi.org/10.1002/cjs.11574

[31] BETSCH, S., EBNER, B. and NESTMANN, F. (2022). Characterizations of non-normalized discrete probability distributions and their application in statistics. *Electron. J. Stat.* **16** 1303–1329. MR4381061 https://doi.org/10.1214/22-ejs1983

[32] CARMELI, C., DE VITO, E., TOIGO, A. and UMANITÀ, V. (2010). Vector valued reproducing kernel Hilbert spaces and universality. *Anal. Appl.* (*Singap.*) **8** 19–61. MR2603770 https://doi.org/10.1142/S0219530510001503

[33] CHATTERJEE, S. (2014). A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians—Seoul* 2014. *Vol. IV* 1–24. Kyung Moon Sa, Seoul. MR3727600

[34] CHEN, C., ZHANG, R., WANG, W., LI, B. and CHEN, L. (2018). A unified particle-optimization framework for scalable Bayesian sampling. In *Uncertainty in Artificial Intelligence* (*UAI*).

[35] CHEN, L. H. and RÖLLIN, A. (2010). Stein couplings for normal approximation. ArXiv preprint. Available at arXiv:1003.6039.

[36] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387 https://doi.org/10.1214/aop/1176996359

[37] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Springer, Heidelberg. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

[38] CHEN, P., WU, K., CHEN, J., O'LEARY-ROSEBERRY, T. and GHATTAS, O. (2019). Projected Stein variational Newton: A fast and scalable Bayesian inference method in high dimensions. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 15130–15139.

[39] CHEN, W. Y., BARP, A. A., BRIOL, F.-X., GORHAM, J., GIROLAMI, M., MACKEY, L. and OATES, C. J. (2019). Stein point Markov chain Monte Carlo. In *International Conference on Machine Learning* (*ICML*) 1011–1021.

[40] CHEN, W. Y., MACKEY, L., GORHAM, J., BRIOL, F.-X. and OATES, C. J. (2018). Stein points. In *International Conference on Machine Learning* (*ICML*) 844–853.

[41] CHEWI, S., GOUIC, T. L., LU, C., MAUNU, T. and RIGOLLET, P. (2020). SVGD as a kernelized Wasserstein gradient flow of the chi-squared divergence. In *Advances on Neural Information Processing Systems* (*NeurIPS*).

[42] Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. In *International Conference on Machine Learning* (*ICML*) 2606–2615.

[43] Courtade, T. A., Fathi, M. and Pananjady, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898

[44] Dellaportas, P. and Kontoyiannis, I. (2012). Control variates for estimation based on reversible Markov chain Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 133–161. MR2885843 https://doi.org/10.1111/j.1467-9868.2011.01000.x

[45] Detommaso, G., Cui, T., Marzouk, Y., Scheichl, R. and Spantini, A. (2018). A Stein variational Newton method. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 9169–9179.

[46] Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates (with a discussion and a rejoinder by the authors). *Ann. Statist.* **14** 1–67. MR0829555 https://doi.org/10.1214/aos/1176349830

[47] Diaconis, P. and Holmes, S., eds. (2004). *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46**.

[48] Dörr, P., Ebner, B. and Henze, N. (2021). A new test of multivariate normality by a double estimation in a characterizing PDE. *Metrika* **84** 401–427. MR4233599 https://doi.org/10.1007/s00184-020-00795-x

[49] Duncan, A., Nüsken, N. and Szpruch, L. (2019). On the geometry of Stein variational gradient descent. ArXiv preprint. Available at arXiv:1912.00894.

[50] Ebner, B. (2021). On combining the zero bias transform and the empirical characteristic function to test normality. *ALEA Lat. Am. J. Probab. Math. Stat.* **18** 1029–1045. MR4282180 https://doi.org/10.30757/alea.v18-38

[51] Ebner, B. and Henze, N. (2020). Tests for multivariate normality—a critical review with emphasis on weighted L^2-statistics. *TEST* **29** 845–892. MR4182841 https://doi.org/10.1007/s11749-020-00740-0

[52] Erdogdu, M. A., Mackey, L. and Shamir, O. (2018). Global non-convex optimization with discretized diffusions. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 9694–9703.

[53] Fang, X., Shao, Q.-M. and Xu, L. (2019). Multivariate approximations in Wasserstein distance by Stein's method and Bismut's formula. *Probab. Theory Related Fields* **174** 945–979. MR3980309 https://doi.org/10.1007/s00440-018-0874-5

[54] Fathi, M., Goldstein, L., Reinert, G. and Saumard, A. (2020). Relaxing the Gaussian assumption in shrinkage and SURE in high dimension. ArXiv preprint. Available at arXiv:2004.01378.

[55] Feng, Y., Wang, D. and Liu, Q. (2017). Learning to draw samples with amortized Stein variational gradient descent. In *Uncertainty in Artificial Intelligence* (*UAI*).

[56] Fernández, T., Rivera, N., Xu, W. and Gretton, A. (2020). Kernelized Stein discrepancy tests of goodness-of-fit for time-to-event data. In *International Conference on Machine Learning* (*ICML*).

[57] Fisher, M. A., Nolan, T. H., Graham, M. M., Prangle, D. and Oates, C. J. (2021). Measure transport with kernel Stein discrepancy. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*).

[58] Gaunt, R. E. (2017). On Stein's method for products of normal random variables and zero bias couplings. *Bernoulli* **23** 3311–3345. MR3654808 https://doi.org/10.3150/16-BEJ848

[59] Gaunt, R. E. (2022). Bounds for the chi-square approximation of the power divergence family of statistics. *J. Appl. Probab.*

[60] Gaunt, R. E., Pickett, A. M. and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756. MR3655852 https://doi.org/10.1214/16-AAP1213

[61] Gaunt, R. E. and Reinert, G. (2021). Bounds for the chi-square approximation of Friedman's statistic by Stein's method. ArXiv preprint. Available at arXiv:2111.00949.

[62] Ghaderinezhad, F. and Ley, C. (2019). Quantification of the impact of priors in Bayesian statistics via Stein's method. *Statist. Probab. Lett.* **146** 206–212. MR3884714 https://doi.org/10.1016/j.spl.2018.11.012

[63] Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** 419–435.

[64] Goldstein, L. and Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** 237–260. MR2132278 https://doi.org/10.1007/s10959-004-2602-6

[65] Goldstein, L. and Reinert, G. (2013). Stein's method for the beta distribution and the Pólya-Eggenberger urn. *J. Appl. Probab.* **50** 1187–1205. MR3161381 https://doi.org/10.1239/jap/1389370107

[66] Gong, C., Peng, J. and Liu, Q. (2019). Quantile Stein variational gradient descent for parallel Bayesian optimization. In *International Conference on Machine Learning* (*ICML*) 2347–2356.

[67] Gong, W., Li, Y. and Hernández-Lobato, J. M. (2021). Sliced kernelized Stein discrepancy. In *International Conference on Learning Representations* (*ICLR*).

[68] Gorham, J., Duncan, A. B., Vollmer, S. J. and Mackey, L. (2019). Measuring sample quality with diffusions. *Ann. Appl. Probab.* **29** 2884–2928. MR4019878 https://doi.org/10.1214/19-AAP1467

[69] Gorham, J. and Mackey, L. (2015). Measuring sample quality with Stein's method. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 226–234. Curran Associates, Red Hook.

[70] Gorham, J. and Mackey, L. (2017). Measuring sample quality with kernels. In *International Conference on Machine Learning* (*ICML*) 1292–1301.

[71] Gorham, J., Raj, A. and Mackey, L. (2020). Stochastic Stein discrepancies. In *Advances on Neural Information Processing Systems* (*NeurIPS*).

[72] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

[73] Grathwohl, W., Wang, K. C., Jacobsen, J. H., Duvenaud, D. and Zemel, R. (2020). Learning the Stein discrepancy for training and evaluating energy-based models without sampling. In *International Conference on Machine Learning* 9485–9499.

[74] Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B. and Smola, A. J. (2006). A kernel method for the two-sample-problem. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 513–520.

[75] 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

[76] Haarnoja, T., Tang, H., Abbeel, P. and Levine, S. (2017). Reinforcement learning with deep energy-based policies. In *International Conference on Machine Learning* (*ICML*) 1352–1361.

[77] Han, J. and Liu, Q. (2017). Stein variational adaptive importance sampling. In *Uncertainty in Artificial Intelligence* (*UAI*).

[78] HAN, J. and LIU, Q. (2018). Stein variational gradient descent without gradient. In *International Conference on Machine Learning* (*ICML*) 1900–1908.

[79] HENDERSON, S. G. and SIMON, B. (2004). Adaptive simulation using perfect control variates. *J. Appl. Probab.* **41** 859–876. MR2074828 https://doi.org/10.1017/s0021900200020593

[80] HENZE, N., MEINTANIS, S. G. and EBNER, B. (2012). Goodness-of-fit tests for the gamma distribution based on the empirical Laplace transform. *Comm. Statist. Theory Methods* **41** 1543–1556. MR3003807 https://doi.org/10.1080/03610926.2010.542851

[81] HENZE, N. and VISAGIE, J. (2020). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Ann. Inst. Statist. Math.* **72** 1109–1136. MR4137748 https://doi.org/10.1007/s10463-019-00720-8

[82] HINTON, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14** 1771–1800.

[83] HODGKINSON, L., SALOMONE, R. and ROOSTA, F. (2020). The reproducing Stein kernel approach for post-hoc corrected sampling. ArXiv preprint. Available at arXiv:2001.09266.

[84] HOLMES, S. (2004). Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46** 45–67. IMS, Beachwood, OH. MR2118602

[85] HOLMES, S. and REINERT, G. (2004). Stein's method for the bootstrap. In *Stein's Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46** 95–136. IMS, Beachwood, OH. MR2118605

[86] HU, T., CHEN, Z., SUN, H., BAI, J., YE, M. and CHENG, G. (2018). Stein neural sampler. ArXiv preprint. Available at arXiv:1810.03545.

[87] HUGGINS, J. H. and MACKEY, L. (2018). Random feature Stein discrepancies. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 1899–1909.

[88] HYVÄRINEN, A. (2005). Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **6** 695–709. MR2249836

[89] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc.* 4*th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

[90] JITKRITTUM, W., XU, W., SZABO, Z., FUKUMIZU, K. and GRETTON, A. (2017). A linear-time kernel goodness-of-fit test. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 261–270.

[91] KEY, O., FERNANDEZ, T., GRETTON, A. and BRIOL, F.-X. (2021). Composite goodness-of-fit tests with kernels. In *NeurIPS* 2021 *Workshop Your Model Is Wrong: Robustness and Misspecification in Probabilistic Modeling*. Available at arXiv:2111.10275.

[92] KIM, T., YOON, J., DIA, O., KIM, S., BENGIO, Y. and AHN, S. (2018). Bayesian model-agnostic meta-learning. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 7332–7342.

[93] KORATTIKARA, A., CHEN, Y. and WELLING, M. (2014). Austerity in MCMC land: Cutting the Metropolis-Hastings budget. In *Proceedings of International Conference on Machine Learning* (*ICML*). *ICML*'14.

[94] KORBA, A., SALIM, A., ARBEL, M., LUISE, G. and GRETTON, A. (2020). A non-asymptotic analysis for Stein variational gradient descent. In *Advances in Neural Information Processing Systems* (*NeurIPS*) **33**.

[95] KUMAR KATTUMANNIL, S. (2009). On Stein's identity and its application. *Statist. Probab. Lett.* **79** 1444–1449. MR2536504 https://doi.org/10.1016/j.spl.2009.03.021

[96] LEDOUX, M., NOURDIN, I. and PECCATI, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** 256–306. MR3320893 https://doi.org/10.1007/s00039-015-0312-0

[97] LEUCHT, A. and NEUMANN, M. H. (2013). Dependent wild bootstrap for degenerate U- and V-statistics. *J. Multivariate Anal.* **117** 257–280. MR3053547 https://doi.org/10.1016/j.jmva.2013.03.003

[98] LEY, C., REINERT, G. and SWAN, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. MR3595350 https://doi.org/10.1214/16-PS278

[99] LEY, C., REINERT, G. and SWAN, Y. (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics. *Ann. Appl. Probab.* **27** 216–241. MR3619787 https://doi.org/10.1214/16-AAP1202

[100] LEY, C. and SWAN, Y. (2016). Parametric Stein operators and variance bounds. *Braz. J. Probab. Stat.* **30** 171–195. MR3481100 https://doi.org/10.1214/14-BJPS271

[101] LI, L., LI, Y., LIU, J.-G., LIU, Z. and LU, J. (2020). A stochastic version of Stein variational gradient descent for efficient sampling. *Commun. Appl. Math. Comput. Sci.* **15** 37–63. MR4113783 https://doi.org/10.2140/camcos.2020.15.37

[102] LIPPERT, R. A., HUANG, H. and WATERMAN, M. S. (2002). Distributional regimes for the number of k-word matches between two random sequences. *Proc. Natl. Acad. Sci. USA* **99** 13980–13989. MR1944413 https://doi.org/10.1073/pnas.202468099

[103] LIU, A., LIANG, Y. and VAN DEN BROECK, G. (2020). Off-policy deep reinforcement learning with analogous disentangled exploration. In *International Conference on Autonomous Agents and Multiagent Systems* (*AAMAS*).

[104] LIU, C. and ZHU, J. (2018). Riemannian Stein variational gradient descent for Bayesian inference. In *AAAI Conference on Artificial Intelligence* 3627–3634.

[105] LIU, C., ZHUO, J., CHENG, P., ZHANG, R. and ZHU, J. (2019). Understanding and accelerating particle-based variational inference. In *International Conference on Machine Learning* (*ICML*) 4082–4092.

[106] LIU, H., FENG, Y., MAO, Y., ZHOU, D., PENG, J. and LIU, Q. (2018). Action-dependent control variates for policy optimization via Stein's identity. In *International Conference on Learning Representations* (*ICLR*).

[107] LIU, Q. (2017). Stein variational gradient descent as gradient flow. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 3115–3123.

[108] LIU, Q., LEE, J. and JORDAN, M. (2016). A kernelized Stein discrepancy for goodness-of-fit tests. In *International Conference on Machine Learning* (*ICML*) 276–284.

[109] LIU, Q. and LEE, J. D. (2017). Black-box importance sampling. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 952–961.

[110] LIU, Q., LEE, J. D. and JORDAN, M. I. (2016). A kernelized Stein discrepancy for goodness-of-fit tests and model evaluation. In *International Conference on Machine Learning* (*ICML*) 276–284.

[111] LIU, Q. and WANG, D. (2016). Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 2370–2378.

[112] LIU, Q. and WANG, D. (2018). Stein variational gradient descent as moment matching. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 8854–8863.

[113] LIU, S., KANAMORI, T., JITKRITTUM, W. and CHEN, Y. (2019). Fisher efficient inference of intractable models. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 8793–8803.

[114] LIU, Y., RAMACHANDRAN, P., LIU, Q. and PENG, J. (2017). Stein variational policy gradient. In *Uncertainty in Artificial Intelligence* (*UAI*).

[115] LU, J., LU, Y. and NOLEN, J. (2019). Scaling limit of the Stein variational gradient descent: The mean field regime. *SIAM J. Math. Anal.* **51** 648–671. MR3919409 https://doi.org/10.1137/18M1187611

[116] MACKEY, L. and GORHAM, J. (2016). Multivariate Stein factors for a class of strongly log-concave distributions. *Electron. Commun. Probab.* **21** 56. MR3548768 https://doi.org/10.1214/16-ecp15

[117] MATSUBARA, T., KNOBLAUCH, J., BRIOL, F. X. and OATES, C. J. (2021). Robust generalised Bayesian inference for intractable likelihoods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*. To appear. Available at arXiv:2104.07359.

[118] MATSUBARA, T., KNOBLAUCH, J., BRIOL, F. X. and OATES, C. J. (2022). Generalised Bayesian inference for discrete intractable likelihood. Available at arXiv:2206.08420.

[119] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. Springer, London. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

[120] MIJATOVIĆ, A. and VOGRINC, J. (2018). On the Poisson equation for Metropolis-Hastings chains. *Bernoulli* **24** 2401–2428. MR3757533 https://doi.org/10.3150/17-BEJ932

[121] MIJOULE, G., REINERT, G. and SWAN, Y. (2021). Stein's density method for multivariate continuous distributions. ArXiv preprint. Available at arXiv:2101.05079.

[122] MIRA, A., SOLGI, R. and IMPARATO, D. (2013). Zero variance Markov chain Monte Carlo for Bayesian estimators. *Stat. Comput.* **23** 653–662. MR3094805 https://doi.org/10.1007/s11222-012-9344-6

[123] MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938 https://doi.org/10.2307/1428011

[124] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

[125] NÜSKEN, N. and RENGER, D. (2021). Stein variational gradient descent: Many-particle and long-time asymptotics. ArXiv preprint. Available at arXiv:2102.12956.

[126] OATES, C. J., COCKAYNE, J., BRIOL, F.-X. and GIROLAMI, M. (2019). Convergence rates for a class of estimators based on Stein's method. *Bernoulli* **25** 1141–1159. MR3920368 https://doi.org/10.3150/17-bej1016

[127] OATES, C. J., GIROLAMI, M. and CHOPIN, N. (2017). Control functionals for Monte Carlo integration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 695–718. MR3641403 https://doi.org/10.1111/rssb.12185

[128] OATES, C. J., PAPAMARKOU, T. and GIROLAMI, M. (2016). The controlled thermodynamic integral for Bayesian model evidence evaluation. *J. Amer. Statist. Assoc.* **111** 634–645. MR3538693 https://doi.org/10.1080/01621459.2015.1021006

[129] OKSENDAL, B. (2013). *Stochastic Differential Equations*: *An Introduction with Applications*, 6th ed. Springer, Berlin.

[130] PU, Y., GAN, Z., HENAO, R., LI, C., HAN, S. and CARIN, L. (2017). VAE learning via Stein variational gradient descent. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 4236–4245.

[131] RACHEV, S. T., KLEBANOV, L. B., STOYANOV, S. V. and FABOZZI, F. J. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. Springer, New York. MR3024835 https://doi.org/10.1007/978-1-4614-4869-3

[132] RANGANATH, R., TRAN, D., ALTOSAAR, J. and BLEI, D. (2016). Operator variational inference. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 496–504.

[133] REINERT, G. (1998). Couplings for normal approximations with Stein's method. In *Microsurveys in Discrete Probability* (*Princeton*, *NJ*, 1997). *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **41** 193–207. Amer. Math. Soc., Providence, RI. MR1630415 https://doi.org/10.1089/cmb.1998.5.223

[134] REINERT, G. (2005). Three general approaches to Stein's method. In *An Introduction to Stein's Method*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 183–221. Singapore Univ. Press, Singapore. MR2235451 https://doi.org/10.1142/9789812567680_0004

[135] REINERT, G., CHEW, D., SUN, F. and WATERMAN, M. S. (2009). Alignment-free sequence comparison. I. Statistics and power. *J. Comput. Biol.* **16** 1615–1634. MR2578699 https://doi.org/10.1089/cmb.2009.0198

[136] REINERT, G. and ROSS, N. (2019). Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. *Ann. Appl. Probab.* **29** 3201–3229. MR4019886 https://doi.org/10.1214/19-AAP1478

[137] RIABIZ, M., CHEN, W., COCKAYNE, J., SWIETACH, P., NIEDERER, S. A., MACKEY, L. and OATES, C. (2020). Optimal thinning of MCMC output. ArXiv preprint. Available at arXiv:2005.03952.

[138] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

[139] SCHWARTZ, L. (1964). Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *J. Anal. Math.* **13** 115–256. MR0179587 https://doi.org/10.1007/BF02786620

[140] SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics* **162**. Wiley, New York.

[141] SHAO, Q.-M. (2005). An explicit Berry-Esseen bound for Student's t-statistic via Stein's method. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 143–155. Singapore Univ. Press, Singapore. MR2205333 https://doi.org/10.1142/9789812567673_0009

[142] SHAO, Q.-M. (2010). Stein's method, self-normalized limit theory and applications. In *Proceedings of the International Congress of Mathematicians*. *Volume IV* 2325–2350. Hindustan Book Agency, New Delhi. MR2827974

[143] SHAO, Q.-M., ZHANG, K. and ZHOU, W.-X. (2016). Stein's method for nonlinear statistics: A brief survey and recent progress. *J. Statist. Plann. Inference* **168** 68–89. MR3412222 https://doi.org/10.1016/j.jspi.2015.06.008

[144] SI, S., OATES, C. J., DUNCAN, A. B., CARIN, L. and BRIOL, F.-X. (2020). Scalable control variates for Monte Carlo methods via stochastic optimization. ArXiv preprint. Available at arXiv:2006.07487.

[145] SMOLA, A., GRETTON, A., SONG, L. and SCHÖLKOPF, B. (2007). A Hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory* 13–31.

[146] SOHL-DICKSTEIN, J., BATTAGLINO, P. and DEWEESE, M. R. (2011). Minimum probability flow learning. In *International Conference on Machine Learning* 905–912.

[147] SOUTH, L. F., KARVONEN, T., NEMETH, C., GIROLAMI, M. and OATES, C. (2020). Semi-exact control functionals from sard's method. ArXiv preprint. Available at arXiv:2002.00033.

[148] SOUTH, L. F., OATES, C. J., MIRA, A. and DROVANDI, C. (2018). Regularised zero-variance control variates for high-dimensional variance reduction. ArXiv preprint. Available at arXiv:1811.05073.

[149] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley-Los Angeles, CA. MR0084922

[150] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 583–602. MR0402873

[151] STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

[152] STEIN, C., DIACONIS, P., HOLMES, S. and REINERT, G. (2004). Use of exchangeable pairs in the analysis of simulations. In *Stein's Method*: *Expository Lectures and Applications*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46** 1–26. IMS, Beachwood, OH. MR2118600 https://doi.org/10.1214/lnms/1196283797

[153] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[154] SUN, Z., BARP, A. and BRIOL, F.-X. (2021). Vector-valued control variates. Available at arXiv:2109.08944.

[155] TEYMUR, O., GORHAM, J., RIABIZ, M. and OATES, C. (2021). Optimal quantisation of probability measures using maximum mean discrepancy. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 1027–1035.

[156] TIHOMIROV, A. N. (1980). Convergence rate in the central limit theorem for weakly dependent random variables. *Teor. Veroyatn. Primen.* **25** 800–818. MR0595140

[157] WANG, D. and LIU, Q. (2016). Learning to draw samples: With application to amortized MLE for generative adversarial learning. ArXiv preprint. Available at arXiv:1611.01722.

[158] WANG, D. and LIU, Q. (2019). Nonlinear Stein variational gradient descent for learning diversified mixture models. In *International Conference on Machine Learning* (*ICML*) 6576–6585.

[159] WANG, D., TANG, Z., BAJAJ, C. and LIU, Q. (2019). Stein variational gradient descent with matrix-valued kernels. In *Advances on Neural Information Processing Systems* (*NeurIPS*) 7834–7844.

[160] WANG, D., ZENG, Z. and LIU, Q. (2018). Stein variational message passing for continuous graphical models. In *International Conference on Machine Learning* (*ICML*) 5219–5227.

[161] WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *International Conference on Machine Learning* (*ICML*) 681–688.

[162] XU, W. (2022). Standardisation-function kernel Stein discrepancy: A unifying view on kernel Stein discrepancy tests for goodness-of-fit. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 1575–1597.

[163] XU, W. and REINERT, G. (2021). A Stein goodness-of-fit test for exponential random graph models. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 415–423.

[164] YANG, J., LIU, Q., RAO, V. and NEVILLE, J. (2018). Goodness-of-fit testing for discrete distributions via Stein discrepancy. In *International Conference on Machine Learning* (*ICML*) 5561–5570.

[165] YANG, J., RAO, V. and NEVILLE, J. (2019). A Stein–papangelou goodness-of-fit test for point processes. In *International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 226–235.

[166] YANG, Z., BALASUBRAMANIAN, K., WANG, Z. and LIU, H. (2017). Learning non-Gaussian multi-index model via second-order Stein's method. In *Advances in Neural Information Processing Systems* (*NeurIPS*) **30** 6097–6106.

[167] ZHANG, X. and CURTIS, A. (2019). Seismic tomography using variational inference methods. *J. Geophys. Res.*, *Solid Earth* **125** e2019JB018589.

[168] ZHANG, X. and CURTIS, A. (2020). Variational full-waveform inversion. *Geophys. J. Int.* **222** 406–411.

[169] ZHANG, Y. and LEE, A. A. (2019). Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning. *Chem. Sci.* **10** 8154–8163. https://doi.org/10.1039/c9sc00616h

[170] ZHU, Y. and ZABARAS, N. (2018). Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. *J. Comput. Phys.* **366** 415–447. MR3800689 https://doi.org/10.1016/j.jcp.2018.04.018

[171] ZHU, Z., WAN, R. and ZHONG, M. (2018). Neural control variates for variance reduction. ArXiv preprint. Available at arXiv:1806.00159.

[172] ZHUO, J., LIU, C., SHI, J., ZHU, J., CHEN, N. and ZHANG, B. (2018). Message passing Stein variational gradient descent. In *International Conference on Machine Learning* (*ICML*) 6013–6022.

[173] ZOLOTAREV, V. M. (1984). Probability metrics. *Theory Probab. Appl.* **28** 278–302.

Statistical Science
2023, Vol. 38, No. 1, 140–159
https://doi.org/10.1214/22-STS864

Local scale invariance and robustness of proper scoring rules

David Bolin and Jonas Wallin

Abstract. Averages of proper scoring rules are often used to rank probabilistic forecasts. In many cases, the individual terms in these averages are based on observations and forecasts from different distributions. We show that some of the most popular proper scoring rules, such as the continuous ranked probability score (CRPS), give more importance to observations with large uncertainty, which can lead to unintuitive rankings. To describe this issue, we define the concept of local scale invariance for scoring rules. A new class of generalized proper kernel scoring rules is derived and as a member of this class we propose the scaled CRPS (SCRPS). This new proper scoring rule is locally scale invariant and, therefore, works in the case of varying uncertainty. Like the CRPS, it is computationally available for output from ensemble forecasts, and does not require the ability to evaluate densities of forecasts.

We further define robustness of scoring rules, show why this also can be an important concept for average scores unless one is specifically interested in extremes, and derive new proper scoring rules that are robust against outliers. The theoretical findings are illustrated in three different applications from spatial statistics, stochastic volatility models and regression for count data.

Key words and phrases: Probabilistic forecasting, model selection, spatial statistics, forecast ranking.

REFERENCES

BARAN, S. and LERCH, S. (2016). Mixture EMOS model for calibrating ensemble forecasts of wind speed. *Environmetrics* **27** 116–130. MR3481324 https://doi.org/10.1002/env.2380

BERG, C., CHRISTENSEN, J. P. R. and RESSEL, P. (1984). *Harmonic Analysis on Semigroups*: *Theory of positive definite and related functions*. *Graduate Texts in Mathematics* **100**. Springer, New York. MR0747302 https://doi.org/10.1007/978-1-4612-1128-0

BERGHAUSER PONT, M., STAVROULAKI, G. and MARCUS, L. (2019). Development of urban types based on network centrality, built density and their impact on pedestrian movement. *Environ*. *Plan*. *B Urban Anal*. *City Sci*. **46** 1549–1564.

BERGHAUSER PONT, M., BOLIN, D., HÅKANSSON, E., IVARSSON, O., STAVROULAKI, G. and VERENDEL, V. (2019). stepflow – R-Shiny interface for pedestrian flow data and models. http://129.16.20.138:3838/stepflow/stepflow/, retrieved on January 24, 2022.

BERNARDO, J.-M. (1979). Expected information as expected utility. *Ann*. *Statist*. **7** 686–690. MR0527503

BESSAC, J. and NAVEAU, P. (2021). Forecast score distributions with imperfect observations. *Adv*. *Stat*. *Climatol*. *Meteorol*. *Oceanogr*. **7** 53–71.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

BRIER, G. W. et al. (1950). Verification of forecasts expressed in terms of probability. *Mon*. *Weather Rev*. **78** 1–3.

BRÖCKER, J. (2012). Evaluating raw ensembles with the continuous ranked probability score. *Q*. *J*. *R*. *Meteorol*. *Soc*. **138** 1611–1617.

CAMPBELL, S. D. and DIEBOLD, F. X. (2005). Weather forecasting for weather derivatives. *J*. *Amer*. *Statist*. *Assoc*. **100** 6–16. MR2166065 https://doi.org/10.1198/016214504000001051

CANDILLE, G. and TALAGRAND, O. (2005). Evaluation of probabilistic prediction systems for a scalar variable. *Q*. *J*. *R*. *Meteorol*. *Soc*. **131** 2131–2150.

DAWID, A. P. (1998). Coherent Measures of Discrepancy, Uncertainty and Dependence, with Applications to Bayesian Predictive Experimental Design Technical Report No. 139.

DAWID, A. P. (2007). The geometry of proper scoring rules. *Ann*. *Inst*. *Statist*. *Math*. **59** 77–93. MR2396033 https://doi.org/10.1007/s10463-006-0099-8

DAWID, A. P. and MUSIO, M. (2014). Theory and applications of proper scoring rules. *Metron* **72** 169–183. MR3233147 https://doi.org/10.1007/s40300-014-0039-y

DAWID, A. P., MUSIO, M. and VENTURA, L. (2016). Minimum scoring rule inference. *Scand*. *J*. *Stat*. **43** 123–138. MR3466997 https://doi.org/10.1111/sjos.12168

David Bolin is Associate Professor, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia (e-mail: david.bolin@kaust.edu.sa). Jonas Wallin is Associate Professor, Lund University, Lund, Sweden (e-mail: jonas.wallin@stat.lu.se).

DAWID, A. P. and SEBASTIANI, P. (1999). Coherent dispersion criteria for optimal experimental design. *Ann. Statist.* **27** 65–81. MR1701101 https://doi.org/10.1214/aos/1018031101

DESCAMPS, L., LABADIE, C., JOLY, A., BAZILE, E., ARBOGAST, P. and CÉBRON, P. (2015). PEARP, the Météo-France short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **141** 1671–1685.

DIEBOLD, F. X. and MARIANO, R. S. (1995). Comparing predictive accuracy. *J. Bus. Econom. Statist.* **13** 253–263.

EFRON, B. (1991). Regression percentiles using asymmetric squared error loss. *Statist. Sinica* **1** 93–125. MR1101317

FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spat. Stat.* **14** 505–531. MR3431054 https://doi.org/10.1016/j.spasta.2015.10.001

GARRATT, A., LEE, K., PESARAN, M. H. and SHIN, Y. (2003). Forecast uncertainties in macroeconomic modeling: An application to the U.K. economy. *J. Amer. Statist. Assoc.* **98** 829–838. MR2055491 https://doi.org/10.1198/016214503000000765

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 https://doi.org/10.1111/j.1467-9868.2007.00587.x

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GNEITING, T. and RANJAN, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *J. Bus. Econom. Statist.* **29** 411–422. MR2848512 https://doi.org/10.1198/jbes.2010.08110

GOOD, I. J. (1952). Rational decisions. *J. Roy. Statist. Soc. Ser. B* **14** 107–114. MR0077033

HAGELIN, S., SON, J., SWINBANK, R., MCCABE, A., ROBERTS, N. and TENNANT, W. (2017). The Met Office convective-scale ensemble, MOGREPS-UK. *Q. J. R. Meteorol. Soc.* **143** 2846–2861.

HAIDEN, T., JANOUSEK, M., VITART, F., FERRANTI, L. and PRATES, F. (2019). Evaluation of ECMWF forecasts, including the 2019 upgrade Technical Memo No. 853 ECMWF. https://doi.org/10.21957/mlvapkke

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A Case Study Competition Among Methods for Analyzing Large Spatial Data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

HILLIER, B., PENN, A., HANSON, J., GRAJEWSKI, T. and XU, J. (1993). Natural movement: Or, configuration and attraction in urban pedestrian movement. *Environ. Plan. B, Plan. Des.* **20** 29–66.

HYVÄRINEN, A. (2005). Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **6** 695–709. MR2249836

INGEBRIGTSEN, R., LINDGREN, F., STEINSLAND, I. and MARTINO, S. (2015). Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. *Spat. Stat.* **14** 338–364. MR3431045 https://doi.org/10.1016/j.spasta.2015.07.003

JUUTILAINEN, I., TAMMINEN, S. and RÖNING, J. (2012). Exceedance probability score: A novel measure for comparing probabilistic predictions. *J. Stat. Theory Pract.* **6** 452–467. MR3196559 https://doi.org/10.1080/15598608.2012.695663

LEHMANN, E. L. (1997). *Theory of Point Estimation*. Springer, New York. Reprint of the 1983 original. MR1451376

LERCH, S. and THORARINSDOTTIR, T. L. (2013). Comparison of non-homogeneous regression models for probabilistic wind speed forecasting. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* **65** 21206.

LERCH, S., THORARINSDOTTIR, T. L., RAVAZZOLO, F. and GNEITING, T. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statist. Sci.* **32** 106–127. MR3634309 https://doi.org/10.1214/16-STS588

DEGROOT, M. H. and FIENBERG, S. E. (1983). The comparison and evaluation of forecasters. *J. R. Stat. Soc., Ser. D, Stat.* **32** 12–22.

MOYEED, R. A. and PAPRITZ, A. (2002). An empirical comparison of kriging methods for nonlinear spatial point prediction. *Math. Geol.* **34** 365–386. MR1951788 https://doi.org/10.1023/A:1015085810154

MURPHY, A. H. (1972). Scalar and vector partitions of the probability score: Part I. Two-state situation. *J. Appl. Meteorol.* **11** 273–282.

MURPHY, A. H. (1973). Hedging and Skill Scores for Probability Forecasts. *J. Appl. Meteorol.* **12** 215–223.

NOLDE, N. and ZIEGEL, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.* **11** 1833–1874. MR3743276 https://doi.org/10.1214/17-AOAS1041

NOWOTARSKI, J. and WERON, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* **81** 1548–1568.

OPSCHOOR, A., VAN DIJK, D. and VAN DER WEL, M. (2017). Combining density forecasts using focused scoring rules. *J. Appl. Econometrics* **32** 1298–1313. MR3734489 https://doi.org/10.1002/jae.2575

PALMER, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Q. J. R. Meteorol. Soc.* **128** 747–774.

PARRY, M., DAWID, A. P. and LAURITZEN, S. (2012). Proper local scoring rules. *Ann. Statist.* **40** 561–592. MR3014317 https://doi.org/10.1214/12-AOS971

PATTON, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *J. Econometrics* **160** 246–256. MR2745881 https://doi.org/10.1016/j.jeconom.2010.03.034

ROULSTON, M. S. and SMITH, L. A. (2003). Combining dynamical and statistical ensembles. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* **55** 16–30.

SELTEN, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Exp. Econ.* **1** 43–61.

SHEPHARD, N. (1994). Partial non-Gaussian state space. *Biometrika* **81** 115–131. MR1279661 https://doi.org/10.1093/biomet/81.1.115

STAVROULAKI, G., BOLIN, D., BERGHAUSER PONT, M., MARCUS, L. and HÅKANSSON, E. (2019). Statistical Modelling and Analysis of Big Data on Pedestrian Movement. In *Proceedings of the* 12*th Space Syntax Symposium* 1–24.

TAILLARDAT, M., FOUGÈRES, A.-L., NAVEAU, P. and DE FONDEVILLE, R. (2019). Extreme events evaluation using CRPS distributions. Preprint. Available at arXiv:1905.04022.

TÖDTER, J. and AHRENS, B. (2012). Generalization of the ignorance score: Continuous ranked version and its decomposition. *Mon. Weather Rev.* **140** 2005–2017.

VENABLES, W. N. and RIPLEY, B. D. (2002). *Modern Applied Statistics with S*, 4th ed. Springer, New York.

WILKS, D. S. (2005). *Statistical Methods in the Atmospheric Sciences: An Introduction*. Elsevier Science and Technology, Burlington.

WINKLER, R. L. (1996). Scoring rules and the evaluation of probabilities. *TEST* **5** 1–60. With comments and a rejoinder by the author. MR1410455 https://doi.org/10.1007/BF02562681

ZIMMERMAN, D. L. and STEIN, M. (2010). Classical Geostatistical Methods. In *Handbook of Spatial Statistics*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 517–539. CRC Press, Boca Raton, FL. MR2730964 https://doi.org/10.1201/9781420072884-c29

Statistical Science
2023, Vol. 38, No. 1, 160–183
https://doi.org/10.1214/22-STS866

In Praise (and Search) of J. V. Uspensky

Persi Diaconis and Sandy Zabell

Abstract. The two of us have shared a fascination with James Victor Uspensky's 1937 textbook *Introduction to Mathematical Probability* ever since our graduate student days: it contains many interesting results not found in other books on the same subject in the English language, together with many non-trivial examples, all clearly stated with careful proofs. We present some of Uspensky's gems to a modern audience hoping to tempt others to read Uspensky for themselves, as well as report on a few of the other mathematical topics he also wrote about (e.g., his book on number theory contains early results about perfect shuffles).

Uspensky led an interesting life: a member of the Russian Academy of Sciences, he spoke at the 1924 International Congress of Mathematicians in Toronto before leaving Russia in 1929 and coming to the US and Stanford. Comparatively little has been written about him in English; the second half of this paper attempts to remedy this.

Key words and phrases: J. V. Uspensky, A. A. Markov, probability theory, Markov's method of continued fractions, Bernoulli's theorem, Lexis ratio, card shuffling, Russian mathematics, Stanford Mathematics Department.

REFERENCES

ALEXANDERSON, G. L. (2011). Harold M. Bacon. In *Fascinating Mathematical People*: *Interviews and Memoirs* (D. J. Albers and G. L. Alexanderson, eds.) Chapter 3. Princeton Univ. Press, Princeton, NJ.

BAHADUR, R. R. (1960). Some approximations to the binomial distribution function. *Ann*. *Math*. *Stat*. **31** 43–54. MR0120675 https://doi.org/10.1214/aoms/1177705986

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann*. *Math*. *Stat*. **37** 577–580. MR0189095 https://doi.org/10.1214/aoms/1177699450

BEATTY, S. (1926). Problem 3173. *Amer. Math. Monthly* **33** 159.

BEATTY, S., OSTROWSKI, A., HYSLOP, J. and AITKEN, A. C. (1927). Problems and Solutions: Solutions: 3177. *Amer. Math*. *Monthly* **34** 159–160. MR1521129 https://doi.org/10.2307/2298716

BELL, E. T. (1951). Hans Frederick Blichfeldt 1873–1945. In *Biographical Memoirs* **26** 181–189. National Academy of Science, Washington, DC.

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J*. *Amer*. *Statist*. *Assoc*. **57** 33–45.

BERNOULLI, J. (1913). *Ars Conjectandi*, *Part* 4. Imp. Akaemiya Nauk, St. Petersburg. Translated from the Latin by Ya. V. Uspenskii, with an introduction by A. A. Markov. Reprinted 1986, with additional notes by Yu. V. Prokhorov.

BERNSTEIN, S. N. (1924). On a variant of Chebyshev's inequality and on the accuracy of Laplace's formula (in Russian). *Uč*. *Zap*. *Nauchn*.-*Issled*. *Kafedr Ukrainy*, *Otd*. *Mat*. **1** 38–48.

BERTRAND, J. (1888). *Calcul des Probabilités*. Gauthier-Villars, Paris. Reprinted 1889, 1907. Reprinted, 1972 by Chelsea, New York.

BHATIA, R. (2008). A conversation with S. R. S. Varadhan. *Math*. *Intelligencer* **30** 24–42. MR2410123 https://doi.org/10.1007/BF02985734

BLACKWELL, D. (1954). On optimal systems. *Ann*. *Math*. *Stat*. **25** 394–397. MR0061776 https://doi.org/10.1214/aoms/1177728798

BLICHFELDT, H. F. (1900). On a certain class of groups of transformation in space of three dimensions. *Amer*. *J*. *Math*. **22** 113–120. MR1505825 https://doi.org/10.2307/2369749

BOHR, H. and MOLLERUP, J. (1922). *Laerebog i Matematisk Analyse* **3**. In Danish.

BORWEIN, J. M., CHOI, K.-K. S. and PIGULLA, W. (2005). Continued fractions of tails of hypergeometric series. *Amer*. *Math*. *Monthly* **112** 493–501. MR2142600 https://doi.org/10.2307/30037519

BURNSIDE, W. (1928). *Theory of Probability*. Cambridge Univ. Press, Cambridge. 2nd ed., 1936.

BUTTER, F. A. JR. (1945). Recent Publications: Aircraft Analytic Geometry. *Amer. Math. Monthly* **52** 338–340. MR1526233 https://doi.org/10.2307/2305297

CASSELS, J. W. S. and VAUGHAN, R. C. (1985). Obituary: Ivan Matveevich Vinogradov. *Bull*. *Lond*. *Math*. *Soc*. **17** 584–600. MR0813744 https://doi.org/10.1112/blms/17.6.584

CHAMBERLAIN, L. (2006). *Lenin's Private War*. St. Martin's Press, New York.

COCHRAN, W. G. (1952). The χ^2 test of goodness of fit. *Ann*. *Math*. *Stat*. **23** 315–345. MR0049531 https://doi.org/10.1214/aoms/1177729380

Persi Diaconis is the Mary V. Sunseri Professor of Statistics and Mathematics, Stanford University, Stanford, California 94305, USA (e-mail: diaconis@math.stanford.edu). Sandy Zabell is Professor, Departments of Mathematics and Statistics, Northwestern University, Evanston, Illinois 60208, USA (e-mail: zabell@math.northwestern.edu).

COOLIDGE, J. L. (1925). *An Introduction to Mathematical Probability*. Clarendon Press, Oxford.

CRAIG, C. C. (1933). On the Tchebychef inequality of Bernstein. *Ann. Math. Stat.* **4** 94–102.

CRAMÉR, H. (1976). Half a century with probability theory: Some personal recollections. *Ann. Probab.* **4** 509–546. MR0402837 https://doi.org/10.1214/aop/1176996025

DAVID, F. N. (1938). Review of introduction to mathematical probability by J. V. Uspensky. *Biometrika* **30** 194–195.

DE MOIVRE, A. (1967). *The Doctrine of Chances*: *A Method of Calculating the Probabilities of Events in Play*. Chelsea Publishing Co., New York. MR0216923

DEHEUVELS, P. and PFEIFER, D. (1988). On a relationship between Uspensky's theorem and Poisson approximations. *Ann. Inst. Statist. Math.* **40** 671–681. MR0996692 https://doi.org/10.1007/BF00049425

DEMIDOV, S. S. (2015). World War I and mathematics in "the Russian world". *Czas. Tech.* **112** 77–92.

DESALVO, S. (2021). Will the real Hardy–Ramanujan formula please stand up? *Integers* **21** A88. MR4304355

DEWITT, N. (1961). *Education and Professional Employment in the U.S.S.R.* National Science Foundation, Washington, DC.

DIACONIS, P. and GRAHAM, R. (2012). *Magical Mathematics*: *The Mathematical Ideas That Animate Great Magic Tricks*. Princeton Univ. Press, Princeton, NJ. MR2858033

DIACONIS, P., GRAHAM, R. L. and KANTOR, W. M. (1983). The mathematics of perfect shuffles. *Adv. in Appl. Math.* **4** 175–196. MR0700845 https://doi.org/10.1016/0196-8858(83)90009-X

DIACONIS, P. and ZABELL, S. (1991). Closed form summation for classical distributions: Variations on a theme of de Moivre. *Statist. Sci.* **6** 284–302. MR1144242

DICKSON, L. E. (1947). Hans Frederik Blichfeldt: 1873–1945. *Bull. Amer. Math. Soc.* **53** 882–883.

DUDLEY, R. M. (1987). Some inequalities for continued fractions. *Math. Comp.* **49** 585–593. MR0906191 https://doi.org/10.2307/2008331

DUTKA, J. (1981). The incomplete beta function—a historical profile. *Arch. Hist. Exact Sci.* **24** 11–29. MR0618150 https://doi.org/10.1007/BF00327713

DWYER, W. A. (1937). On Certain Fundamental Identities Due to Uspensky. *Amer. J. Math.* **59** 290–294. MR1507239 https://doi.org/10.2307/2371411

ERMOLAEVA, N. (1997). Uspensky, Yakov Viktorovich. In *Russian Diaspora*: *The Golden Book of Emigration*, *the First Third of the* 20*th Century*, *an Encyclopedic Biographical Dictionary* (V. V. Shelokhaeva and N. I. Kanishcheva, eds.) ROSSPEN, Moscow.

ETHIER, S. N. (2010). *The Doctrine of Chances*. *Probability and Its Applications* (*New York*). Springer, Berlin. MR2656351 https://doi.org/10.1007/978-3-540-78783-9

ETHIER, S. N. and KHOSHNEVISAN, D. (2002). Bounds on gambler's ruin probabilities in terms of moments. *Methodol. Comput. Appl. Probab.* **4** 55–68. MR1918933 https://doi.org/10.1023/A:1015705430513

FELLER, W. (1950). *An Introduction to Probability Theory and Its Applications*. Wiley, New York. 2nd ed. 1957, 3rd ed. 1968.

FINKEL, S. (2003). Purging the public intellectual: The 1922 expulsions from Soviet Russia. *Russ. Rev.* **62** 589–613.

FISHER, A. (1922). *The Mathematical Theory of Probabilities & Its Application to Frequency Curves & Statistical Methods*. Macmillan, New York. 1st ed., 1915.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh. 14th (posthumous) ed., 1970, Hafner, New York.

FORSYTH, C. H. (1925). Simple derivations of the formulas for the dispersion of a statistical series. *Amer. Math. Monthly* **31** 190–196.

FRY, T. C. (1965). *Probability and Its Engineering Uses*, 2nd ed. D. Van Nostrand Co., Princeton, NJ–Toronto–London. MR0181027

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

GRAHAM, R. L. (1963). On a theorem of Uspensky. *Amer. Math. Monthly* **70** 407–409. MR0148555 https://doi.org/10.2307/2311859

GRAHAM, R. L., LIN, S. and LIN, C. S. (1978). Spectra of numbers. *Math. Mag.* **51** 174–176. MR0491580 https://doi.org/10.2307/2689998

GRAHAM, R. and O'BRYANT, K. (2005). A discrete Fourier kernel and Fraenkel's tiling conjecture. *Acta Arith.* **118** 283–304. MR2168767 https://doi.org/10.4064/aa118-3-4

HALD, A. (1990). *A History of Probability and Statistics and Their Applications Before* 1750. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR1029276 https://doi.org/10.1002/0471725161

HALD, A. (1998). *A History of Mathematical Statistics from* 1750 *to* 1930. *Wiley Series in Probability and Statistics*: *Texts and References Section*. Wiley, New York. MR1619032

HARDY, G. H. and RAMANUJAN, S. (1918). Asymptotic formulaae in combinatory analysis. *Proc. Lond. Math. Soc.* **17** 75–115. MR1575586 https://doi.org/10.1112/plms/s2-17.1.75

HARDY, G. H. and WRIGHT, E. M. (1960). *The Theory of Numbers*, 4th ed. Clarendon Press, Oxford.

HARTMAN, E. P. (1970). *Adventures in Research*: *A History of Ames Research Center* 1940–1965. National Aeronautics and Space Administration, Washington, DC.

HEYDE, C. C. and SENETA, E. (1977). *I. J. Bienaymé*. *Statistical Theory Anticipated*. *Studies in the History of Mathematics and Physical Sciences*, *No*. 3. Springer, New York. MR0462888

HILLE, E. (1926). On Laguerre's series. *Proc. Natl. Acad. Sci. USA* **12** 261–265, 265–269, 348–352.

JACKSON, D. (1935). Mathematical principles in the theory of small samples. *Amer. Math. Monthly* **72** 344–364.

JACKSON, B. W. (no date). A History of the Mathematics Department at San Jose State. (accessed January 30, 2022).

JELLISON, C. (1997). The prisoner and the professor. *Stanford Mag.* **72**. March/April issue, https://stanfordmag.org/contents/the-prisoner-and-the-professor.

KAPLANSKY, I. (2013). Integers uniquely represented by certain ternary forms. In *The Mathematics of Paul Erdős I* (R. L. Graham et al., eds.) 71–79. Springer, New York.

KARATSUBA, A. A. (1981). Ivan Matveevich Vinogradov (on his ninetieth birthday). *Russian Math. Surveys* **36** 1–17.

KRANDICK, W. and MEHLHORN, K. (2006). New bounds for the Descartes method. *J. Symbolic Comput.* **41** 49–66. MR2194885 https://doi.org/10.1016/j.jsc.2005.02.004

LORENTZ, G. G. (2002). Mathematics and politics in the Soviet Union from 1928 to 1953. *J. Approx. Theory* **116** 169–223. MR1911079 https://doi.org/10.1006/jath.2002.3670

MALYSHEV, A. V. and FADDEEV, D. K. (1961). Boris Alekseevich Venkov (on his sixtieth birthday). *Uspekhi Mat. Nauk* **16** 235–240.

MARKOV, A. A. (1888). *Table des valeurs de l'intégrale* $\int_x^\infty e^{-t^2}\,dt$. Imperial Academy of Sciences, St. Petersburg.

MARKOV, A. A. (1900). *Ischislenie Veroiatnostei* (*Calculus of Probabilities*). Tipografija Imperatorskoj Akademii Nauk, Sanktpeterburg. Three subsequent editions: 2nd ed., 1908, and 3rd ed., 1913; 4th (posthumous) ed., 1924, Gosudarstvennoe Izdatel'stvo, Moscow. German translation of 2nd ed.: *Wahrscheinlichkeitsrechnung* (Heinrich Liebmann, trans.), Teubner, Leipzig and Berlin, 1912.

MARKOV, A., STEKLOV, V. and KRYLOV, N. (1921). Note on the scientific work of Petrograd University Professor Yakov Viktorovich Uspensky. *Izv. Rossiyskoy Akad. Nauk* (*Ser.* 6) **15** 4–6.

MCKAY, B. D. (1989). On Littlewood's estimate for the binomial distribution. *Adv. in Appl. Probab.* **21** 475–478. MR0997736 https://doi.org/10.2307/1427172

MILLER, G. H. (1970). Blichfeldt, Hans Frederick. In *Dictionary of Scientific Biography* (C. C. Gillispie, ed.) **2**. Charles Scribner's Son, New York. Reprinted 2008 in the ebook *Complete Dictionary of Scientific Biography*.

MOLINA, E. C. (1944). Arne Fisher, 1887–1944. *J. Amer. Statist. Assoc.* **39** 251.

MUNROE, M. E. (1951). *Theory of Probability*. McGraw-Hill, Inc., New York–Toronto–London. MR0040604

NAZAROV, A. I. and SINKEVICH, G. I. (2019). History of Leningrad mathematics in the first half of the 20th century. *ICCM Not.* **7** 47–63. MR3995130 https://doi.org/10.4310/ICCM.2019.v7.n2.a5

NEVAI, P. (1986). Géza Freud, orthogonal polynomials and Christoffel functions. A case study. *J. Approx. Theory* **48** 3–167. MR0862231 https://doi.org/10.1016/0021-9045(86)90016-X

OLDS, C. D. (1963). *Continued Fractions*. Random House, New York. MR0146146

OLDS, C. D. (1970). The simple continued fraction expansion of e. *Amer. Math. Monthly* **77** 968–974. MR1536091 https://doi.org/10.2307/2318113

ONDAR, K. O. (1981). *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. Springer, New York. Translated by Charles and Margaret Stein.

OSTROWSKI, A. M. (1950). Note on Vincent's theorem. *Ann. of Math.* (2) **52** 702–707. MR0038481 https://doi.org/10.2307/1969443

PEARSON, K. (1931). Historical note on the distribution of the standard deviations of samples of any size drawn from an indefinitely large normal parent population. *Biometrika* **23** 416–418.

PEARSON, E. (1990). *'Student': A Statistical Biography of William Sealy Gosset*. Clarendon Press, Oxford. R. L. Plackett and G. A.. Barnard, eds.

PEIZER, D. B. and PRATT, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities. I. *J. Amer. Statist. Assoc.* **63** 1416–1456. MR0235650

PERLMAN, M. D. and WICHURA, M. J. (1975). Sharpening Buffon's needle. *Amer. Statist.* **29** 157–163. MR0464479 https://doi.org/10.2307/2683484

PÓLYA, G., SZEGŐ, G. and YOUNG, D. H. (1947). Memorial Resolution: James Victor Uspensky. Stanford University Archives.

RASHDALL, H. (1895). *The Universities of Europe in the Middle Ages. Volume* 1: *Salerno—Bologna—Paris*. Clarendon Press, Oxford.

REEDS, J. A., DIFFIE, W. and FIELD, J. V. (2015). *Breaking Teleprinter Ciphers at Bletchley Park*. Wiley–IEEE Press, Piscataway, NJ.

RIETZ, H. L. (1924). *Handbook of Mathematical Statistics*. Houghton Mifflin, Boston, MA.

ROUCHÉ, E. (1888). Sur un problème relatif à la durée du jeu. *C. R. Acad. Sci.* **116** 47–49.

ROYDEN, H. (1989). The history of the mathematics department at Stanford. In *A Century of Mathematics in America, Part II* (P. Duren, ed.) 237–278. Amer. Math. Soc., Providence, RI.

ROYDEN, H., LYMAN, R., PHILLIPS, R., SUNSERI, M. and WINBIGLER, H. D. (1992). Memorial resolution: Harold M. Bacon (1907–1992). Department of Special Collections and University Archives. Available at https://purl.stanford.edu/ws360gk0273 (accessed February 29, 2020).

SANSONE, G. (1950). La formula di approssimazione asintotica dei polinomi di Tchebychef-Laguerre col procedimento di J. V. Uspensky. *Math. Z.* **53** 97–105. MR0037940 https://doi.org/10.1007/BF01162402

SEABASS, R. A. (2002). Harvard Lomax. In *MAM*2006: *Markov Anniversary Meeting* (A. N. Langville and W. J. Stewart, eds.) 169–172. The National Academies Press.

SEAL, H. (1966). The random walk of a simple risk business. *ASTIN Bull. J. IAA* **4** 19–28.

SENETA, E. (2006). Markov and the the creation of Markov chains. In *MAM*2006: *Markov Anniversary Meeting* (A. N. Langville and W. J. Stewart, eds.) 1–20. Boson Books, Raleigh, NC.

SENETA, E. (2013). A tricentenary history of the law of large numbers. *Bernoulli* **19** 1088–1121. MR3102545 https://doi.org/10.3150/12-BEJSP12

SHEYNIN, O. (2005). Jakob Bernoulli: On the Law of Large Numbers. Unpublished translation.

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

STANFORD (1992). Long-time calculus professor Harold Bacon dies at 85. Stanford University News Service.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR0852410

STRUTT, J. W. (1894). *The Theory of Sound, Volume* 1, 2nd ed. Macmillan, London and New York. (3rd Baron Rayleigh).

SVIDERSKAYA, G. E. (2002). Rodion Osievich Kuzmin.

SYLLA, E. (2006). *Jacob Bernoulli: The Art of Conjecturing*. Johns Hopkins Univ. Press, Baltimore, MD. Translation of Jacob Bernoulli's *Ars conjectandi*, with an introduction and notes, by Edith Dudley Sylla.

SZEGÖ, G. (1939). *Orthogonal Polynomials. American Mathematical Society Colloquium Publications, Vol.* 23. Amer. Math. Soc., New York. MR0000077

TIMOSHENKO, S. (1947). Typescript. Unpublished manuscript, presumably prepared shortly after Uspensky's death and in the Stanford University Archives.

TIMOSHENKO, S. (1968). *As I Remember: The Autobiography of Stephen P. Timoshenko*. Van Nostrand, Princeton, NJ. Translated from the Russian by Robert Addis.

TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability*. Macmillan, Cambridge and London. Reprinted, 1949 and 1964, Chelsea, New York. ["A work of true learning, beyond criticism"—John Maynard Keynes.].

USPENSKY, J. V. (1906). On whole numbers formed with the 5th root of unity. *Mat. Sb.* **26** 1–17.

USPENSKY, J. V. (1920). Asymptotic formulae for numerical functions which occur in the theory of the partition of numbers into summands. *Bull. Acad. Sci. Russ.* **14** 199–218.

USPENSKY, J. V. (1927). On a problem arising out of the theory of a certain game. *Amer. Math. Monthly* **34** 516–521. MR1521314 https://doi.org/10.2307/2299838

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.

USPENSKY, J. V. (1948). *Theory of Equations*. McGraw-Hill, New York.

USPENSKY, J. V. and HEASLET, M. A. (1939). *Elementary Number Theory*. McGraw-Hill, Inc., New York. MR0000236

USPENSKY, J. V. and VENKOV, B. (1928). On some new class-number relations. *Bull. Acad. Sci. Russ.* **1** 315–317.

VENKOV, B. A. and NATANSON, I. P. (1949). Rodion Osievich Kuzmin (1891–1949) (obituary). *Uspekhi Mat. Nauk* **4** 148–155.

VERGER, J. (2003). Teachers. In *A History of the University in Europe. Volume* 1: *Universities in the Middle Ages* (H. D. Ridder-Symoens, ed.) 144–168. Cambridge Univ. Press, Cambridge.

WALL, H. S. (1948). *Analytic Theory of Continued Fractions*. D. Van Nostrand, New York, NY. MR0025596

WENKOV, B. A. (1931). Über die Klassenanzahl positiver binärer quadratischer Formen. *Math*. *Z*. **33** 350–374. MR1545216 https://doi.org/10.1007/BF01174358

WYTHOFF, W. A. (1907). A modification of the game of Nim. *Nieuw Arch*. *Wiskd*. (5) **7** 199–202.

ZABELL, S. L. (2008). On student's 1908 article "The probable error of a mean". *J*. *Amer*. *Statist*. *Assoc*. **103** 1–7.

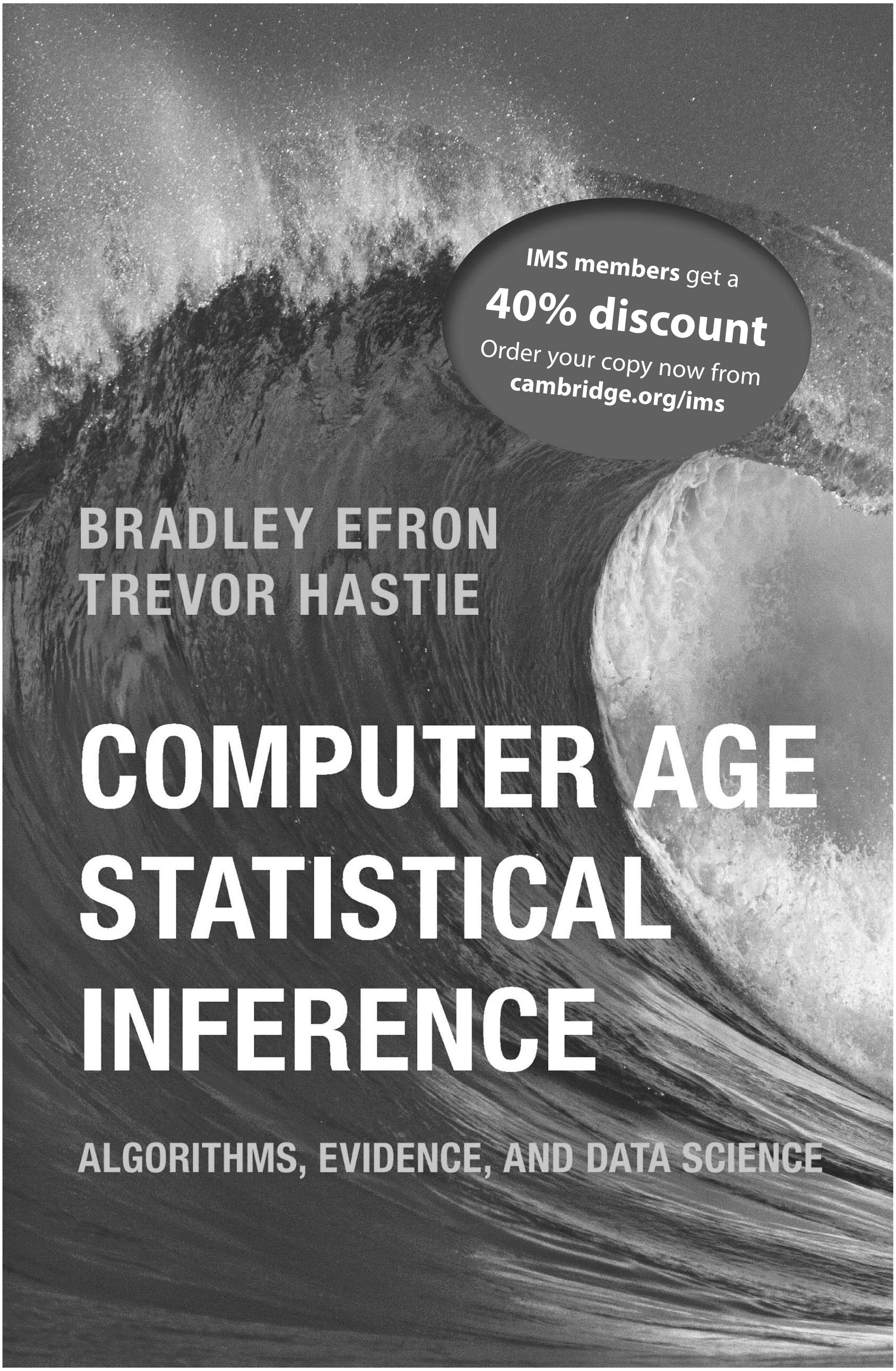
INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

IMS EDITORS



IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE