STATISTICAL SCIENCE

Volume 34, Number 1 February 2019

***Statistical Science* [ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 34, Number 1, February 2019. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, 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, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Statistical Science Volume 34, Number 1 (1–175) February 2019

Volume 34 Number 1 February 2019

Statistical Science
2019, Vol. 34, No. 1, 1–22
https://doi.org/10.1214/18-STS660

Probabilistic Integration: A Role in Statistical Computation?

François-Xavier Briol, Chris J. Oates, Mark Girolami, Michael A. Osborne and Dino Sejdinovic

Abstract. A research frontier has emerged in scientific computation, wherein discretisation error is regarded as a source of epistemic uncertainty that can be modelled. This raises several statistical challenges, including the design of statistical methods that enable the coherent propagation of probabilities through a (possibly deterministic) computational work-flow, in order to assess the impact of discretisation error on the computer output. This paper examines the case for probabilistic numerical methods in routine statistical computation. Our focus is on numerical integration, where a *probabilistic integrator* is equipped with a full distribution over its output that reflects the fact that the integrand has been discretised. Our main technical contribution is to establish, for the first time, rates of posterior contraction for one such method. Several substantial applications are provided for illustration and critical evaluation, including examples from statistical modelling, computer graphics and a computer model for an oil reservoir.

Key words and phrases: Computational statistics, nonparametric statistics, probabilistic numerics, uncertainty quantification.

REFERENCES

BAUER, B., DEVROYE, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2017). Nonparametric estimation of a function from noiseless observations at random points. *J. Multivariate Anal.* **160** 93–104. MR3688691

BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. MR2239907

BOGACHEV, V. I. (1998). *Gaussian Measures*. *Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391

BONDARENKO, A., RADCHENKO, D. and VIAZOVSKA, M. (2013). Optimal asymptotic bounds for spherical designs. *Ann. of Math.* (2) **178** 443–452. MR3071504

BRIOL, F.-X., OATES, C. J., GIROLAMI, M. and OSBORNE, M. A. (2015). Frank-Wolfe Bayesian quadrature: Probabilistic integration with theoretical guarantees. In *Proc. Adv. Neur. In.* (*NIPS*) 1162–1170.

BRIOL, F.-X., COCKAYNE, J., TEYMUR, O., YOO, W. W., SCHOBER, M. and HENNIG, P. (2016a). Contributed discussion on article by Chkrebtii, Campbell, Calderhead, and Girolami [MR3577378]. *Bayesian Anal.* **11** 1285–1293. MR3577382

BRIOL, F.-X., OATES, C. J., GIROLAMI, M., OSBORNE, M. A and SEJDINOVIC, D. (2019). Supplement to "Probabilistic integration: A role in statistical computation?" DOI:10.1214/18-STS660SUPP.

BROUILLAT, J., BOUVILLE, C., LOOS, B., HANSEN, C. and BOUATOUCH, K. (2009). A Bayesian Monte Carlo approach to global illumination. *Comput. Graph. Forum* **28** 2315–2329.

BUCHHOLZ, A. and CHOPIN, N. (2017). Improving approximate Bayesian computation via quasi Monte Carlo. arXiv:1710:01057.

CALDERHEAD, B. and GIROLAMI, M. (2009). Estimating Bayes

François-Xavier Briol is Research Associate, Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (e-mail: f.briol16@imperial.ac.uk). Chris J. Oates is Chair in Statistics, School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK (e-mail: chris.oates@ncl.ac.uk). Mark Girolami is Chair in Statistics, Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (e-mail: m.girolami@imperial.ac.uk). Michael A. Osborne is Dyson Associate Professor in Machine Learning, Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK (e-mail: mosb@robots.ox.ac.uk). Dino Sejdinovic is Associate Professor in Statistics, Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK (e-mail: dino.sejdinovic@stats.ox.ac.uk).

factors via thermodynamic integration and population MCMC. *Comput. Statist. Data Anal.* **53** 4028–4045. MR2744303

CIALENCO, I., FASSHAUER, G. E. and YE, Q. (2012). Approximation of stochastic partial differential equations by a kernel-based collocation method. *Int. J. Comput. Math.* **89** 2543–2561. MR3004665

DIACONIS, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol.* 1 (*West Lafayette, Ind.*, 1986) 163–175. Springer, New York. MR0927099

DICK, J., KUO, F. Y. and SLOAN, I. H. (2013). High-dimensional integration: The quasi-Monte Carlo way. *Acta Numer.* **22** 133–288. MR3038697

DICK, J. and PILLICHSHAMMER, F. (2010). *Digital Nets and Sequences*. Cambridge Univ. Press, Cambridge. Discrepancy theory and quasi-Monte Carlo integration. MR2683394

DUVENAUD, D. (2014). Automatic model construction with Gaussian processes Ph.D. thesis, Univ. Cambridge.

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

EFTANG, J. L. and STAMM, B. (2012). Parameter multi-domain '*hp*' empirical interpolation. *Internat. J. Numer. Methods Engrg.* **90** 412–428. MR2911317

FRIEL, N., HURN, M. and WYSE, J. (2014). Improving power posterior estimation of statistical evidence. *Stat. Comput.* **24** 709–723. MR3229692

FRIEL, N. and PETTITT, A. N. (2008). Marginal likelihood estimation via power posteriors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 589–607. MR2420416

GELMAN, A. and MENG, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13** 163–185. MR1647507

GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 509–579. MR3351446

GUNTER, T., GARNETT, R., OSBORNE, M., HENNIG, P. and ROBERTS, S. (2014). Sampling for inference in probabilistic models with fast Bayesian quadrature. In *Proc. Adv. Neur. In* 2789–2797.

HAJIZADEH, Y., CHRISTIE, M. and DEMYANOV, V. (2011). Ant colony optimization for history matching and uncertainty quantification of reservoir models. *J. Pet. Sci. Eng.* **77** 78–92.

HENNIG, P., OSBORNE, M. A. and GIROLAMI, M. (2015). Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **471** 20150142, 17. MR3378744

HICKERNELL, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67** 299–322. MR1433265

HICKERNELL, F. J., LEMIEUX, C. and OWEN, A. B. (2005). Control variates for quasi-Monte Carlo. *Statist. Sci.* **20** 1–31. MR2182985

HUG, S., SCHWARZFISCHER, M., HASENAUER, J., MARR, C. and THEIS, F. J. (2016). An adaptive scheduling scheme for calculating Bayes factors with thermodynamic integration using Simpson's rule. *Stat. Comput.* **26** 663–677. MR3489863

HUSZAR, F. and DUVENAUD, D. (2012). Optimally-weighted herding is Bayesian quadrature. In *Proc. Uncertainty in Artificial Intelligence* (*UAI*) 377–385.

KADANE, J. B. (1985). Parallel and sequential computation: A statistician's view. *J. Complexity* **1** 256–263. MR0925431

KADANE, J. B. and WASILKOWSKI, G. W. (1985). Average case ϵ-complexity in computer science. A Bayesian view. In *Bayesian Statistics*, 2 (*Valencia*, 1983) 361–374. North-Holland, Amsterdam. MR0862497

KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169

KANAGAWA, M., SRIPERUMBUDUR, B. K. and FUKUMIZU, K. (2016). Convergence guarantees for kernel-based quadrature rules in misspecified settings. In *Proc. Adv. Neur. In.* (*NIPS*).

KANAGAWA, M., SRIPERUMBUDUR, B. K. and FUKUMIZU, K. (2017). Convergence analysis of deterministic kernel-based quadrature rules in misspecified settings. arXiv:1709.00147.

KARVONEN, T. and SÄRKKÄ, S. (2017). Classical quadrature rules via Gaussian processes. *IEEE I Workshop Mach. Learn. Signal Proc.* to appear.

KARVONEN, T. and SÄRKKÄ, S. (2018). Fully symmetric kernel quadrature. *SIAM J. Sci. Comput.* **40** A697–A720. MR3769708

KENNEDY, M. (1998). Bayesian quadrature with non-normal approximating functions. *Stat. Comput.* **8** 365–375.

KRISTOFFERSEN, S. (2013). The empirical interpolation method Master's thesis, Dept. Mathematical Sciences, Norwegian Univ. Science and Technology, Trondheim.

KUO, F. Y., DUNSMUIR, W. T. M., SLOAN, I. H., WAND, M. P. and WOMERSLEY, R. S. (2008). Quasi-Monte Carlo for highly structured generalised response models. *Methodol. Comput. Appl. Probab.* **10** 239–275. MR2399683

LAN, S., BUI-THANH, T., CHRISTIE, M. and GIROLAMI, M. (2016). Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems. *J. Comput. Phys.* **308** 81–101. MR3448239

LARKIN, F. M. (1972). Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mountain J. Math.* **2** 379–421. MR0303193

MARQUES, R., BOUVILLE, C., RIBARDIERE, M., SANTOS, L. P. and BOUATOUCH, K. (2013). A spherical Gaussian framework for Bayesian Monte Carlo rendering of glossy surfaces. *IEEE Trans. Vis. Comput. Graph.* **19** 1619–1632.

MARQUES, R., BOUVILLE, C., SANTOS, L. and BOUATOUCH, K. (2015). Efficient quadrature rules for illumination integrals: From quasi Monte Carlo to Bayesian Monte Carlo. *Synth. Lect. Comput. Graph. Animat.* **7** 1–92.

MINKA, T. (2000). Deriving quadrature rules from Gaussian processes Technical report, Statistics Dept., Carnegie Mellon Univ.

MOHAMED, L., CHRISTIE, M. and DEMYANOV, V. (2010). Comparison of stochastic sampling algorithms for uncertainty quantification. *SPE J.* **15** 31–38.

MOSBACH, S. and TURNER, A. G. (2009). A quantitative probabilistic investigation into the accumulation of rounding errors in numerical ODE solution. *Comput. Math. Appl.* **57** 1157–1167. MR2508546

NOVAK, E. and WOŹNIAKOWSKI, H. (2008). *Tractability of Multivariate Problems. Vol.* 1: *Linear Information. EMS Tracts in Mathematics* **6**. European Mathematical Society (EMS), Zürich. MR2455266

NOVAK, E. and WOŹNIAKOWSKI, H. (2010). *Tractability of Multivariate Problems. Volume II*: *Standard Information for Functionals. EMS Tracts in Mathematics* **12**. European Mathematical Society (EMS), Zürich. MR2676032

O'HAGAN, A. (1991). Bayes-Hermite quadrature. *J. Statist. Plann. Inference* **29** 245–260. MR1144171

O'HAGAN, A. (1992). Some Bayesian numerical analysis. In *Bayesian Statistics*, 4 (*Peñíscola*, 1991) 345–363. Oxford Univ. Press, New York. MR1380285

OATES, C. J., COCKAYNE, J. and AYKROYD, R. G. (2017). Bayesian probabilistic numerical methods for industrial process monitoring. arXiv:1707.06107.

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

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

OATES, C. J., COCKAYNE, J., BRIOL, F.-X. and GIROLAMI, M. (2016). Convergence rates for a class of estimators based on Stein's identity. Bernoulli (to appear), arXiv:1603.03220.

OATES, C. J., NIEDERER, S., LEE, A., BRIOL, F. X. and GIROLAMI, M. (2017). Probabilistic models for integration error in assessment of functional cardiac models. In *Proc. Adv. Neur. In.* (*NIPS*). to appear.

OETTERSHAGEN, J. (2017). Construction of optimal cubature algorithms with applications to econometrics and uncertainty quantification Ph.D. Thesis, Univ. Bonn.

OSBORNE, M. A. (2010). Bayesian Gaussian processes for sequential prediction, optimisation and quadrature Ph.D. thesis, Univ. Oxford.

OSBORNE, M. A., DUVENAUD, D., GARNETT, R., RASMUSSEN, C. E., ROBERTS, S. and GHAHRAMANI, Z. (2012). Active learning of model evidence using Bayesian quadrature. In *Proc. Adv. Neur. In.* (*NIPS*) 46–54.

PARK, H., SCHEIDT, C., FENWICK, D., BOUCHER, A. and CAERS, J. (2013). History matching and uncertainty quantification of facies models with multiple geological interpretations. *Comput. Geosci.* **17** 609–621.

PHARR, M. and HUMPHREYS, G. (2004). *Physically Based Rendering*: *From Theory to Implementation*. Kaufmann, Los Altos.

RASMUSSEN, C. E. and GHAHRAMANI, Z. (2002). Bayesian Monte Carlo. In *Proc. Adv. Neur. Inf.* (*NIPS*) 489–496.

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

ROBERT, C. P. and CASELLA, G. (1999). *Monte Carlo Statistical Methods*. *Springer Texts in Statistics*. Springer, New York. MR1707311

SÄRKKA, S., HARTIKAINEN, J., SVENSSON, L. and SANDBLOM, F. (2016). On the relation between Gaussian process quadratures and sigma-point methods. *J. Adv. Inf. Fusion* **11** 31–46.

SCHÖLKOPF, B. and SMOLA, A. (2002). *Learning with Kernels*: *Support Vector Machines*, *Regularization*, *Optimization and Beyond*. MIT Press, Cambridge.

SICKEL, W. and ULLRICH, T. (2009). Tensor products of Sobolev–Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Theory* **161** 748–786. MR2563079

SINESCU, V., KUO, F. Y. and SLOAN, I. H. (2012). On the choice of weights in a function space for quasi-Monte Carlo methods for a class of generalised response models in statistics. In *Proc. Monte Carlo and Quasi-Monte Carlo Methods*.

SMOLA, A., GRETTON, A., SONG, L. and SCHÖLKOPF, B. (2007). A Hilbert space embedding for distributions. In *Proc.* 18*th I. Conf. Algorithmic Learn. Theory* 13–31.

SOMMARIVA, A. and VIANELLO, M. (2006). Numerical cubature on scattered data by radial basis functions. *Computing* **76** 295–310. MR2210099

SONG, L. (2008). Learning via Hilbert space embedding of distributions Ph.D. thesis, School of Information Technologies, Univ. Sydney.

STEIN, M. L. (1995a). Predicting integrals of random fields using observations on a lattice. *Ann. Statist.* **23** 1975–1990. MR1389861

STEIN, M. L. (1995b). Locally lattice sampling designs for isotropic random fields. *Ann. Statist.* **23** 1991–2012. MR1389862

STEIN, M. L. (1999). *Interpolation of Spatial Data*. *Springer Series in Statistics*. Springer, New York. Some theory for Kriging. MR1697409

STEINWART, I. and CHRISTMANN, A. (2008). *Support Vector Machines*. *Information Science and Statistics*. Springer, New York. MR2450103

SULDIN, A. B. (1959). Wiener measure and its applications to approximation methods. I. Izvestiya Vysshikh Uchebnykh Zavedenii. *Matematika* **6** 145–158.

SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

TRAUB, J. F., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1988). *Information-Based Complexity*. *Computer Science and Scientific Computing*. Academic Press, Boston, MA. MR0958691

WAHBA, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442

WENDLAND, H. (2005). *Scattered Data Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724

XI, X., BRIOL, F.-X. and GIROLAMI, M. (2018). Bayesian quadrature for multiple related integrals. International Conference on Machine Learning, PMLR 80:5369–5378, arXiv:1801.04153.

XU, W. and STEIN, M. L. (2017). Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA J. Uncertain. Quantificat.* **5** 138–175. MR3601677

YANG, Y. and DUNSON, D. B. (2016). Bayesian manifold regression. *Ann. Statist.* **44** 876–905. MR3476620

YU, Y., DEBEVEC, P., MALIK, J. and HAWKINS, T. (1999). Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proc. Ann. Conf. Comput. Graph. Int. Tech.* 215–224.

Statistical Science
2019, Vol. 34, No. 1, 23–28
https://doi.org/10.1214/18-STS685

Comment on “Probabilistic Integration: A Role in Statistical Computation?”

Fred J. Hickernell and R. Jagadeeswaran

Abstract. Probabilistic integration provides a criterion for stopping a simulation when a specified error tolerance is satisfied with high confidence. We comment on some of the modeling assumptions and implementation issues involved in designing an automatic Bayesian cubature.

Key words and phrases: Bayesian, fast algorithms, quasi-Monte Carlo.

REFERENCES

BRIOL, F.-X., OATES, C. J., GRIOLAMI, M., OSBORNE, M. A. and SEJDINOVIC, D. (2019). Probabilistic integration: A role in statistical computation? *Statist. Sci.* **34** 1–22.

CHOI, S. C. T., DING, Y., HICKERNELL, F. J., JIANG, L., JIMÉNEZ RUGAMA, LL. A., LI, D., JAGADEESWARAN, R., TONG, X., ZHANG, K., ZHANG, Y. and ZHOU, X. (2013–2017). GAIL: Guaranteed Automatic Integration Library (Versions 1.0–2.2).

DICK, J., KUO, F. Y. and SLOAN, I. H. (2013). High-dimensional integration: The quasi-Monte Carlo way. *Acta Numer.* **22** 133–288. MR3038697

HICKERNELL, F. J. (1996). Quadrature error bounds with applications to lattice rules. *SIAM J. Numer. Anal.* **33** 1995–2016. MR1411860

HICKERNELL, F. J. (1998). Lattice rules: How well do they measure up? In *Random and Quasi-Random Point Sets*. *Lect. Notes Stat.* **138** 109–166. Springer, New York. MR1662841

JAGADEESWARAN, R. and HICKERNELL, F. J. (2018). Fast automatic Bayesian cubature using lattice sampling. Submitted for publication.

OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F., CLARK, C. W. and DALHUIS, A. B. O. (2018). Digital Library of Mathematical Functions.

SHAH, A., WILSON, A. G. and GHAHRAMANI, Z. (2014). Student-*t* processes as alternatives to Gaussian processes. In *Proceedings of the* 17*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 2014, *Reykjavik*, *Iceland*, *PLMR* (S. Kaski and J. Corander, eds.) **33** 877–885.

SLOAN, I. H. and JOE, S. (1994). *Lattice Methods for Multiple Integration*. *Oxford Science Publications*. Oxford University Press, New York. MR1442955

Fred J. Hickernell is Professor of Applied Mathematics and Vice Provost for Research, Illinois Institute of Technology, RE 208, 10 W. 32nd Street, Chicago, Illinois 60616, USA (e-mail: hickernell@iit.edu; URL: iit.edu/~hickernell). R. Jagadeeswaran is a Ph.D. student in Applied Mathematics, Illinois Institute of Technology, RE 208, 10 W. 32nd Street, Chicago, Illinois 60616, USA (e-mail: jrathin1@hawk.iit.edu).

Statistical Science
2019, Vol. 34, No. 1, 29–33
https://doi.org/10.1214/18-STS676

Comment: Unreasonable Effectiveness of Monte Carlo

Art B. Owen

Abstract. There is a role for statistical computation in numerical integration. However, the competition from incumbent methods looks to be stiffer for this problem than for some of the newer problems being handled by probabilistic numerics. One of the challenges is the unreasonable effectiveness of the central limit theorem. Another is the unreasonable effectiveness of pseudorandom number generators. A third is the common $O(n^3)$ cost of methods based on Gaussian processes. Despite these advantages, the classical methods are weak in places where probabilistic methods could bring an improvement.

Key words and phrases: Probabilistic numerics, quasi-Monte Carlo.

REFERENCES

COCKAYNE, J., OATES, C., SULLIVAN, T. and GIROLAMI, M. (2017). Bayesian probabilistic numerical methods. Technical report. Available at arXiv:1702.03673.

DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. MR0836973

DIACONIS, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics*, *IV*, *Vol*. 1 (*West Lafayette*, *Ind*., 1986) 163–175. Springer, New York. MR0927099

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist*. **14** 1–67. MR0829555

DICK, J. (2011). Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *Ann. Statist*. **39** 1372–1398. MR2850206

DICK, J. and PILLICHSHAMMER, F. (2010). *Digital Nets and Sequences*: *Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge Univ. Press, Cambridge. MR2683394

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

FERRENBERG, A. M. LANDAU, D. P. and WONG, Y. J. (1992). Monte Carlo simulations: Hidden errors from "good" random number generators. *Phys. Rev. Lett*. **69** 3382.

GELMAN, A. and SHIRLEY, K. (2011). Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 163–174. CRC Press, Boca Raton, FL. MR2858448

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist*. **16** 927–985. MR0959185

HENNIG, P., OSBORNE, M. A. and GIROLAMI, M. (2015). Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Statistical Society*, *A* **471** 20150142, 17. MR3378744

JAGADEESWARAN, R. and HICKERNELL, F. J. (2018). Fast automatic Bayesian cubature using lattice sampling. Technical report. Available at arXiv:1809.09803.

L'ECUYER, P. and LEMIEUX, C. (2002). A survey of randomized quasi-Monte Carlo methods. In *Modeling Uncertainty*: *An Examination of Stochastic Theory*, *Methods*, *and Applications* (M. Dror, P. L'Ecuyer and F. Szidarovszki, eds.) 419–474. Kluwer Academic, Boston, MA.

L'ECUYER, P. and SIMARD, R. (2007). TestU01: A C library for empirical testing of random number generators. *ACM Trans*. *Math*. *Software* **33** Art. 22. MR2404400

LAVINE, M. and HODGES, J. (2019). Intuition for an old curiosity and an implication for MCMC. *Amer. Statist*. To appear. DOI: 10.1080/00031305.2018.1518267.

LEWIS, P. A. W., GOODMAN, A. S. and MILLER, J. M. (1969). A pseudo-random number generator for the System/360. *IBM System Journal* **8** 136–146.

NIEDERREITER, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull*. *Amer. Math*. *Soc*. **84** 957–1041. MR0508447

NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **63**. SIAM, Philadelphia, PA. MR1172997

OWEN, A. B. (1992). Empirical likelihood and small samples. In *Computing Science and Statistics* 79–88. Springer, Berlin.

OWEN, A. B. (2018). Effective dimension of some weighted pre-Sobolev spaces with dominating mixed partial derivatives. Technical report. Available at arXiv:1709.06695.

Art B. Owen is Professor, Department of Statistics, Stanford University, Sequoia Hall, Stanford, California 94305, USA (e-mail: owen@stanford.edu).

PENG, L. (2004). Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. *Ann. Statist.* **32** 1192–1214. MR2065202

RITTER, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Springer, Berlin. MR1763973

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. MR1041765

SLOAN, I. H. (2007). Finite-order integration weights can be dangerous. *Comput. Methods Appl. Math.* **7** 239–254. MR2404133

SURJANOVIC, S. and BINGHAM, D. (2014). Virtual library of simulation experiments: Test functions and datasets. Available at http://www.sfu.ca/~ssurjano.

TAN, Z. (2004). On a likelihood approach for Monte Carlo integration. *J. Amer. Statist. Assoc.* **99** 1027–1036. MR2109492

VATS, D., FLEGAL, J. M. and JONES, G. L. (2015). Multivariate output analysis for Markov chain Monte Carlo. Technical report. Available at arXiv:1512.07713.

XU, W. and STEIN, M. L. (2017). Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA J. Uncertain. Quantificat.* **5** 138–175. MR3601677

Statistical Science
2019, Vol. 34, No. 1, 34–37
https://doi.org/10.1214/18-STS677

Comment on "Probabilistic Integration: A Role in Statistical Computation?"

Michael L. Stein and Ying Hung

REFERENCES

BLACKWELL, D. and DUBINS, L. (1962). Merging of opinions with increasing information. *Ann. Math. Stat.* **33** 882–886. MR0149577

CHILÈS, J.-P. and DELFINER, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2850475

FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89** 197–210. MR1888368

GNEITING, T., ŠEVČÍKOVÁ, H., PERCIVAL, D. B., SCHLATHER, M. and JIANG, Y. (2006). Fast and exact simulation of large Gaussian lattice systems in $\mathbb{R}^2$: Exploring the limits. *J. Comput. Graph. Statist.* **15** 483–501. MR2291260

GRAMACY, R. B. and LEE, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *J. Amer. Statist. Assoc.* **103** 1119–1130. MR2528830

GRAMACY, R. B. and LEE, H. K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics* **51** 130–145. MR2668170

JOSEPH, V. R. (2012). Bayesian computation using design of experiments-based interpolation technique. *Technometrics* **54** 209–225. MR2967968

SACKS, J. and YLVISAKER, D. (1970). Statistical designs and integral approximation. In *Proc. Twelfth Biennial Sem. Canad. Math. Congr. on Time Series and Stochastic Processes; Convexity and Combinatorics* (*Vancouver, B.C.*, 1969) 115–136. Canad. Math. Congr., Montreal, QC. MR0277069

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. MR1041765

SAMPSON, P. D. and GUTTORP, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* **87** 108–119.

ZHU, Z. and STEIN, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *J. Agric. Biol. Environ. Sci.* **11** 24–44.

Michael Stein is Professor of Statistics, University of Chicago, Chicago, Illinois, USA (e-mail: stein@galton.uchicago.edu). Ying Hung is Assistant Professor of Statistics, Rutgers University, New Brunswick, New Jersey, USA (e-mail: yhung@stat.rutgers.edu).

Statistical Science
2019, Vol. 34, No. 1, 38–42
https://doi.org/10.1214/18-STS683

Rejoinder: Probabilistic Integration: A Role in Statistical Computation?

François-Xavier Briol, Chris J. Oates, Mark Girolami, Michael A. Osborne and Dino Sejdinovic

Abstract. This article is the rejoinder for the paper "Probabilistic Integration: A Role in Statistical Computation?" (*Statist. Sci.* **34** (2019) 1–22). We would first like to thank the reviewers and many of our colleagues who helped shape this paper, the Editor for selecting our paper for discussion, and of course all of the discussants for their thoughtful, insightful and constructive comments. In this rejoinder, we respond to some of the points raised by the discussants and comment further on the fundamental questions underlying the paper: (i) Should Bayesian ideas be used in numerical analysis? and (ii) If so, what role should such approaches have in statistical computation?

Key words and phrases: Computational statistics, nonparametric statistics, probabilistic numerics, uncertainty quantification.

REFERENCES

BRIOL, F-X., OATES, C. J., GIROLAMI, M., OSBORNE, M. A. and SEJDINOVIC, D. (2019). Probabilistic integration: A role in statistical computation? *Statist. Sci.* **34** 1–22.

COCKAYNE, J., OATES, C. J., SULLIVAN, T. and GIROLAMI, M. (2017). Bayesian probabilistic numerical methods. arXiv:1701.04006.

COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540

COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525

DIACONIS, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics*, *IV*, *Vol.* 1 (*West Lafayette*, *Ind.*, 1986) 163–175. Springer, New York. MR0927099

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–67. MR0829555

DUNLOP, M. M., GIROLAMI, M., STUART, A. M. and TECKENTRUP, A. L. (2018). How deep are deep Gaussian processes? *J. Mach. Learn. Res.* **19** 1–46.

JAGADEESWARAN, R. and HICKERNELL, F. J. (2018). Fast automatic Bayesian cubature using lattice sampling. arXiv:1809.09803.

KANAGAWA, M., SRIPERUMBUDUR, B. and FUKUMIZU, K. (2016). Convergence guarantees for kernel-based quadrature rules in misspecified settings. In *Advances in Neural Information Processing Systems* 3288–3296.

KANAGAWA, M., HENNIG, P., SEJDINOVIC, D. and SRIPERUMBUDUR, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. arXiv:1807.02582.

KARVONEN, T. and SÄRKKÄ, S. (2018). Fully symmetric kernel quadrature. *SIAM J. Sci. Comput.* **40** A697–A720. MR3769708

OATES, C. J., COCKAYNE, J. and AYKROYD, R. G. (2017). Bayesian probabilistic numerical methods for industrial process monitoring. arXiv:1707.06107.

OWHADI, H., SCOVEL, C. and SULLIVAN, T. (2015). On the brittleness of Bayesian inference. *SIAM Rev.* **57** 566–582. MR3419871

RITTER, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Springer, Berlin. MR1763973

SNIEKERS, S. and VAN DER VAART, A. (2015). Adaptive Bayesian

François-Xavier Briol is Research Associate, Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (e-mail: f.briol16@imperial.ac.uk). Chris J. Oates is Chair of Statistics, School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK (e-mail: chris.oates@newcastle.ac.uk). Mark Girolami is Chair of Statistics, Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (e-mail: m.girolami@imperial.ac.uk). Michael A. Osborne is Dyson Associate Professor in Machine Learning, Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK (e-mail: mosb@robots.ox.ac.uk). Dino Sejdinovic is Associate Professor, Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK (e-mail: dino.sejdinovic@stats.ox.ac.uk).

credible sets in regression with a Gaussian process prior. *Electron. J. Stat.* **9** 2475–2527. MR3425364

SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

XI, X., BRIOL, F.-X. and GIROLAMI, M. (2018). Bayesian quadrature for multiple related integrals. In *International Conference on Machine Learning*, *PMLR* 80 5369–5378.

ZHU, Z. and STEIN, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *J. Agric. Biol. Environ. Stat.* **11** 24–44.

Statistical Science
2019, Vol. 34, No. 1, 43–68
https://doi.org/10.1214/18-STS667

Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition

Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone

Abstract. Statisticians have made great progress in creating methods that reduce our reliance on parametric assumptions. However, this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to at best 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, "Is Your SATT Where It's At?", launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from 30 competitors across the two versions of the competition (black-box algorithms and do-it-yourself analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies and settings that affect performance. The most consistent conclusion was that methods that flexibly model the response surface perform better overall than methods that fail to do so. Finally new methods are proposed that combine features of several of the top-performing submitted methods.

Key words and phrases: Causal inference, competition, machine learning, automated algorithms, evaluation.

REFERENCES

ABADIE, A. and IMBENS, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74** 235–267. MR2194325

THE H2O. AI TEAM (2016). h2o: R Interface for H2O. R package version 3.10.0.10.

ATHANASOPOULOS, G. and HYNDMAN, R. J. (2011). The value of feedback in forecasting competitions. *Int. J. Forecast.* **27** 845–849.

ATHEY, S. and IMBENS, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113** 7353–7360. MR3531135

AUSTIN, P. C., GROOTENDORST, P. and ANDERSON, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and un-

Vincent Dorie is Associate Research Scientist, Data Science Institute, Columbia University, 475 Riverside Drive, Room 320L, New York, New York 10115, USA (e-mail: vdorie@gmail.com). Jennifer Hill is Professor of Applied Statistics and Data Science, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: jennifer.hill@nyu.edu). Uri Shalit is Assistant Professor, Faculty of Industrial Engineering and Management, Technion, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: urishalit@technion.ac.il). Marc Scott is Professor of Applied Statistics, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: marc.scott@nyu.edu). Dan Cervone is Director of Quantitative Research, Los Angeles Dodgers, Dodger Stadium, 1000 Vin Scully Ave., Los Angeles, California 90012, USA (e-mail: dcervone@gmail.com).

treated subjects: A Monte Carlo study. *Stat. Med.* **26** 734–753. MR2339171

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189

BARNOW, B. S., CAIN, G. G. and GOLDBERGER, A. S. (1980). Issues in the analysis of selectivity bias. In *Evaluation Studies* (E. Stromsdorfer and G. Farkas, eds.) **5** 42–59. Sage, San Francisco, CA.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

CARPENTER, J. (2011). May the best analyst win. *Science* **331** 698–699.

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172

COHEN, J. (1962). The statistical power of abnormal-social psychological research: A review. *J. Abnorm. Soc. Psychol.* **65** 145–163.

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. and MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96** 187–199. MR2482144

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

DIETTERICH, T. G. (2000). Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems* 1–15. Springer, Berlin.

DING, P. and MIRATRIX, L. (2014). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *J. Causal Inference* **3** 41–57.

DORIE, V., HARADA, M., CARNEGIE, N. B. and HILL, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Stat. Med.* **35** 3453–3470. MR3537215

DORIE, V., HILL, J., SHALIT, U., SCOTT, M. and CERVONE, D. (2019). Supplement to "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." DOI:10.1214/18-STS667SUPP.

ENDERS, C. K. and TOFIGHI, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychol. Methods* **12** 121–138.

FELLER, A. and HOLMES, C. C. (2009). Beyond toplines: Heterogeneous treatment effects in randomized experiments. Unpublished manuscript, Oxford Univ.

GELMAN, A. and HILL, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, New York.

GREENLAND, S. and ROBINS, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *Int. J. Epidemiol.* **15** 413–419.

GREENLAND, S., ROBINS, J. M. and PEARL, J. (1999). Confounding and collapsibility in causal inference. *Statist. Sci.* **14** 29–46.

GUYON, I., ALIFERIS, C. F., COOPER, G. F., ELISSEEFF, A., PELLET, J.-P., SPIRTES, P. and STATNIKOV, A. R. (2008). Design and analysis of the causation and prediction challenge. In *WCCI Causation and Prediction Challenge* 1–33.

HABERMAN, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.* **12** 971–988. MR0751286

HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242

HAHN, P. R., MURRAY, J. S. and CARVALHO, C. (2017). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. Preprint. Available at arXiv:1706.09523.

HARTMAN, E., GRIEVE, R., RAMSAHAI, R. and SEKHON, J. S. (2015). From sample average treatment effect to population average treatment effect on the treated: Combining experimental with observational studies to estimate population treatment effects. *J. Roy. Statist. Soc. Ser. A* **178** 757–778. MR3348358

HILL, J. (2008). Discussion of research using propensity-score matching: Comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin. *Stat. Med.* **27** 2055–2061. MR2439884

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546

HILL, J. L., REITER, J. P. and ZANUTTO, E. L. (2004). A comparison of experimental and observational data analyses. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (X.-L. Meng and A. Gelman, eds.) 49–60. Wiley, Chichester. MR2134801

HILL, J. and SU, Y.-S. (2013). Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *Ann. Appl. Stat.* **7** 1386–1420. MR3127952

HIRANO, K. and IMBENS, G. W. (2001). Estimation of causal effects using propensity score weighting: An application of data on right ear catheterization. *Health Serv. Outcomes Res. Methodol.* **1** 259–278.

HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826

IMAI, K., and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. Roy. Statist. Soc. Ser. B* **76** 243–263.

IMBENS, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

KERN, H. L., STUART, E. A., HILL, J. L. and GREEN, D. P. (2016). Assessing methods for generalizing experimental impact estimates to target samples. *J. Res. Educ. Eff.* **9** 103–127.

KURTH, T., WALKER, A. M., GLYNN, R. J., CHAN, K. A., GAZIANO, J. M., BERGER, K. and ROBINS, J. M. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of non-uniform effect. *Am. J. Epidemiol.* **163** 262–270.

LALONDE, R. and MAYNARD, R. (1987). How precise are evaluations of employment and training programs: Evidence from a field experiment. *Eval. Rev.* **11** 428–451.

LECHNER, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric Evaluation of Labour Market Policies* (M. Lechner and F. Pfeiffer, eds.). *ZEW Economic Studies* **13** 43–58. Physica-Verlag, Heidelberg.

LEDELL, E. (2016). h2oEnsemble: H2O Ensemble Learning. R package version 0.1.8.

LEE, B. K., LESSLER, J. and STUART, E. A. (2010). Improving propensity score weighting using machine learning. *Stat. Med.* **29** 337–346. MR2750549

LITTLE, R. J. (1988). Missing-data adjustments in large surveys. *J. Bus. Econom. Statist.* **6** 287–296.

MIDDLETON, J., SCOTT, M., DIAKOW, R. and HILL, J. (2016). Bias amplification and bias unmasking. *Polit. Anal.* **24** 307–323.

NISWANDER, K. R. and GORDON, M. (1972). *The Collaborative Perinatal Study of the National Institute of Neurological Diseases and Stroke*: *The Women and Their Pregnancies*. W.B. Saunders, Philadelphia, PA.

PAULHAMUS, B., EBAUGH, A., BOYLLS, C., BOS, N., HIDER, S. and GIGUERE, S. (2012). Crowdsourced cyber defense: Lessons from a large-scale, game-based approach to threat identification on a live network. In *International Conference on Social Computing*, *Behavioral–Cultural Modeling*, *and Prediction* 35–42. Springer, Berlin.

PEARL, J. (2009a). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

PEARL, J. (2009b). Causal inference in statistics: An overview. *Stat*. *Surv.* **3** 96–146. MR2545291

PEARL, J. (2010). On a class of bias-amplifying variables that endanger effect estimates. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence* 425–432. Accessed 02/02/2016.

POLLEY, E., LEDELL, E., KENNEDY, C. and VAN DER LAAN, M. (2016). SuperLearner: Super Learner Prediction. R package version 2.0-21.

RANARD, B. L., HA, Y. P., MEISEL, Z. F., ASCH, D. A., HILL, S. S., BECKER, L. B., SEYMOUR, A. K. and MERCHANT, R. M. (2014). Crowdsourcing—Harnessing the masses to advance health and medicine, a systematic review. *J. Gen*. *Intern*. *Med.* **29** 187–203.

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA. MR2514435

ROBINS, J. M. (1999). Association, causation, and marginal structural models. *Synthese* **121** 151–179. MR1766776

ROBINS, J. M. and ROTNITZKY, A. (2001). Comment on 'Inference for semiparametric models: Some questions and an answer,' by P. J. Bickel and J. Kwon. *Statist*. *Sinica* **11** 920–936.

ROKACH, L. (2009). Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. *Comput*. *Statist*. *Data Anal*. **53** 4046–4072. MR2744304

ROSENBAUM, P. R. (1987). Model-based direct adjustment. *J. Amer. Statist*. *Assoc*. **82** 387–394.

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. Springer, New York. MR1899138

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

ROSENBAUM, P. R. and RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist*. *Assoc*. **79** 516–524.

ROSENTHAL, R. (1979). The file drawer problem and tolerance for null results. *Psychol*. *Bull*. **86** 638–641.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann*. *Statist*. **6** 34–58. MR0472152

RUBIN, D. B. (2006). *Matched Sampling for Causal Effects*. Cambridge Univ. Press, Cambridge. MR2307965

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

SEKHON, J. S. (2007). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *J. Stat*. *Softw.*.

SHADISH, W. R., CLARK, M. H. and STEINER, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J. Amer. Statist*. *Assoc*. **103** 1334–1343. MR2655714

STEINER, P. and KIM, Y. (2016). The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases. *J. Causal Inference* **4**.

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist*. *Sci*. **25** 1–21. MR2741812

TADDY, M., GARDNER, M., CHEN, L. and DRAPER, D. (2016). A nonparametric Bayesian analysis of heterogenous treatment effects in digital experimentation. *J. Bus*. *Econom*. *Statist*. **34** 661–672. MR3548002

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

VANSCHOREN, J., VAN RIJN, J. N., BISCHL, B. and TORGO, L. (2014). OpenML: Networked science in machine learning. *ACM SIGKDD Explor. Newsl*. **15** 49–60.

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int*. *J. Biostat*. **2** Art. 11, 40. MR2306500

WAGER, S. and ATHEY, S. (2015). Estimation and inference of heterogeneous treatment effects using random forests. Preprint. Available at arXiv:1510.04342.

WESTREICH, D., LESSLER, J. and FUNK, M. J. (2010). Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J. Clin*. *Epidemiol*. **63** 826–833.

WIND, D. K. and WINTHER, O. (2014). Model selection in data analysis competitions. In *MetaSel@ ECAI* 55–60.

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat*. *Soc*. *Ser. B*. *Stat*. *Methodol*. **67** 301–320. MR2137327

Statistical Science
2019, Vol. 34, No. 1, 69–71
https://doi.org/10.1214/18-STS684

Comment: Spherical Cows in a Vacuum: Data Analysis Competitions for Causal Inference

Miguel A. Hernán

Abstract. A recent data analysis competition compared the performance of several methods for causal inference from observational data. However, sound causal inference requires not only adequate data analysis techniques but also subject-matter expertise about the causal structure of the problem under study. Therefore, until a methodology is developed to combine data analysis and subject-matter knowledge, causal inference competitions may only provide advice to practitioners under ideal conditions.

Key words and phrases: Causal inference, data analysis competitions.

REFERENCES

GREENLAND, S., PEARL, J. and ROBINS, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology* **10** 37–48.

HERNÁN, M. A., HERNÁNDEZ-DÍAZ, S. and ROBINS, J. M. (2004). A structural approach to selection bias. *Epidemiology* **15** 615–625.

HERNÁN, M. A., HSU, J. and HEALY, B. (2019). Data science is science's second chance to get causal inference right. A classification of data science tasks. *Chance* **32** 42–49.

DORIE et al. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statist. Sci.* **34** 43–68.

NISWANDER, K. R. and GORDON, M. (1972). *The Collaborative Perinatal Study of the National Institute of Neurological Diseases and Stroke*: *The Women and Their Pregnancies*. W.B. Saunders, Philadelphia, PA.

PEARL, J. (2011). Understanding bias amplification. *Am. J. Epidemiol.* **174** 1223–1227.

ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. (Errata appeared in *Comput. Math. Appl.* **14**(1987), 917–921).

SPHERICAL COW. In Wikipedia. Retrieved November 6, 2018, from. https://en.wikipedia.org/wiki/Spherical_cow.

Miguel A. Hernán is Professor, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, and Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts 02115, USA (e-mail: miguel_hernan@post.harvard.edu).

Statistical Science
2019, Vol. 34, No. 1, 72–76
https://doi.org/10.1214/18-STS680

Comment: Will Competition-Winning Methods for Causal Inference Also Succeed in Practice?

Qingyuan Zhao, Luke J. Keele and Dylan S. Small

Abstract. First, we would like to congratulate the authors for successfully hosting the causal inference data competition (referred to as Competition henceforth) and contributing a unique and thought-provoking article to the literature. The authors have provided a comprehensive and timely platform to evaluate the ever-growing number of methods used for covariate adjustment in observational studies. In our comment, we don't generally question the results of the competition, but we do wish to emphasize several other key elements about the role statistics plays in causal inference and observational studies.

Key words and phrases: Observational studies, machine learning, study design.

REFERENCES

ANGRIST, J. D. and KRUEGER, A. B. (1999). Empirical strategies in labor economics. In *Handbook of Labor Economics* (O. Ashenfelter and D. Card, eds.) **3A** 1277–1366. Elsevier, Amsterdam.

BOX, G. E. (1979). Some problems of statistics and everyday life. *J. Amer. Statist. Assoc.* **74** 1–4.

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152

COOK, T. D., CAMPBELL, D. T. and SHADISH, W. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, Boston, MA.

COOK, T. D., SHADISH, W. R. and WONG, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *J. Policy Anal. Manage.* **27** 724–750.

HILL, A. B. (1965). The environment and disease: Association or causation? *J. R. Soc. Med.* **58** 295–300.

IMBENS, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *Am. Econ. Rev. Pap. Proc.* **93** 126–132.

KEELE, L. and SMALL, D. (2018). Comparing covariate prioritization via matching to machine learning methods for causal inference using five empirical applications. Preprint. Available at arXiv:1805.03743.

LIPSITCH, M., TCHETGEN TCHETGEN, E. and COHEN, T. (2010). Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* **21** 383–388.

PIMENTEL, S. D., KELZ, R. R., SILBER, J. H. and ROSENBAUM, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *J. Amer. Statist. Assoc.* **110** 515–527. MR3367244

ROSENBAUM, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74** 13–26. MR0885915

ROSENBAUM, P. R. (2001). Replicating effects and biases. *Amer. Statist.* **55** 223–227. MR1963397

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138

ROSENBAUM, P. R. (2005). Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *Amer. Statist.* **59** 147–152. MR2133562

ROSENBAUM, P. R. (2006). Differential effects and generic biases in observational studies. *Biometrika* **93** 573–586. MR2261443

Qingyuan Zhao is Postdoctoral fellow, Department of Statistics, The Wharton School, University of Pennsylvania, 400 Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104, USA (e-mail: qyzhao@wharton.upenn.edu). Luke J. Keele is Associate Professor of Applied Statistics, Hospital of the University of Pennsylvania, Silverstein 4, 3400 Spruce Street Philadelphia, Pennsylvania 19104, USA (e-mail: luke.keele@gmail.com). Dylan S. Small is Professor of Statistics, Department of Statistics, Wharton School, University of Pennsylvania, 400 Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu).

RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2** 808–804. MR2516795

SHADISH, W. R., CLARK, M. H. and STEINER, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J. Amer. Statist. Assoc.* **103** 1334–1343. MR2655714

ZHAO, Q. (2019). Covariate balancing propensity score by tailored loss functions. *Ann. Statist.* **47** 965–993. MR3909957

ZUBIZARRETA, J. R. (2012). Using mixed integer programming for matching in an observational study of kidney failure after surgery. *J. Amer. Statist. Assoc.* **107** 1360–1371. MR3036400

Statistical Science
2019, Vol. 34, No. 1, 77–81
https://doi.org/10.1214/18-STS690

Comment: Strengthening Empirical Evaluation of Causal Inference Methods

David Jensen

Abstract. This is a contribution to the discussion of the paper by Dorie et al. (*Statist. Sci.* **34** (2019) 43–68), which reports the lessons learned from 2016 Atlantic Causal Inference Conference Competition. My comments strongly support the authors' focus on empirical evaluation, using examples and experience from machine learning research, particularly focusing on the problem of algorithmic complexity. I argue that even broader and deeper empirical evaluation should be undertaken by the researchers who study causal inference. Finally, I highlight a few key conclusions that suggest where future research should focus.

Key words and phrases: Causal inference, empirical evaluation, machine learning, algorithmic complexity, constructed observational studies, alignment.

REFERENCES

BRADLEY, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit*. **30** 1145–1159.

COHEN, P. R. and HOWE, A. E. (1988). How evaluation guides AI research: The message still counts more than the medium. *AI Mag*. **9** 35.

DARPA (2017). Ground Truth (GT). Broad agency announcement. Defense Sciences Office. Defense Advanced Research Projects Agency, U.S. Dept. Defense. HR001117S0031.

DHEERU, D. and KARRA TANISKIDOU, E. (2017). UCI machine learning repository.

DOMINGOS, P. and PAZZANI, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn*. **29** 103–130.

DORIE, V. and HILL, J. and SHALIT, U. and SCOTT, M. and CERVONE, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statist. Sci*. **34** 43–68.

GARANT, D. and JENSEN, D. (2016). Evaluating causal models by comparing interventional distributions. ArXiv Preprint arXiv:1608.04698.

GELMAN, A. and LOKEN, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Dept. Statistics, Columbia Univ., New York, NY.

GUYON, I., JANZING, D. and SCHÖLKOPF, B. (2010). Causality: Objectives and assessment. In *Causality*: *Objectives and Assessment* 1–42.

GUYON, I., STATNIKOV, A. and BATU, B. (2019). *Cause-Effect Pairs in Machine Learning*. *Springer Series on Challenges in Machine Learning*. Springer. To appear.

GUYON, I., ALIFERIS, C., COOPER, G., ELISSEEFF, A., PELLET, J.-P., SPIRTES, P. and STATNIKOV, A. (2008). Design and analysis of the causation and prediction challenge. In *Causation and Prediction Challenge* 1–33.

HAHN, P. R., DORIE, V. and MURRAY, J. S. (2018). Atlantic Causal Inference Conference (ACIC) data analysis challenge 2017.

HILL, J. L., REITER, J. P. and ZANUTTO, E. L. (2004). A comparison of experimental and observational data analyses. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. *Wiley Ser. Probab. Stat*. 49–60. Wiley, Chichester. MR2134801

LALONDE, R. and MAYNARD, R. (1987). How precise are evaluations of employment and training programs: Evidence from a field experiment. *Evaluation Review* **11** 428–451.

LEWIS, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *European Conference on Machine Learning* 4–15. Springer.

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

PROVOST, F. J., FAWCETT, T., KOHAVI, R. et al. (1998). The case against accuracy estimation for comparing induction algo-

David Jensen is Professor of Computer Science and Director of the Computational Social Science Institute, University of Massachusetts Amherst, Amherst, Massachusetts 01003, USA (e-mail: jensen@cs.umass.edu; URL: https://people.cs.umass.edu/~jensen/).

rithms. In *Proceedings of the International Conference on Machine Learning* **98** 445–453.

SCHAFFTER, T., MARBACH, D. and FLOREANO, D. (2011). GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27** 2263–2270.

SHADISH, W. R., CLARK, M. H. and STEINER, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J. Amer. Statist. Assoc.* **103** 1334–1343. MR2655714

SHIMONI, Y., YANOVER, C., KARAVANI, E. and GOLDSCHMNIDT, Y. (2018). Benchmarking framework for performance-evaluation of causal inference analysis. ArXiv Preprint arXiv:1802.05046.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR1815675

VAN'T VEER, A. E. and GINER-SOROLLA, R. (2016). Preregistration in social psychology—A discussion and suggested template. *J. Exp. Soc. Psychol.* **67** 2–12.

Statistical Science
2019, Vol. 34, No. 1, 82–85
https://doi.org/10.1214/18-STS689

Comment on "Automated Versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition"

Susan Gruber and Mark J. van der Laan

Abstract. Dorie and co-authors (DHSSC) are to be congratulated for initiating the ACIC Data Challenge. Their project engaged the community and accelerated research by providing a level playing field for comparing the performance of a priori specified algorithms. DHSSC identified themes concerning characteristics of the DGP, properties of the estimators, and inference. We discuss these themes in the context of targeted learning.

Key words and phrases: Targeted learning, causal inference, TMLE.

REFERENCES

BENKESER, D. and VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. *Proc. Int. Conf. Data Sci. Adv. Anal.* **2016** 689–696.

GRUBER, S. and VAN DER LAAN, M. J. (2010). An application of collaborative targeted maximum likelihood estimation in causal inference and genomics. *Int. J. Biostat.* **6** Art. 18, 31. MR2653847

JU, C., SCHWAB, J. and VAN DER LAAN, M. J. (2017). On adaptive propensity score truncation in causal inference. Preprint. Available at arXiv:1707.05861.

JU, C., GRUBER, S., LENDLE, S. D., CHAMBAZ, A., FRANKLIN, J. M., WYSS, R., SCHNEEWEISS, S. and VAN DER LAAN, M. J. (2017). Scalable collaborative targeted learning for high-dimensional data. *Stat. Methods Med. Res.*

POLLEY, E. C. and VAN DER LAAN, M. J. (2010). Super learner in prediction. Technical Report 200, Univ. California Berkeley Division of Biostatistics, Working Paper Series.

VAN DER LAAN, M. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. *Int. J. Biostat.* **13** 20150097, 35. MR3724476

VAN DER LAAN, M. J. and GRUBER, S. (2010). Collaborative double robust targeted maximum likelihood estimation. *Int. J. Biostat.* **6** Art. 17, 70. MR2653848

VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. 25, 23. MR2349918

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

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500

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

Susan Gruber is Principal, Putnam Data Sciences, LLC, 85 Putnam Avenue, Cambridge, Massachusetts 02139, USA (e-mail: sgruber@putnamds.com). Mark J. van der Laan is Professor in Biostatistics & Statistics, University of California at Berkeley, Berkeley Way West, 2121 Berkeley Way, #5302, Berkeley, California 94720-7360, USA (e-mail: laan@berkeley.edu).

Statistical Science
2019, Vol. 34, No. 1, 86–89
https://doi.org/10.1214/18-STS679

Comment: Causal Inference Competitions: Where Should We Aim?

Ehud Karavani, Tal El-Hay, Yishai Shimoni and Chen Yanover

Abstract. Data competitions proved to be highly beneficial to the field of machine learning, and thus expected to provide similar advantages in the field of causal inference. As participants in the 2016 and 2017 Atlantic Causal Inference Conference (ACIC) data competitions and co-organizers of the 2018 competition, we discuss the strengths of simulation-based competitions and suggest potential extensions to address their limitations. These suggested augmentations aim at making the data generating processes more realistic and gradually increase in complexity, allowing thorough investigations of algorithms' performance. We further outline a community-wide competition framework to evaluate an end-to-end causal inference pipeline, beginning with a causal question and a database, and ending with causal estimates.

Key words and phrases: Causal inference, competition, data challenge, machine learning, automated algorithms, evaluation.

REFERENCES

BENNETT, J., LANNING, S. et al. (2007). The Netflix prize. In *Proceedings of KDD Cup and Workshop* **2007** 35. New York, NY.

DANAEI, G., RODRÍGUEZ, L. A. G., CANTERO, O. F., LOGAN, R. W. and HERNÁN, M. A. (2018). Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J. Clin. Epidemiol.* **96** 12–22.

DORIE, V., HILL, J., SHALIT, U., SCOTT, M. and CERVONE, D. (2017). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. ArXiv Preprint. Available at arXiv:707.02641.

HAHN, P. R., DORIE, V. and MURRAY, J. S. (2018). Atlantic Causal Inference Conference (ACIC) Data Analysis Challenge 2017.

HERNÁN, M. A. and ROBINS, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183** 758–764.

HILL, J. L., REITER, J. P. and ZANUTTO, E. L. (2004). A comparison of experimental and observational data analyses. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. *Wiley Ser. Probab. Stat.* 49–60. Wiley, Chichester. MR2134801

HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. MR0867618

JONES, D., MOLITOR, D. and REIF, J. (2018). What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. Technical report, National Bureau of Economic Research, Cambridge, MA.

LENSINK, M. F., VELANKAR, S. and WODAK, S. J. (2017). Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* **85** 359–377.

MALINSKY, D. and DANKS, D. (2018). Causal discovery algorithms: A practical guide **13** e12470.

MOULT, J., FIDELIS, K., KRYSHTAFOVYCH, A., SCHWEDE, T. and TRAMONTANO, A. (2018). Critical assessment of methods of protein structure prediction (CASP)—Round XII **86 Suppl. 1** 7–15.

RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATHY, A., KHOSLA, A. et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115** 211–252.

SHADISH, W. R., CLARK, M. H. and STEINER, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J. Amer. Statist. Assoc.* **103** 1334–1343. MR2655714

SHIMONI, Y., YANOVER, C., KARAVANI, E. and GOLDSCHMNIDT, Y. (2018). Benchmarking framework for performance-evaluation of causal inference analysis. ArXiv Preprint. Available at arXiv:1802.05046.

SILBERZAHN, R., UHLMANN, E., MARTIN, D., ANSELMI, P., AUST, F., AWTREY, E., BAHNIK, Š., BAI, F., BANNARD, C.

Ehud Karavani is researcher, Healthcare Informatics Department, IBM Research—Haifa, Israel (e-mail: ehudk@ibm.com). Tal El-Hay is researcher, Healthcare Informatics Department, IBM Research—Haifa, Israel (e-mail: talelh@il.ibm.com). Yishai Shimoni is researcher, Healthcare Informatics Department, IBM Research—Haifa, Israel (e-mail: yishais@il.ibm.com). Chen Yanover is researcher, Healthcare Informatics Department, IBM Research—Haifa, Israel (e-mail: cheny@il.ibm.com).

et al. (2017). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. PsyArXiv.

Statistical Science
2019, Vol. 34, No. 1, 90–93
https://doi.org/10.1214/18-STS682

Comment: Contributions of Model Features to BART Causal Inference Performance Using ACIC 2016 Competition Data

Nicole Bohme Carnegie

Abstract. With a thorough exposition of the methods and results of the 2016 Atlantic Causal Inference Competition, Dorie et al. have set a new standard for reproducibility and comparability of evaluations of causal inference methods. In particular, the open-source **R** package `aciccomp2016`, which permits reproduction of all datasets used in the competition, will be an invaluable resource for evaluation of future methodological developments.

Building upon results from Dorie et al., we examine whether a set of potential modifications to Bayesian Additive Regression Trees (BART)—multiple chains in model fitting, using the propensity score as a covariate, targeted maximum likelihood estimation (TMLE), and computing symmetric confidence intervals—have a stronger impact on bias, RMSE, and confidence interval coverage in combination than they do alone. We find that bias in the estimate of SATT is minimal, regardless of the BART formulation. For purposes of CI coverage, however, all proposed modifications are beneficial—alone and in combination—but use of TMLE is least beneficial for coverage and results in considerably wider confidence intervals.

Key words and phrases: Bayesian additive regression trees, TMLE, propensity score.

REFERENCES

[1] CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172

[2] DORIE, V. (2017). aciccomp2016: Atlantic Causal Inference Conference Competition 2016 Simulation. R package version 0.1-0.

[3] DORIE, V., HILL, J., SHALIT, U., SCOTT, M. and CERVONE, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statist. Sci.* **34** 43–68.

[4] HAHN, P. R., MURRAY, J. S. and CARVALHO, C. M. (2017). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. Preprint. Available at arXiv:1706.09523 [stat.ME].

[5] R CORE TEAM (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[6] VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500

Nicole Bohme Carnegie is Assistant Professor of Statistics, Department of Mathematical Sciences, Montana State University, PO Box 172400, Bozeman, Montana 59717, USA (e-mail: nicole.carnegie@montana.edu).

Statistical Science
2019, Vol. 34, No. 1, 94–99
https://doi.org/10.1214/18-STS688

Rejoinder: Response to Discussions and a Look Ahead

Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone

Abstract. Response to discussion of Dorie (2017), in which the authors of that piece express their gratitude to the discussants, rebut some specific criticisms, and argue that the limitations of the 2016 Atlantic Causal Inference Competition represent an exciting opportunity for future competitions in a similar mold.

Key words and phrases: Causal inference, competition, machine learning, automated algorithms, evaluation.

REFERENCES

DORIE, V. (2017). aciccomp2016: Atlantic Causal Inference Conference Competition 2016 Simulation. R package version 0.1-0.

LAZER, D., KENNEDY, R., KING, G. and VESPIGNANI, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science* **343** 1203–1205.

Vincent Dorie is an Associate Research Scientist, Columbia University, 475 Riverside Drive, New York, New York 10027, USA (e-mail: vdorie@gmail.com). Jennifer Hill is a Professor of Applied Statistics and Data Science, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: jennifer.hill@nyu.edu). Uri Shalit is an Assistant Professor, Technion, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: urishalit@technion.ac.il). Marc Scott is a Professor of Applied Statistics, New York University, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: marc.scott@nyu.edu). Dan Cervone is Director of Quantitative Research, Los Angeles Dodgers, Dodger Stadium, 1000 Vin Scully Avenue, Los Angeles, California 90012, USA (e-mail: dcervone@gmail.com).

Statistical Science
2019, Vol. 34, No. 1, 100–128
https://doi.org/10.1214/18-STS662

Gaussian Integrals and Rice Series in Crossing Distributions—to Compute the Distribution of Maxima and Other Features of Gaussian Processes

Georg Lindgren

Abstract. We describe and compare how methods based on the classical Rice's formula for the expected number, and higher moments, of level crossings by a Gaussian process stand up to contemporary numerical methods to accurately deal with crossing related characteristics of the sample paths.

We illustrate the relative merits in accuracy and computing time of the Rice moment methods and the exact numerical method, developed since the late 1990s, on three groups of distribution problems, the maximum over a finite interval and the waiting time to first crossing, the length of excursions over a level, and the joint period/amplitude of oscillations.

We also treat the notoriously difficult problem of dependence between successive zero crossing distances. The exact solution has been known since at least 2000, but it has remained largely unnoticed outside the ocean science community.

Extensive simulation studies illustrate the accuracy of the numerical methods. As a historical introduction an attempt is made to illustrate the relation between Rice's original formulation and arguments and the exact numerical methods.

Key words and phrases: Computational statistics, distribution of maximum, Durbin's formula, excursion length distribution, first passage, independent interval assumption, level crossings, multivariate normal probabilities, period/amplitude distribution, Rice's formula, RIND program, statistical computation, stochastic process, successive crossing distance distribution, truncated normal moments, Wafo toolbox.

REFERENCES

ABRAHAMS, J. (1982). The zero crossing problem for some nonstationary Gaussian procceses. *IEEE Trans. Inform. Theory* **IT-28** 677–678.

ABRAHAMS, J. (1986). A survey of recent progress on level-crossing problems for random processes. In *Communications and Networks* (I. F. Blake and H. V. Poor, eds.) 6–25. Springer, New York.

ARELLANO-VALLE, R. B. and GENTON, M. G. (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statist. Probab. Lett.* **78** 27–35. MR2381270

AZAÏS, J.-M., DE CASTRO, Y. and MOURAREAU, S. (2018). Power of the spacing test for least-angle regression. *Bernoulli* **24** 465–492. MR3706766

AZAÏS, J.-M. and GENZ, A. (2013). Computation of the distribution of the maximum of stationary Gaussian processes. *Methodol. Comput. Appl. Probab.* **15** 969–985. MR3117635

AZAÏS, J.-M., LEÓN, J. R. and WSCHEBOR, M. (2011). Rice formulae and Gaussian waves. *Bernoulli* **17** 170–193. MR2797987

AZAÏS, J.-M. and WSCHEBOR, M. (2002). The distribution of the maximum of a Gaussian process: Rice method revisited. In *In and Out of Equilibrium* (*Mambucaba*, 2000). *Progress in Probability* **51** 321–348. Birkhäuser, Boston, MA. MR1901961

Georg Lindgren is Professor emeritus, Mathematical Statistics, Lund University, Box 118, SE-221 00, Lund, Sweden (e-mail: georg@maths.lth.se).

AZAÏS, J.-M. and WSCHEBOR, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Wiley, Hoboken, NJ. MR2478201

BLAKE, I. F. and LINDSEY, W. C. (1973). Level-crossing problems for random processes. *IEEE Trans. Inform. Theory* **IT-19** 295–315. MR0370729

BRAININA, I. S. (2013). *Applications of Random Process Excursion Analysis*. Elsevier, London. MR3286421

BRODTKORB, P. A. (2006). Evaluating nearly singular multinormal expectations with application to wave distributions. *Methodol. Comput. Appl. Probab.* **8** 65–91. MR2253077

BRODTKORB, P. A., JOHANNESSON, P., LINDGREN, G., RYCHLIK, I., RYDÉN, J. and SJÖ, E. (2000). WAFO—a Matlab toolbox for analysis of random waves and loads. In *Proc.* 10*th Int. Offshore Polar Engng Conf.* **3** 343–350.

CARTWRIGHT, D. E. and LONGUET-HIGGINS, M. S. (1956). The statistical distribution of the maxima of a random function. *Proc. R. Soc. Lond. Ser. A* **237** 212–232. MR0082227

CAVANIÉ, A., ARHAN, M. and EZRATY, R. (1976). A statistical relationship between individual heights and periods of storm waves. In *Proceedings, BOSS'*76 354–360, Trondheim.

CIERCO-AYROLLES, C., CROQUETTE, A. and DELMAS, C. (2003). Computing the distribution of the maximum of Gaussian random processes. *Methodol. Comput. Appl. Probab.* **5** 427–438. MR2018904

CRAMÉR, H. (1940). On the theory of stationary random processes. *Ann. of Math.* (2) **41** 215–230. MR0000920

CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.

CRANDALL, S. H. (1958). *Random Vibration*. Wiley and M.I.T Technology Press.

DE MARÉ, J. (1974). When are the successive zero-crossing intervals of a Gaussian process independent? Technical report, Lund Univ.

DURBIN, J. (1985). The first-passage density of a continuous Gaussian process to a general boundary. *J. Appl. Probab.* **22** 99–122. MR0776891

EFRON, B. (1997). The length heuristic for simultaneous hypothesis tests. *Biometrika* **84** 143–157. MR1450198

ESTRADE, A., IRIBARREN, I. and KRATZ, M. (2012). Chord-length distribution functions and Rice formulae. Application to random media. *Extremes* **15** 333–352. MR2969063

GADRICH, T. and ADLER, R. J. (1993). Slepian models for nonstationary Gaussian processes. *J. Appl. Probab.* **30** 98–111. MR1206355

GENZ, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.* **1** 141–149.

GENZ, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. In *Proc.* 25*th Symp. Interface Computer Science and Statistics* **25** 400–405.

GENZ, A. (2012). Matlab software for multivariate normal probability computation. http://www.math.wsu.edu/faculty/genz/software/software.html.

GENZ, A. and BRETZ, F. (2009). *Computation of Multivariate Normal and t Probabilities*. *Lecture Notes in Statistics* **195**. Springer, Dordrecht. MR2840595

GENZ, A. and KWONG, K.-S. (2000). Numerical evaluation of singular multivariate normal distributions. *J. Stat. Comput. Simul.* **68** 1–21. MR1820599

GENZ, A., BRETZ, F., MIWA, T. and LEISCH, F. (2017). Multivariate normal and t distribution. http://mvtnorm.R-forge.R-project.org.

GRIGORIU, M. (1989). Reliability of Daniels systems subject to quasistatic and dynamic nonstationary Gaussian load processes. *Probab. Eng. Mech.* **4** 128–134.

KAC, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* **49** 314–320. MR0007812

KAC, M. and SLEPIAN, D. (1959). Large excursions of Gaussian processes. *Ann. Math. Stat.* **30** 1215–1228. MR0115222

KAN, R. and ROBOTTI, C. (2017). On moments of folded and truncated multivariate normal distributions. *J. Comput. Graph. Statist.* **26** 930–934. MR3765356

KHINTCHINE, A. (1934). Korrelationstheorie der stationären stochastischen Prozesze. *Math. Ann.* **109** 604–615. MR1512911

KRATZ, M. F. (2006). Level crossings and other level functionals of stationary Gaussian processes. *Probab. Surv.* **3** 230–288. MR2264709

LANING, J. H. JR. and BATTIN, R. H. (1956). *Random Processes in Automatic Control*. McGraw-Hill, New York. MR0079362

LAZAROV, B. and DITLEVSEN, O. (2005). Slepian simulation of distributions of plastic displacements of earthquake excited shear frames with a large number of stories. *Probab. Eng. Mech.* **20** 251–262.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. MR0691492

LINDGREN, G. (1972). Wave-length and amplitude in Gaussian noise. *Adv. in Appl. Probab.* **4** 81–108. MR0314104

LINDGREN, G. (2013). *Stationary Stochastic Processes—Theory and Applications*. CRC Press, Boca Raton, FL. MR3024784

LINDGREN, G. (2019). Supplement to "Gaussian integrals and Rice series in crossing distributions—to compute the distribution of maxima and other features of Gaussian processes." DOI:10.1214/18-STS662SUPP.

LINDGREN, G. and BROBERG, K. B. (2004). Cycle range distributions for Gaussian processes—exact and approximative results. *Extremes* **7** 69–89. MR2201192

LINDGREN, G. and RYCHLIK, I. (1982). Wave characteristic distributions for Gaussian waves—wave-length, amplitude and steepness. *Ocean Eng.* **9** 411–432.

LINDGREN, G. and RYCHLIK, I. (1991). Slepian models and regression approximations in crossing and extreme value theory. *Int. Stat. Rev.* **59** 195–225.

LONGUET-HIGGINS, M. S. (1957). The statistical analysis of a random, moving surface. *Philos. Trans. R. Soc. Lond. Ser. A* **249** 321–387. MR0087257

LONGUET-HIGGINS, M. S. (1962). The distribution of intervals between zeros of a stationary random function. *Philos. Trans. R. Soc. Lond. Ser. A* **254** 557–599. MR0158431

LONGUET-HIGGINS, M. S. (1963). Bounding approximations to the distribution of intervals between zeros of a stationary Gaussian process. In *Proc. Sympos. Time Series Analysis* (*Brown Univ.*, 1962) 63–88. Wiley, New York. MR0148124

LONGUET-HIGGINS, M. S. (1975). On the joint distribution of wave periods and amplitudes of sea waves. *J. Geophys. Res.* **80** 2688–2694.

LONGUET-HIGGINS, M. S. (1983). On the joint distribution of wave periods and amplitudes in a random wave field. *Proc. R. Soc. Lond. Ser. A* **389** 241–258.

LYON, R. H. (1956). Response of strings to random noise fields. *J. Acoust. Soc. Am.* **28** 391–398. MR0083342

MCFADDEN, J. A. (1958). The axis-crossing intervals of random functions. II. *IRE Trans.* **4** 14–24. MR0098437

MERCADIER, C. (2006a). Numerical bounds for the distributions of the maxima of some one- and two-parameter Gaussian processes. *Adv. in Appl. Probab.* **38** 149–170. MR2213969

MERCADIER, C. (2006b). MAGP (Maximum Analysis of Gaussian Processes). http://math.univ-lyon1.fr/mercadier/MAGP/index.html.

MIDDLETON, D. (1988). S. O. Rice and the theory of random noise: Some personal recollections. *IEEE Trans. Inform. Theory* **34** 1367–1373. MR0993430

MILES, J. W. (1954). On the transonic drag of accelerated bodies. *J. Aeronaut. Sci.* **21** 644–645. MR0063883

MILLMAN, S., ed. (1984). *A History of Engineering & Science in the Bell System: Communication Sciences*, 1925–1980, 1–91. AT&T Bell Laboratories, Indianapolis, IN.

MIMAKI, T. (1973). Zero-crossing intervals of Gaussian processes. *J. Appl. Phys.* **44** 477–485.

MIMAKI, T., MYOKEN, H. and KAWABATA, T. (1985). Level-crossing problem of a Gaussian process having Gaussian power spectrum density. *Jpn. J. Appl. Phys.* **24** L278–L280.

MIMAKI, T., TANABE, M. and WOLF, D. (1981). The multipeak property of the distribution densities of the level-crossing intervals of a Gaussian random process. *IEEE Trans. Inform. Theory* **IT-27**.

MUNAKATA, T. (1998). Some unsolved problems on the level crossing of random processes. In *Unsolved Problems of Noise in Physics, Biology, Electronic Technology and Information Technology* (C. R. Coering, L. B. Kiss and M. F. Shlesinger, eds.) 213–222. World Scientific, Singapore.

O'HAGAN, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* **91** 1290–1300.

PODGÓRSKI, K., RYCHLIK, I. and MACHADO, U. E. B. (2000). Exact distributions for apparent waves in irregular seas. *Ocean Eng.* **27** 979–1016.

PODGÓRSKI, K., RYCHLIK, I. and WALLIN, J. (2015). Slepian noise approach for Gaussian and Laplace moving average processes. *Extremes* **18** 665–695. MR3418772

PRÜHER, J. and SÄRKKÄ, S. (2016). On the use of gradient information in Gaussian process quadratures. In *Proc.* 2016 *IEEE Int. Workshop on Machine Learning for Signal Processing*.

RAINAL, A. J. (1988). Origin of Rice's formula. *IEEE Trans. Inform. Theory* **34** 1383–1387. MR0993433

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA. MR2514435

RICE, S. O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23** 282–332. MR0010932

RICE, S. O. (1945). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **24** 46–156. MR0011918

ROSENBAUM, S. (1961). Moments of a truncated bivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **23** 405–408. MR0132630

RYCHLIK, I. (1987a). A note on Durbin's formula for the first-passage density. *Statist. Probab. Lett.* **5** 425–428. MR0903802

RYCHLIK, I. (1987b). Joint distribution of successive zero crossing distances for stationary Gaussian processes. *J. Appl. Probab.* **24** 378–385. MR0889802

RYCHLIK, I. (1987c). Regression approximations of wavelength and amplitude distributions. *Adv. in Appl. Probab.* **19** 396–430. MR0889943

RYCHLIK, I. (1990). New bounds for the first passage, wave-length and amplitude densities. *Stochastic Process. Appl.* **34** 313–339. MR1047649

RYCHLIK, I. (1992a). Confidence bands for linear regressions. *Comm. Statist. Simulation Comput.* **21** 333–352. MR1173064

RYCHLIK, I. (1992b). The two-barrier problem for continuously differentiable processes. *Adv. in Appl. Probab.* **24** 71–94. MR1146520

SIRE, C. (2008). Crossing intervals of non-Markovian Gaussian processes. *Phys. Rev. E* (3) **78** 011121. MR2496170

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183

SLEPIAN, D. (1963). On the zeros of Gaussian noise. In *Proc. Sympos. Time Series Analysis* (*Brown Univ.*, 1962). Wiley, New York. MR0148128

SMITH, J. M., HOPCRAFT, K. I. and JAKEMAN, E. (2008). Fluctuations in the zeros of differentiable Gaussian processes. *Phys. Rev. E* (3) **77** 031112. MR2495410

STDENIS, M. and PIERSON, W. J. (1953). *On the Motion of Ships in Confused Seas. SNAME Trans.* **61**.

WAFO-GROUP (2017a). WAFO—A Matlab Toolbox for Analysis of Random Waves and Loads—Tutorial for WAFO 2017. http://www.maths.lth.se/matstat/wafo.

WAFO-GROUP (2017b). Wafo-project. https://github.com/wafo-project.

WAX, N. (1954). *Selected Papers on Noise and Stochastic Processes*. Dover, New York.

WIENER, N. (1930). Generalized harmonic analysis. *Acta Math.* **55** 117–258. MR1555316

WILSON, L. R. M. and HOPCRAFT, K. I. (2017). Periodicity in the autocorrelation function as a mechanism for regularly occurring zero crossings or extreme values of a Gaussian process. *Phys. Rev. E* **96** 062129.

WONG, E. (1966). Some results concerning the zero-crossings of Gaussian noise. *SIAM J. Appl. Math.* **14** 1246–1254. MR0207059

YOUSSEF, N., MUNAKATA, T. and TAKEDA, M. (1996). Rice probability functions for level-crossing intervals of speckle intensity fields. *Opt. Commun.* **123** 55–62.

Statistical Science
2019, Vol. 34, No. 1, 129–155
https://doi.org/10.1214/18-STS668

Generalized Multiple Importance Sampling

Víctor Elvira, Luca Martino, David Luengo and Mónica F. Bugallo

Abstract. Importance sampling (IS) methods are broadly used to approximate posterior distributions or their moments. In the standard IS approach, samples are drawn from a single proposal distribution and weighted adequately. However, since the performance in IS depends on the mismatch between the targeted and the proposal distributions, several proposal densities are often employed for the generation of samples. Under this multiple importance sampling (MIS) scenario, extensive literature has addressed the selection and adaptation of the proposal distributions, interpreting the sampling and weighting steps in different ways. In this paper, we establish a novel general framework with sampling and weighting procedures when more than one proposal is available. The new framework encompasses most relevant MIS schemes in the literature, and novel valid schemes appear naturally. All the MIS schemes are compared and ranked in terms of the variance of the associated estimators. Finally, we provide illustrative examples revealing that, even with a good choice of the proposal densities, a careful interpretation of the sampling and weighting procedures can make a significant difference in the performance of the method.

Key words and phrases: Monte Carlo methods, multiple importance sampling, Bayesian inference.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York. MR1225604

BUGALLO, M. F., ELVIRA, V., MARTINO, L., LUENGO, D., MÍGUEZ, J. and DJURIC, P. M. (2017). Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Process. Mag.* **34** 60–79.

CAPPÉ, O., GUILLIN, A., MARIN, J. M. and ROBERT, C. P. (2004). Population Monte Carlo. *J. Comput. Graph. Statist.* **13** 907–929. MR2109057

CAPPÉ, O., DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2008). Adaptive importance sampling in general mixture classes. *Stat. Comput.* **18** 447–459. MR2461888

CORNUET, J.-M., MARIN, J.-M., MIRA, A. and ROBERT, C. P. (2012). Adaptive multiple importance sampling. *Scand. J. Stat.* **39** 798–812. MR3000850

DOUC, R. and CAPPÉ, O. (2005). Comparison of resampling schemes for particle filtering. In *ISPA* 2005. *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis* 64–69. IEEE, New York.

DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2007a). Convergence of adaptive mixtures of importance sampling schemes. *Ann. Statist.* **35** 420–448. MR2332281

DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2007b). Minimum variance importance sampling via population Monte Carlo. *ESAIM Probab. Stat.* **11** 427–447. MR2339302

ELVIRA, V., MARTINO, L., LUENGO, D. and BUGALLO, M. F. (2015). Efficient multiple importance sampling estimators. *IEEE Signal Process. Lett.* **22** 1757–1761.

ELVIRA, V., MARTINO, L., LUENGO, D. and CORANDER, J. (2015b). A gradient adaptive population importance sampler. In *IEEE International Conf. on Acoustics, Speech and Signal Processing* (*ICASSP*) 4075–4079.

ELVIRA, V., MARTINO, L., LUENGO, D. and BUGALLO, M. F.

Víctor Elvira is Associate Professor, IMT Lille Douai & CRIStAL Laboratory (UMR 9189), Rue Guglielmo Marconi, Villeneuve d'Ascq 59653, France (e-mail: victor.elvira@imt-lille-douai.fr). Luca Martino is Visiting Professor, Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain (e-mail: lmartino@ing.uc3m.es). David Luengo is Associate Professor, Department of Signal Theory and Communications, Universidad Politécnica de Madrid, C/ Nikola Tesla, 28031 Madrid, Spain (e-mail: david.luengo@upm.es). Mónica F. Bugallo is Professor, Department of Electrical & Computer Engineering, Stony Brook University, Stony Brook, New York 11794-2350, USA (e-mail: monica.bugallo@stonybrook.edu).

(2016). Heretical multiple importance sampling. *IEEE Signal Process*. *Lett*. **23** 1474–1478.

ELVIRA, V., MARTINO, L., LUENGO, D. and BUGALLO, M. F. (2017). Improving Population Monte Carlo: Alternative weighting and resampling schemes. *Signal Process*. **131** 77–91.

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57** 1317–1339. MR1035115

GORDON, N., SALMOND, D. and SMITH, A. F. M. (1993). Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proc*. *F*, *Commun*. *Radar Signal Process*. **140** 107–113.

GWANYAMA, P. W. (2004). The HM-GM-AM-QM inequalities. *College Math*. *J*. **35** 47–50.

HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Comput*. *Statist*. **14** 375–396.

HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, Cambridge. MR0046395

HE, H. Y. and OWEN, A. B. (2014). Optimal mixture weights in multiple importance sampling. Preprint. Available at arXiv:1411.3954.

HESTERBERG, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37** 185–194.

KAHN, H. and MARSHALL, A. W. (1953). Methods of reducing sample size in Monte Carlo computations. *J*. *Oper*. *Res*. *Soc*. *Am*. **1** 263–278.

KONG, A. (1992). A note on importance sampling using standardized weights. Technical Report 348, Dept. Statistics, Univ. Chicago, Chicago, IL.

KONG, A., LIU, J. S. and WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J*. *Amer*. *Statist*. *Assoc*. **9** 278–288.

KONG, A., MCCULLAGH, P., MENG, X.-L., NICOLAE, D. and TAN, Z. (2003). A theory of statistical models for Monte Carlo integration. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **65** 585–618. MR1998624

LIANG, F. (2002). Dynamically weighted importance sampling in Monte Carlo computation. *J*. *Amer*. *Statist*. *Assoc*. **97** 807–821. MR1941411

LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer, New York. MR2401592

MARTINO, L., ELVIRA, V., LUENGO, D. and CORANDER, J. (2015a). An adaptive population importance sampler: Learning from uncertainty. *IEEE Trans*. *Signal Process*. **63** 4422–4437. MR3368395

MARTINO, L., ELVIRA, V., LUENGO, D. and CORANDER, J. (2017). Layered adaptive importance sampling. *Stat*. *Comput*. **27** 599–623. MR3613588

NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **63**. SIAM, Philadelphia, PA. MR1172997

OWEN, A. (2013). Monte Carlo Theory, Methods and Examples. Available at http://statweb.stanford.edu/~owen/mc/.

OWEN, A. and ZHOU, Y. (2000). Safe and effective importance sampling. *J*. *Amer*. *Statist*. *Assoc*. **95** 135–143. MR1803146

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. MR2080278

TAN, Z. (2004). On a likelihood approach for Monte Carlo integration. *J*. *Amer*. *Statist*. *Assoc*. **99** 1027–1036. MR2109492

VEACH, E. and GUIBAS, L. (1995). Optimally combining sampling techniques for Monte Carlo rendering. In *SIGGRAPH* 1995 *Proceedings* 419–428.

Statistical Science
2019, Vol. 34, No. 1, 156–168
https://doi.org/10.1214/18-STS663

A Conversation with Piet Groeneboom

Geurt Jongbloed

Abstract. Petrus (Piet) Groeneboom was born in Scheveningen in 1941 and grew up in Voorburg. Both villages are located near The Hague in The Netherlands; Scheveningen actually being part of The Hague. He attended the gymnasium of the Huygens lyceum. In 1959, he entered the University of Amsterdam, where he studied psychology. After his "candidate" exam (comparable to BSc) in 1963, he worked at the psychological laboratory of the University of Amsterdam until 1966. In 1965, he took up mathematics as a part-time study. After having obtained his master's degree in 1971, he had a position at the psychological laboratory again until 1973, when he was appointed to the Mathematical Center in Amsterdam. There, he wrote between 1975 and 1979 his Ph.D. thesis with Kobus Oosterhoff as advisor, graduating in 1979. After a period of two years as visiting professor at the University of Washington (UW) in Seattle, Piet moved back to the Mathematical Center until he was appointed full professor of statistics at the University of Amsterdam in 1984. Four years later, he moved to Delft University of Technology where he became professor of statistics and stayed until his retirement in 2006. Between 2000 and 2006 he also held a part-time professorship at the Vrije Universiteit in Amsterdam. From 1999 till 2013 he was Affiliate Professor at the statistics department of UW, Seattle. Apart from being visiting professor at the UW in Seattle, he was also visiting professor at Stanford University, Université Paris 6 and ETH Zürich.

Piet is well known for his work on shape constrained statistical inference. He worked on asymptotic theory for these problems, created algorithms to compute nonparametric estimates in such models and applied these models to real data. He also worked on interacting particle systems, extreme value analysis and efficiency theory for testing procedures. Piet (co-)authored four books and 64 papers and served as promotor of 13 students. He is the recipient of the 1985 Rollo Davidson prize, a fellow of the IMS and elected member of the ISI. In 2015, he delivered the Wald lecture at the Joint Statistical Meeting in Montreal.

Piet and his wife Marijke live in Naarden. He has two sons, Thomas and Tim, and (since June 12, 2018) one grandson, Tarik. This conversation was held at Piet's house in Naarden, on February 28 and April 24, 2018.

Key words and phrases: University of Amsterdam, Mathematical Center (CWI), Delft University of Technology, violin playing.

Geurt Jongbloed is Professor of Statistics and Head of the Mathematics Department of Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands (e-mail: G.Jongbloed@tudelft.nl).

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. *The Theory and Application of Isotonic Regression*. Wiley, London. MR0326887

BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann*. *Statist*. **11** 432–452. MR0696057

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BUYTENDIJK, F. J. J. (1951). *De Vrouw*. Het Spectrum, Utrecht.

CATOR, E. A., JONGBLOED, G., KRAAIKAMP, C., LOPUHAÄ, H. P. and WELLNER, J. A., eds. (2007) *Asymptotics*: *Particles*, *Processes and Inverse Problems*. *Festschrift for Piet Groeneboom*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55**. IMS, Beachwood, OH. MR2459927

DIEUDONNÉ, J. (1960). *Foundations of Modern Analysis*. *Pure and Applied Mathematics*, *Vol*. *X*. Academic Press, New York. MR0120319

DIEUDONNÉ, J. (1968). *Calcul Infinitésimal*. Hermann, Paris. MR0226971

FELLER, W. (1957). *An Introduction to Probability Theory and Its Applications*. *Vol*. *I*. 2nd ed. Wiley, New York; Chapman & Hall, London. MR0088081

GILL, R. D., GROENEBOOM, P. and DE JONG, P. (2018). Elementary statistics on trial (the case of Lucia de Berk). *Chance*. To appear.

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. *A roughness penalty approach*. *Monographs on Statistics and Applied Probability* **58**. Chapman & Hall, London. MR1270012

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann*. *Probab*. **11** 1016–1027. MR0714964

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol*. *II* (*Berkeley*, *Calif*., 1983). *Wadsworth Statist*./*Probab*. *Ser*. 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann*. *Statist*. **29** 1620–1652. MR1891741

GROENEBOOM, P., LEPAGE, Y. and RUYMGAART, F. H. (1976). Rank tests for independence with best strong exact Bahadur slope. *Z*. *Wahrsch*. *Verw*. *Gebiete* **36** 119–127. MR0418330

GROENEBOOM, P. and PYKE, R. (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann*. *Probab*. **11** 328–345. MR0690131

GROENEBOOM, P. and SHORACK, G. R. (1981). Large deviations of goodness of fit statistics and linear combinations of order statistics. *Ann*. *Probab*. **9** 971–987. MR0632970

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237

KELLEY, J. L. (1975). *General Topology*. Springer, New York. MR0370454

KUIPER, N. H. (1960). Tests concerning random points on a circle. *Proc*. *K*. *Ned*. *Akad*. *Wet*., *Ser*. *A*, *Indag*. *Math*. **63** 38–47.

LANDAU, E. (1951). *Foundations of Analysis*. Chelsea, New York.

MEERTENS, L. (1968). *Quartet No*. 1 *in C Major for* 2 *Violins*, *Viola and Violoncello*. Mathematisch Centrum, Amsterdam. Mathematical Centre Report MR **96**.

MEIHUIZEN, J. (2010). *Smalle Marges*. Boom, Amsterdam.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. Springer, New York. MR0762984

REICHENBACH, H. (1951). *The Rise of Scientific Philosophy*. Univ. California Press, Berkeley, CA.

SCHUH, F. (1951). *Playing with Numbers*. (in Dutch: *Spelen met Getallen*). Thieme.

TARSKI, A. (1941). *Introduction to Logic*. Oxford Univ. Press, London.

VAN DER VAART, A. (1991). On differentiable functionals. *Ann*. *Statist*. **19** 178–204. MR1091845

WICKHAM, H. (2015). *R Packages*. O'Reilly Media, Sebastopol, CA.

Statistical Science
2019, Vol. 34, No. 1, 169–175
https://doi.org/10.1214/18-STS678

A Conversation with Dick Dudley

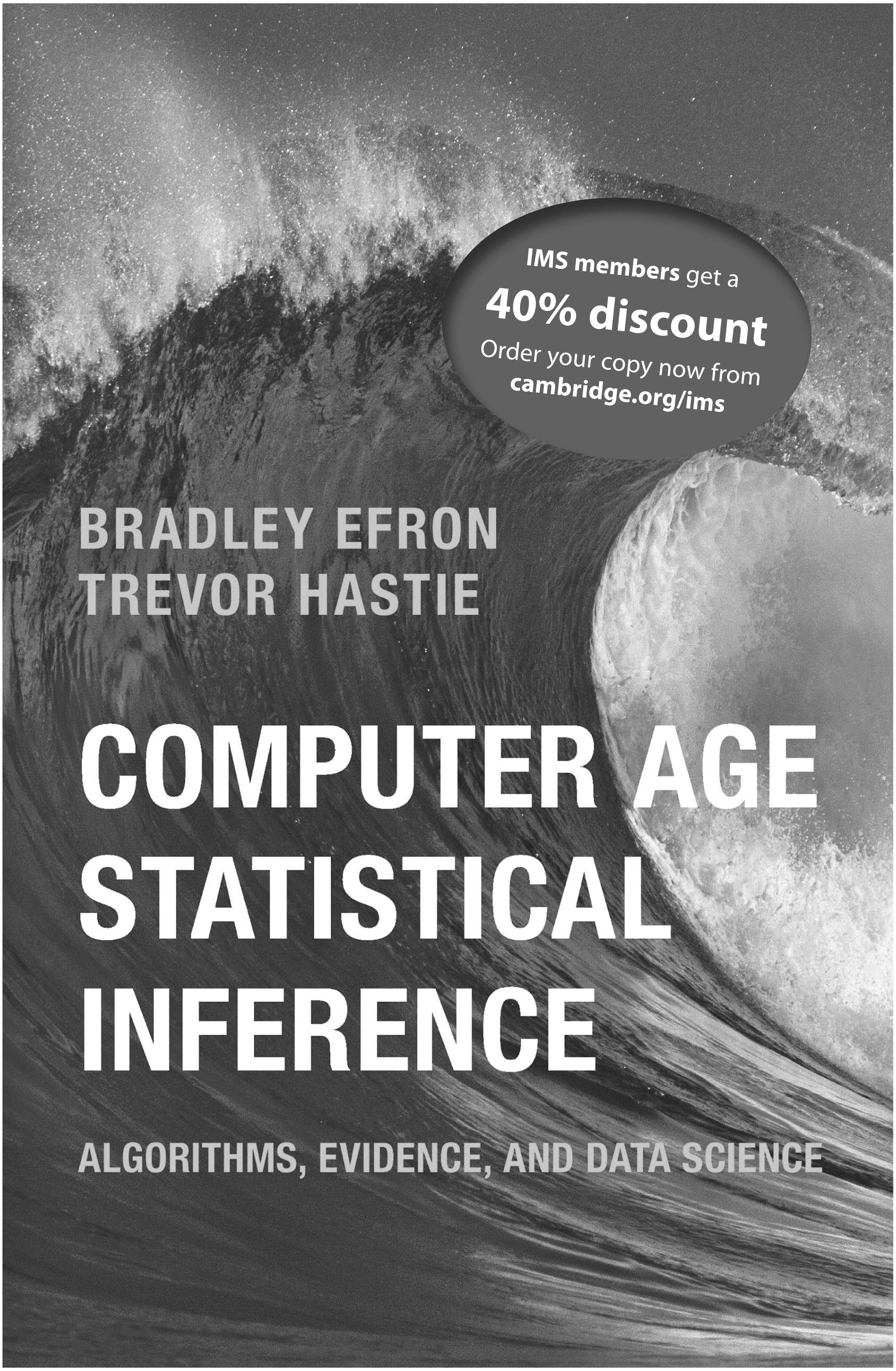
Vladimir Koltchinskii, Richard Nickl and Philippe Rigollet

Abstract. Richard Mansfield Dudley (Dick Dudley) was born in 1938. He received the A.B. from Harvard in 1952 and the Ph.D. from Princeton in 1962 (under the supervision of Gilbert Hunt and Edward Nelson). Following an appointment at UC Berkeley as an assistant professor, he joined the Department of Mathematics at MIT in 1967. Dick Dudley has made fundamental contributions to the theory of Gaussian processes and Probability in Banach Spaces. Among his major achievements is the development of a general framework for empirical processes theory, in particular, for uniform central limit theorems. These results have had and continue having tremendous impact in contemporary statistics and in mathematical foundations of machine learning. A more extensive biographical sketch is contained in the preface to the *Selected works of R. M. Dudley* (editors: E. Giné, V. Koltchinskii and R. Norvaisa) published in 2010.

This conversation took place (mostly, via email) in the fall of 2017.

Key words and phrases: Biography, probability, statistics.

Vladimir Koltchinskii is Professor of Mathematics, School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. Richard Nickl is Professor of Mathematical Statistics, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, United Kingdom (e-mail: r.nickl@statslab.cam.ac.uk). Philippe Rigollet is Associate Professor of Mathematics, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA.


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