STATISTICAL SCIENCE

Volume 39, Number 1 February 2024

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 39, Number 1, February 2024. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

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

Statistical Science Volume 39, Number 1 (1–220) February 2024

Volume 39 Number 1 February 2024

Statistical Science
2024, Vol. 39, No. 1, 1–2
https://doi.org/10.1214/23-STS920

Editorial: Bayesian Computations in the 21st Century

Christian P. Robert and Dennis Prangle

REFERENCES

[1] ANASTASIOU, A., BARP, A., BRIOL, F.-X., EBNER, B., GAUNT, R. E., GHADERINEZHAD, F., GORHAM, J., GRETTON, A., LEY, C. et al. (2023). Stein's method meets computational statistics: A review of some recent developments. *Statist. Sci.* **38** 120–139. MR4534646 https://doi.org/10.1214/22-sts863

[2] CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

[3] FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. MR3843382 https://doi.org/10.1214/18-STS648

Christian Robert is a Professor in Statistics, Université Paris–Dauphine, 75775 Paris Cedex 16, France (e-mail: xian@ceremade.dauphine.fr). Dennis Prangle is a Senior Lecturer in Statistics, University of Bristol, School of Mathematics, Woodland Rd, Bristol BS8 1UG, UK (e-mail: dennis.prangle@bristol.ac.uk).

Statistical Science
2024, Vol. 39, No. 1, 3–19
https://doi.org/10.1214/22-STS876

Computing Bayes: From Then 'Til Now

Gael M. Martin, David T. Frazier and Christian P. Robert

Abstract. This paper takes the reader on a journey through the history of Bayesian computation, from the 18th century to the present day. Beginning with the one-dimensional integral first confronted by Bayes in 1763, we highlight the key contributions of: Laplace, Metropolis (and, importantly, his coauthors), Hammersley and Handscomb, and Hastings, all of which set the foundations for the computational revolution in the late 20th century—led, primarily, by Markov chain Monte Carlo (MCMC) algorithms. A very short outline of 21st century computational methods—including pseudo-marginal MCMC, Hamiltonian Monte Carlo, sequential Monte Carlo and the various "approximate" methods—completes the paper.

Key words and phrases: History of Bayesian computation, Laplace approximation, Metropolis–Hastings algorithm, importance sampling, Markov chain Monte Carlo, pseudo-marginal methods, Hamiltonian Monte Carlo, sequential Monte Carlo, approximate Bayesian methods.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

ANDRAL, C. (2022). An attempt to trace the birth of importance sampling. Available at https://hal.archives-ouvertes.fr/hal-03696061.

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

ANDRIEU, C., DOUCET, A. and ROBERT, C. P. (2004). Computational advances for and from Bayesian analysis. *Statist. Sci.* **19** 118–127. MR2082151 https://doi.org/10.1214/088342304000000071

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

ARDIA, D., BAŞTÜRK, N., HOOGERHEIDE, L. and VAN DIJK, H. K. (2012). A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Comput. Statist. Data Anal.* **56** 3398–3414. MR2943902 https://doi.org/10.1016/j.csda.2010.09.001

BAKER, J., FEARNHEAD, P., FOX, E. B. and NEMETH, C. (2019). Control variates for stochastic gradient MCMC. *Stat. Comput.* **29** 599–615. MR3969063 https://doi.org/10.1007/s11222-018-9826-2

BANTERLE, M., GRAZIAN, C., LEE, A. and ROBERT, C. P. (2019). Accelerating Metropolis–Hastings algorithms by delayed acceptance. *Found. Data Sci.* **1** 103–128.

BARDENET, R., DOUCET, A. and HOLMES, C. (2017). On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.* **18** Paper No. 47. MR3670492

BARNARD, G. A. and BAYES, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika* **45** 293–315.

BAYES, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Lond., B* **53** 370–418.

BEAUMONT, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160.

BEAUMONT, M. A. (2019). Approximate Bayesian computation. *Annu. Rev. Stat. Appl.* **6** 379–403. MR3939526 https://doi.org/10.1146/annurev-statistics-030718-105212

BEAUMONT, M. A., CORNUET, J.-M., MARIN, J.-M. and ROBERT, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* **96** 983–990. MR2767283 https://doi.org/10.1093/biomet/asp052

BÉDARD, M., DOUC, R. and MOULINES, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Process. Appl.* **122** 758–786. MR2891436 https://doi.org/10.1016/j.spa.2011.11.004

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611 https://doi.org/10.1007/978-1-4757-4286-2

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B* **55** 25–37. MR1210422

BETANCOURT, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. Available at arXiv:1701.02434v2.

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

Gael M. Martin is Professor, Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia (e-mail: gael.martin@monash.edu). David T. Frazier is Associate Professor, Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia (e-mail: david.frazier@monash.edu). Christian P. Robert is Professor, Ceremade, Université Paris-Dauphine, Paris, France, and Department of Statistics, Warwick University, Coventry, UK (e-mail: xian@ceremade.dauphine.fr).

BIERKENS, J., BOUCHARD-CÔTÉ, A., DOUCET, A., DUNCAN, A. B., FEARNHEAD, P., LIENART, T., ROBERTS, G. and VOLLMER, S. J. (2018). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Probab. Lett.* **136** 148–154. MR3806858 https://doi.org/10.1016/j.spl.2018.02.021

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BORNN, L., SHEPHARD, N. and SOLGI, R. (2019). Moment conditions and Bayesian non-parametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 5–43. MR3904778 https://doi.org/10.1111/rssb.12294

BROOKS, S., GELMAN, A., JONES, G. and MENG, X. L. (2011). *Handbook of Markov Chain Monte Carlo*. Taylor & Francis.

CALVET, L. E. and CZELLAR, V. (2015). Accurate methods for approximate Bayesian computation filtering. *J. Financ. Econom.* **13** 798–838.

CAPPÉ, O., GUILLIN, A., MARIN, J. M. and ROBERT, C. P. (2004). Population Monte Carlo. *J. Comput. Graph. Statist.* **13** 907–929. MR2109057 https://doi.org/10.1198/106186004X12803

CARPENTER, B., GELMAN, A., HOFFMAN, M., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.

CARTER, C. K. and KOHN, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81** 541–553. MR1311096 https://doi.org/10.1093/biomet/81.3.541

CASELLA, G. and GEORGE, E. I. (1992). Explaining the Gibbs sampler. *Amer. Statist.* **46** 167–174. MR1183069 https://doi.org/10.2307/2685208

CASELLA, G., LAVINE, M. and ROBERT, C. P. (2001). Explaining the perfect sampler. *Amer. Statist.* **55** 299–305. MR1939363 https://doi.org/10.1198/000313001753272240

CASELLA, G. and ROBERT, C. P. (1996). Rao–Blackwellisation of sampling schemes. *Biometrika* **83** 81–94. MR1399157 https://doi.org/10.1093/biomet/83.1.81

CERUZZI, P. (2003). *A History of Modern Computing*, 2nd ed. MIT Press, Cambridge, MA.

CHEN, S., DICK, J. and OWEN, A. B. (2011). Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. *Ann. Statist.* **39** 673–701. MR2816335 https://doi.org/10.1214/10-AOS831

CHIB, S. (2011). Introduction to simulation and MCMC methods. In *The Oxford Handbook of Bayesian Econometrics* 183–217. Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780199559084.013.0006

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis–Hastings algorithm. *Amer. Statist.* **49** 327–335.

CHIB, S. and GREENBERG, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* **12** 409–431. MR1409939 https://doi.org/10.1017/S0266466600006794

CHIB, S. and RAMAMURTHY, S. (2010). Tailored randomized block MCMC methods with application to DSGE models. *J. Econometrics* **155** 19–38. MR2592847 https://doi.org/10.1016/j.jeconom.2009.08.003

CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

CHRISTEN, J. A. and FOX, C. (2005). Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Statist.* **14** 795–810. MR2211367 https://doi.org/10.1198/106186005X76983

CRAIU, R. V. and MENG, X.-L. (2005). Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Ann. Statist.* **33** 661–697. MR2163156 https://doi.org/10.1214/009053604000001075

CRAIU, R. V. and MENG, X.-L. (2011). Perfection within reach: Exact MCMC sampling. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 199–226. CRC Press, Boca Raton, FL. MR2858450

DAMIEN, P., WAKEFIELD, J. and WALKER, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 331–344. MR1680334 https://doi.org/10.1111/1467-9868.00179

DE BRUIJN, N. G. (1961). *Asymptotic Methods in Analysis*, 2nd ed. *Bibliotheca Mathematica*, *Vol. IV*. North-Holland, Amsterdam. MR0177247

DEAN, T. A., SINGH, S. S., JASRA, A. and PETERS, G. W. (2014). Parameter estimation for hidden Markov models with intractable likelihoods. *Scand. J. Stat.* **41** 970–987. MR3277033 https://doi.org/10.1111/sjos.12077

DELIGIANNIDIS, G., DOUCET, A. and PITT, M. K. (2018). The correlated pseudomarginal method. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 839–870. MR3874301 https://doi.org/10.1111/rssb.12280

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

DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. MR0836973 https://doi.org/10.1007/978-1-4613-8643-8

DIEBOLT, J. and ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B* **56** 363–375. MR1281940

DONGARRA, J. and SULLIVAN, F. (2000). Guest editors' introduction: The top 10 algorithms. *Comput. Sci. Eng.* **2** 22–23.

DOUC, R. and ROBERT, C. P. (2011). A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *Ann. Statist.* **39** 261–277. MR2797846 https://doi.org/10.1214/10-AOS838

DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

DUNSON, D. B. and JOHNDROW, J. E. (2020). The Hastings algorithm at fifty. *Biometrika* **107** 1–23. MR4064137 https://doi.org/10.1093/biomet/asz066

FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. MR3843382 https://doi.org/10.1214/18-STS648

FIENBERG, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Anal.* **1** 1–40. MR2227361 https://doi.org/10.1214/06-BA101

FRAZIER, D. T. and DROVANDI, C. (2021). Robust approximate Bayesian inference with synthetic likelihood. *J. Comput. Graph. Statist.* **30** 958–976. MR4356598 https://doi.org/10.1080/10618600.2021.1875839

FRAZIER, D. T., ROBERT, C. P. and ROUSSEAU, J. (2020). Model misspecification in approximate Bayesian computation: Consequences and diagnostics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 421–444. MR4084170 https://doi.org/10.1111/rssb.12356

FRAZIER, D. T., LOAIZA-MAYA, R., MARTIN, G. M. and KOO, B. (2021). Loss-based variational Bayes prediction. ArXiv preprint. Available at arXiv:2104.14054.

FRIGESSI, A., GÅSEMYR, J. and RUE, H. (2000). Antithetic coupling of two Gibbs sampler chains. *Ann. Statist.* **28** 1128–1149. MR1810922 https://doi.org/10.1214/aos/1015956710

FRÜHWIRTH-SCHNATTER, S. (1994). Data augmentation and dynamic linear models. *J. Time Series Anal.* **15** 183–202. MR1263889 https://doi.org/10.1111/j.1467-9892.1994.tb00184.x

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

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). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 163–174. CRC Press, Boca Raton, FL. MR2858448 https://doi.org/10.1201/b10905

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741. https://doi.org/10.1109/tpami.1984.4767596

GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 509–579. MR3351446 https://doi.org/10.1111/rssb.12104

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57** 1317–1339. MR1035115 https://doi.org/10.2307/1913710

GEWEKE, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2166278 https://doi.org/10.1002/0471744735

GEWEKE, J., KOOP, G. and VAN DIJK, H. (2011). *The Oxford Handbook of Bayesian Econometrics*. OUP.

GEYER, C. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics*: *Proceedings of the* 32*rd Symposium on the Interface* (E. M. Keramigas, ed.) 156–163. Interface Foundation, Fairfax, VA.

GEYER, C. J. (2011a). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 3–48. CRC Press, Boca Raton, FL. MR2858443

GEYER, C. J. (2011b). Importance sampling, simulated tempering, and umbrella sampling. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 295–311. CRC Press, Boca Raton, FL. MR2858453

GHOSAL, S., GHOSH, J. K. and SAMANTA, T. (1995). On convergence of posterior distributions. *Ann. Statist.* **23** 2145–2152. MR1389869 https://doi.org/10.1214/aos/1034713651

GLYNN, P. W. (2016). Exact Simulation vs. Exact Estimation. *Proceedings of the 2016 Winter Simulation Conference WSC16*, 193–205.

GLYNN, P. W. and RHEE, C.-H. (2014). Exact estimation for Markov chain equilibrium expectations. *J. Appl. Probab.* **51A** 377–389. MR3317370 https://doi.org/10.1239/jap/1417528487

GOERTZEL, G. and KAHN, H. (1950). Monte Carlo Methods for Shield Computation. U.S. Atomic Energy Commission, Technical Information Division.

GOLIGHTLY, A., HENDERSON, D. A. and SHERLOCK, C. (2015). Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Stat. Comput.* **25** 1039–1055. MR3375634 https://doi.org/10.1007/s11222-014-9469-x

GORDON, N., SALMOND, J. and SMITH, A. F. M. (1993). A novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process.* **140** 107–113.

GRAMACY, R., SAMWORTH, R. and KING, R. (2010). Importance tempering. *Stat. Comput.* **20** 1–7. MR2578072 https://doi.org/10.1007/s11222-008-9108-5

GREEN, P. J., ŁATUSZYŃSKI, K., PEREYRA, M. and ROBERT, C. P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.* **25** 835–862. MR3360496 https://doi.org/10.1007/s11222-015-9574-5

GUBERNATIS, J. E. (2005). Marshall Rosenbluth and the Metropolis algorithm. *Phys. Plasmas* **12** 057303. MR2146373 https://doi.org/10.1063/1.1887186

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Wiley, New York.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

HIGDON, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.* **93** 585–595.

HITCHCOCK, D. B. (2003). A history of the Metropolis–Hastings algorithm. *Amer. Statist.* **57** 254–257. MR2037852 https://doi.org/10.1198/0003130032413

HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

HOOPER, M. (2013). Richard Price, Bayes' theorem, and God. *Significance* **10** 36–39.

HUBER, M. L. (2016). *Perfect Simulation*. *Monographs on Statistics and Applied Probability* **148**. CRC Press, Boca Raton, FL. MR3443710

IBA, Y. (2000). Population-based Monte Carlo algorithms. *Trans. Japanese Soc. Artificial Intell.* **16** 279–286.

JACOB, P. E., O'LEARY, J. and ATCHADÉ, Y. F. (2020). Unbiased Markov chain Monte Carlo methods with couplings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 543–600. MR4112777 https://doi.org/10.1111/rssb.12336

JACOB, P., ROBERT, C. P. and SMITH, M. H. (2011). Using parallel computation to improve independent Metropolis–Hastings based estimation. *J. Comput. Graph. Statist.* **20** 616–635. MR2878993 https://doi.org/10.1198/jcgs.2011.10167

JACQUIER, R., POLSON, N. G. and ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility models. *J. Bus. Econom. Statist.* **12** 371–389.

JAHAN, F., ULLAH, I. and MENGERSEN, K. L. (2020). A survey of Bayesian statistical approaches for big data. In *Case Studies in Applied Bayesian Data Science* (K. Mengersen, P. Pudlo and C. P. Robert, eds.). *Lecture Notes in Math.* **2259** 17–44. Springer, Cham. MR4175775 https://doi.org/10.1007/978-3-030-42553-1_2

JASRA, A. (2015). Approximate Bayesian computation for a class of time series models. *Int. Stat. Rev.* **83** 405–435. MR3429284 https://doi.org/10.1111/insr.12089

JASRA, A., SINGH, S. S., MARTIN, J. S. and MCCOY, E. (2012). Filtering via approximate Bayesian computation. *Stat. Comput.* **22** 1223–1237. MR2992296 https://doi.org/10.1007/s11222-010-9185-0

JOHNDROW, J. E., SMITH, A., PILLAI, N. and DUNSON, D. B. (2019). MCMC for imbalanced categorical data. *J. Amer. Statist. Assoc.* **114** 1394–1403. MR4011787 https://doi.org/10.1080/01621459.2018.1505626

JORDAN, M., GHAHRAMANI, Z. and JAAKKOLA, T. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **6** 183–233.

KAHN, H. (1949). Stochastic (Monte Carlo) Attenuation Analysis. Technical Report, RAND Corporation R-163.

KIM, S., SHEPHARD, N. and CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65** 361–393.

KLOEK, T. and VAN DIJK, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* **46** 1–19.

KNOBLAUCH, J., JEWSON, J. and DAMOULAS, T. (2022). An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *J. Mach. Learn. Res.* **23** 1–109.

KOOP, G. M. (2003). *Bayesian Econometrics*. Wiley, New York.

LEMIEUX, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling*. *Springer Series in Statistics*. Springer, New York. MR2723077

LINDLEY, D. V. (1980). Approximate Bayesian methods. In *Bayesian Statistics*: *Proceedings of the First International Meeting Held in Valencia* (*Spain*), *May* 28 *to June* 2, 1979.

LIU, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. Springer, New York. MR1842342

LIU, J. S., LIANG, F. and WONG, W. H. (2000). The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc.* **95** 121–134. MR1803145 https://doi.org/10.2307/2669532

LLORENTE, F., MARTINO, L., DELGADO, D. and LOPEZ-SANTIAGO, J. (2021). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. Available at arXiv:2005.08334.

LUO, X. and TJELMELAND, H. (2019). A multiple-try Metropolis–Hastings algorithm with tailored proposals. *Comput. Statist.* **34** 1109–1133. MR3982858 https://doi.org/10.1007/s00180-019-00878-y

LYNE, A.-M., GIROLAMI, M., ATCHADÉ, Y., STRATHMANN, H. and SIMPSON, D. (2015). On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Statist. Sci.* **30** 443–467. MR3432836 https://doi.org/10.1214/15-STS523

MARIN, J.-M., PUDLO, P., ROBERT, C. P. and RYDER, R. J. (2012). Approximate Bayesian computational methods. *Stat. Comput.* **22** 1167–1180. MR2992292 https://doi.org/10.1007/s11222-011-9288-2

MARINARI, E. and PARISI, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.* **19** 451–458.

MARTIN, G. M., FRAZIER, D. T. and ROBERT, C. P. (2020). Computing Bayes: Bayesian Computation from 1763 to the 21st Century. Available at arXiv:2004.06425.

MARTIN, G. M., FRAZIER, D. T. and ROBERT, C. P. (2023). Approximating Bayes in the 21st century. *Statist. Sci.* **38**. https://doi.org/10.1214/22-STS875

MARTINO, L. (2018). A review of multiple try MCMC algorithms for signal processing. *Digit. Signal Process.* **75** 134–152. MR3765591 https://doi.org/10.1016/j.dsp.2018.01.004

MARTINO, S. and RIEBLER, A. (2019). Integrated nested Laplace approximations (INLA). Available at arXiv:1907.01248.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MIDDLETON, L., DELIGIANNIDIS, G., DOUCET, A. and JACOB, P. E. (2020). Unbiased Markov chain Monte Carlo for intractable target distributions. *Electron. J. Stat.* **14** 2842–2891. MR4132645 https://doi.org/10.1214/20-EJS1727

NAESSETH, C. A., LINDSTEN, F., SCHÖN, T. B. et al. (2019). Elements of sequential Monte Carlo. *Found. Trends Mach. Learn.* **12** 307–392.

NEAL, R. (1996). Sampling from multimodal distributions using tempered transitions. *Stat. Comput.* **6** 353–366.

NEAL, R. (1998). Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation. *Learning in Graphical Models* 205–228.

NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

NEAL, R. M. (2011a). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

NEAL, R. (2011b). MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression. Available at arXiv:1101.0387.

NEISWANGER, W., WANG, C. and XING, E. (2013). Asymptotically Exact, Embarrassingly Parallel MCMC. Available at arXiv:1311.4780.

NEMETH, C. and FEARNHEAD, P. (2021). Stochastic gradient Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **116** 433–450. MR4227705 https://doi.org/10.1080/01621459.2020.1847120

NOTT, D. J. and KOHN, R. (2005). Adaptive sampling for Bayesian variable selection. *Biometrika* **92** 747–763. MR2234183 https://doi.org/10.1093/biomet/92.4.747

ORMEROD, J. T. and WAND, M. P. (2010). Explaining variational approximations. *Amer. Statist.* **64** 140–153. MR2757005 https://doi.org/10.1198/tast.2010.09058

OWEN, A. B. (2017). Statistically efficient thinning of a Markov chain sampler. *J. Comput. Graph. Statist.* **26** 738–744. MR3698682 https://doi.org/10.1080/10618600.2017.1336446

PITT, M. K., SILVA, R. D. S., GIORDANI, P. and KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Econometrics* **171** 134–151. MR2991856 https://doi.org/10.1016/j.jeconom.2012.06.004

POLLOCK, M., FEARNHEAD, P., JOHANSEN, A. M. and ROBERTS, G. O. (2020). Quasi-stationary Monte Carlo and the ScaLE algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 1167–1221. MR4176340

POLSON, N. G., CARLIN, B. P. and STOFFER, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state–space modeling. *J. Amer. Statist. Assoc.* **87** 493–500.

PRICE, R. (1764). A demonstration of the second rule in the essay towards the solution of a problem in the doctine of chances. *Philos. Trans. R. Soc. Lond., B* **54** 296–325.

PRICE, L. F., DROVANDI, C. C., LEE, A. and NOTT, D. J. (2018). Bayesian synthetic likelihood. *J. Comput. Graph. Statist.* **27** 1–11. MR3788296 https://doi.org/10.1080/10618600.2017.1302882

PRITCHARD, J. K., SEIELSTAD, M. T., PEREZ-LEZAUN, A. and FELDMAN, M. W. (1999). Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16** 1791–1798. https://doi.org/10.1093/oxfordjournals.molbev.a026091

PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms* (*Atlanta*, *GA*, 1995) **9** 223–252. MR1611693 https://doi.org/10.1002/(SICI)1098-2418(199608/09)9:1/2<223::AID-RSA14>3.3.CO;2-R

QUIROZ, M., TRAN, M.-N., VILLANI, M. and KOHN, R. (2018). Speeding up MCMC by delayed acceptance and data subsampling. *J. Comput. Graph. Statist.* **27** 12–22. MR3788297 https://doi.org/10.1080/10618600.2017.1307117

QUIROZ, M., KOHN, R., VILLANI, M. and TRAN, M.-N. (2019). Speeding up MCMC by efficient data subsampling. *J. Amer. Statist. Assoc.* **114** 831–843. MR3963184 https://doi.org/10.1080/01621459.2018.1448827

RITTER, C. and TANNER, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy–Gibbs sampler. *J. Amer. Statist. Assoc.* **87** 861–868.

ROBERT, C. P. (2007). *The Bayesian Choice*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR2723361

ROBERT, C. P. (2015). The Metropolis–Hastings algorithm. In *Wiley StatsRef*: *Statistics Reference Online* 1–15. Wiley StatsRef: Statistics Reference Online. American Cancer Societ.

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR2080278 https://doi.org/10.1007/978-1-4757-4145-2

ROBERT, C. and CASELLA, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statist. Sci.* **26** 102–115. MR2849912 https://doi.org/10.1214/10-STS351

ROBERT, C. P., ELVIRA, V., TAWN, N. and WU, C. (2018). Accelerating MCMC algorithms. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1435. MR3850448 https://doi.org/10.1002/wics.1435

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

ROBERTS, G. O. and ROSENTHAL, J. S. (1999). Convergence of slice sampler Markov chains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 643–660. MR1707866 https://doi.org/10.1111/1467-9868.00198

ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367. MR2749836 https://doi.org/10.1198/jcgs.2009.06134

ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 https://doi.org/10.1111/1467-9868.00070

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

ROSENTHAL, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 93–111. CRC Press, Boca Raton, FL. MR2858446

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

RUE, H., RIEBLER, A., SØRBYE, S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: A review. *Annu. Rev. Stat. Appl.* **4** 395–421.

SCHMON, S. M., CANNON, P. W. and KNOBLAUCH, J. (2020). Generalized posteriors in approximate Bayesian computation. ArXiv preprint. Available at arXiv:2011.08644.

SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *Int. J. Manag. Sci. Eng. Manag.* **11** 78–88.

SISSON, S. A., FAN, Y. and BEAUMONT, M. (2019). *Handbook of Approximate Bayesian Computation*. Chapman & Hall/CRC, Boca Raton, FL.

SISSON, S. A., FAN, Y. and TANAKA, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* **104** 1760–1765. MR2301870 https://doi.org/10.1073/pnas.0607208104

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55** 3–23. MR1210421

STIGLER, S. M. (1975). Studies in the history of probability and statistics. XXXIV. Napoleonic statistics: The work of Laplace. *Biometrika* **62** 503–517. MR0436395 https://doi.org/10.1093/biomet/62.2.503

STIGLER, S. M. (1986a). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR1057346

STIGLER, S. M. (1986b). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378. MR0858515

STIGLER, S. M. (2018). Richard Price, the first Bayesian. *Statist. Sci.* **33** 117–125. MR3757508 https://doi.org/10.1214/17-STS635

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. MR0898357

TAVARÉ, S. D., BALDING, D. J., GRIFFITH, R. C. and DONNELLY, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145** 505–518.

TAWN, N. G., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Weight-preserving simulated tempering. *Stat. Comput.* **30** 27–41. MR4057469 https://doi.org/10.1007/s11222-019-09863-3

TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166 https://doi.org/10.1214/aos/1176325750

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567

TIERNEY, L., KASS, R. E. and KADANE, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** 710–716. MR1132586

TIERNEY, L. and MIRA, A. (1998). Some adaptive Monte Carlo methods for Bayesian inference. *Stat. Med.* **18** 2507–2515.

TRAN, K. T. (2018). A Common Derivation for Markov Chain Monte Carlo Algorithms with Tractable and Intractable Targets. Available at arXiv:1607.01985.

TRAN, M.-N., NOTT, D. J. and KOHN, R. (2017). Variational Bayes with intractable likelihood. *J. Comput. Graph. Statist.* **26** 873–882. MR3765351 https://doi.org/10.1080/10618600.2017.1330205

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

VAN NIEKERK, J., BAKKA, H., RUE, H. and SCHENK, O. (2019). New frontiers in Bayesian modeling using the INLA package in R. ArXiv preprint. Available at arXiv:1907.10426.

WANG, Y. and BLEI, D. (2019). Variational Bayes under model misspecification. In *Advances in Neural Information Processing Systems* 13357–13367.

WANG, X. and DUNSON, D. B. (2013). Parallel MCMC via Weierstrass Sampler. Available at arXiv:1312.4605.

WIQVIST, S., PICCHINI, U., FORMAN, J. L., LINDORFF-LARSEN, K. and BOOMSMA, W. (2018). Accelerating delayed-acceptance Markov chain Monte Carlo algorithms. Available at arXiv:1806.05982.

WOOD, S. N. (2020). Simplified integrated nested Laplace approximation. *Biometrika* **107** 223–230. MR4064150 https://doi.org/10.1093/biomet/asz044

ZANELLA, G. and ROBERTS, G. (2019). Scalable importance tempering and Bayesian variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 489–517. MR3961496

ZHANG, C., BUTEPAGE, J., KJELLSTROM, H. and MANDT, S. (2019). Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 2008–2026. https://doi.org/10.1109/TPAMI.2018.2889774

Statistical Science
2024, Vol. 39, No. 1, 20–45
https://doi.org/10.1214/22-STS875

Approximating Bayes in the 21st Century

Gael M. Martin, David T. Frazier and Christian P. Robert

Abstract. The 21st century has seen an enormous growth in the development and use of approximate Bayesian methods. Such methods produce computational solutions to certain "intractable" statistical problems that challenge exact methods like Markov chain Monte Carlo: for instance, models with unavailable likelihoods, high-dimensional models and models featuring large data sets. These approximate methods are the subject of this review. The aim is to help new researchers in particular—and more generally those interested in adopting a Bayesian approach to empirical work—distinguish between different approximate techniques, understand the sense in which they are approximate, appreciate when and why particular methods are useful and see the ways in which they can can be combined.

Key words and phrases: Approximate Bayesian inference, intractable Bayesian problems, approximate Bayesian computation, Bayesian synthetic likelihood, variational Bayes, integrated nested Laplace approximation.

REFERENCES

ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.* **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

AN, Z., SOUTH, L. F. and DROVANDI, C. (2022). BSL: An R package for efficient parameter estimation for simulation-based models via Bayesian synthetic likelihood. *J. Stat. Softw.* **101** 1–33.

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

ARDIA, D., BAŞTÜRK, N., HOOGERHEIDE, L. and VAN DIJK, H. K. (2012). A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Comput. Statist. Data Anal.* **56** 3398–3414. MR2943902 https://doi.org/10.1016/j.csda.2010.09.001

BARDENET, R., DOUCET, A. and HOLMES, C. (2017). On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.* **18** Paper No. 47, 43 pp. MR3670492

BARTHELMÉ, S. and CHOPIN, N. (2014). Expectation propagation for likelihood-free inference. *J. Amer. Statist. Assoc.* **109** 315–333. MR3180566 https://doi.org/10.1080/01621459.2013.864178

BARTHELMÉ, S., CHOPIN, N. and COTTET, V. (2019). Divide and conquer in ABC: Expectation-propagation algorithms for likelihood-free inference. In *Handbook of Approximate Bayesian Computation* (S. Sisson, Y. Fan and M. Beaumont, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 415–434. CRC Press, Boca Raton, FL. MR3889291

BAUWENS, L. and RICHARD, J. (1985). A 1-1 poly-t random variable generator with application to Monte Carlo integration. *J. Econometrics* **29** 19–46.

BEAUMONT, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160. https://doi.org/10.1093/genetics/164.3.1139

BEAUMONT, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* **41** 379–406.

BEAUMONT, M. A. (2019). Approximate Bayesian computation. *Annu. Rev. Stat. Appl.* **6** 379–403. MR3939526 https://doi.org/10.1146/annurev-statistics-030718-105212

BEAUMONT, M., ZHANG, W. and BALDING, D. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162** 2025–2035.

BEAUMONT, M. A., CORNUET, J.-M., MARIN, J.-M. and ROBERT, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* **96** 983–990. MR2767283 https://doi.org/10.1093/biomet/asp052

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611 https://doi.org/10.1007/978-1-4757-4286-2

BERILD, M. O., MARTINO, S., GÓMEZ-RUBIO, V. and RUE, H. (2022). Importance sampling with the integrated nested Laplace approximation. *J. Comput. Graph. Statist.* **31** 1225–1237. MR4513383 https://doi.org/10.1080/10618600.2022.2067551

BERNTON, E., JACOB, P. E., GERBER, M. and ROBERT, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 235–269. MR3928142 https://doi.org/10.1111/rssb.12312

BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B* **55** 25–37. MR1210422

Gael M. Martin is Professor, Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia (e-mail: gael.martin@monash.edu). David T. Frazier is Associate Professor, Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia (e-mail: david.frazier@monash.edu). Christian P. Robert is Professor, Université Paris-Dauphine, Paris, France, and Department of Statistics, Warwick University, Coventry, UK (e-mail: xian@ceremade.dauphine.fr).

BETANCOURT, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. Preprint. Available at arXiv:1701.02434v2.

BILODEAU, B., STRINGER, A. and TANG, Y. (2021). Stochastic convergence rates and applications of adaptive quadrature in Bayesian inference. Preprint. Available at arXiv:2102.06801.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1103–1130. MR3557191 https://doi.org/10.1111/rssb.12158

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BLUM, M. G. B. (2010). Approximate Bayesian computation: A nonparametric perspective. *J. Amer. Statist. Assoc.* **105** 1178–1187. MR2752613 https://doi.org/10.1198/jasa.2010.tm09448

BLUM, M. G. B. (2019). Regression approaches for ABC. In *Handbook of Approximate Bayesian Computation* (S. Sisson, Y. Fan and M. Beaumont, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 71–85. CRC Press, Boca Raton, FL. MR3889280

BLUM, M. G. B. and FRANÇOIS, O. (2010). Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* **20** 63–73. MR2578077 https://doi.org/10.1007/s11222-009-9116-0

BLUM, M. G. B., NUNES, M. A., PRANGLE, D. and SISSON, S. A. (2013). A comparative review of dimension reduction methods in approximate Bayesian computation. *Statist. Sci.* **28** 189–208. MR3112405 https://doi.org/10.1214/12-sts406

BORNN, L., PILLAI, N. S., SMITH, A. and WOODARD, D. (2017). The use of a single pseudo-sample in approximate Bayesian computation. *Stat. Comput.* **27** 583–590. MR3613586 https://doi.org/10.1007/s11222-016-9640-7

BORTOT, P., COLES, S. G. and SISSON, S. A. (2007). Inference for stereological extremes. *J. Amer. Statist. Assoc.* **102** 84–92. MR2345549 https://doi.org/10.1198/016214506000000988

BRAUN, M. and MCAULIFFE, J. (2010). Variational inference for large-scale models of discrete choice. *J. Amer. Statist. Assoc.* **105** 324–335. MR2757203 https://doi.org/10.1198/jasa.2009.tm08030

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. MR3938958 https://doi.org/10.1214/18-STS660

BROOKS, S., GELMAN, A. JONES, G. and MENG, X. (2011). *Handbook of Markov Chain Monte Carlo*. Taylor & Francis, London.

CALVET, L. E. and CZELLAR, V. (2015). Accurate methods for approximate Bayesian computation filtering. *J. Financ. Econom.* **13** 798–838.

CANALE, A. and RUGGIERO, M. (2016). Bayesian nonparametric forecasting of monotonic functional time series. *Electron. J. Stat.* **10** 3265–3286. MR3572849 https://doi.org/10.1214/16-EJS1190

CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76** 1–32.

CERUZZI, P. (2003). *A History of Modern Computing*, 2nd ed. MIT Press, Cambridge, MA.

CHAN, J. C. C. and YU, X. (2022). Fast and accurate variational inference for large Bayesian VARs with stochastic volatility. *J. Econom. Dynam. Control* **143** Paper No. 104505, 19 pp. MR4470727 https://doi.org/10.1016/j.jedc.2022.104505

CHEN, S., DICK, J. and OWEN, A. B. (2011). Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. *Ann. Statist.* **39** 673–701. MR2816335 https://doi.org/10.1214/10-AOS831

CHIB, S. (2011). Introduction to simulation and MCMC methods. In *The Oxford Handbook of Bayesian Econometrics* (J. Geweke, G. Koop and H. van Dijk, eds.) 183–217. Oxford University Press, London.

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis–Hastings algorithm. *Amer. Statist.* **49** 327–335.

CHIB, S. and GREENBERG, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* **12** 409–431. MR1409939 https://doi.org/10.1017/S0266466600006794

CHIB, S., SHIN, M. and SIMONI, A. (2018). Bayesian estimation and comparison of moment condition models. *J. Amer. Statist. Assoc.* **113** 1656–1668. MR3902236 https://doi.org/10.1080/01621459.2017.1358172

CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

CLARTÉ, G., ROBERT, C. P., RYDER, R. J. and STOEHR, J. (2021). Componentwise approximate Bayesian computation via Gibbs-like steps. *Biometrika* **108** 591–607. MR4298766 https://doi.org/10.1093/biomet/asaa090

DAVIS, P. J. and RABINOWITZ, P. (1975). *Methods of Numerical Integration*. *Computer Science and Applied Mathematics*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York. MR0448814

DEAN, T. A., SINGH, S. S., JASRA, A. and PETERS, G. W. (2014). Parameter estimation for hidden Markov models with intractable likelihoods. *Scand. J. Stat.* **41** 970–987. MR3277033 https://doi.org/10.1111/sjos.12077

DELIGIANNIDIS, G., DOUCET, A. and PITT, M. K. (2018). The correlated pseudomarginal method. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 839–870. MR3874301 https://doi.org/10.1111/rssb.12280

DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. MR0836973 https://doi.org/10.1007/978-1-4613-8643-8

DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

DROVANDI, C. and FRAZIER, D. T. (2022). A comparison of likelihood-free methods with and without summary statistics. *Stat. Comput.* **32** Paper No. 42, 23 pp. MR4426803 https://doi.org/10.1007/s11222-022-10092-4

DROVANDI, C. C., PETTITT, A. N. and FADDY, M. J. (2011). Approximate Bayesian computation using indirect inference. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **60** 317–337. MR2767849 https://doi.org/10.1111/j.1467-9876.2010.00747.x

DROVANDI, C. C., PETTITT, A. N. and LEE, A. (2015). Bayesian indirect inference using a parametric auxiliary model. *Statist. Sci.* **30** 72–95. MR3317755 https://doi.org/10.1214/14-STS498

DUNSON, D. B. and JOHNDROW, J. E. (2020). The Hastings algorithm at fifty. *Biometrika* **107** 1–23. MR4064137 https://doi.org/10.1093/biomet/asz066

ELVIRA, V. and MARTINO, L. (2021). Advances in importance sampling. Preprint. Available at arXiv:2102.054071.

FEARNHEAD, P. (2019). Asymptotics of ABC. In *Handbook of Approximate Bayesian Computation* (S. Sisson, Y. Fan and M. Beaumont, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 269–288. CRC Press, Boca Raton, FL. MR3889287

FEARNHEAD, P. and PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 419–474. MR2925370 https://doi.org/10.1111/j.1467-9868.2011.01010.x

FIENBERG, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Anal*. **1** 1–40. MR2227361 https://doi.org/10.1214/06-BA101

FRAZIER, D. T. (2020). Approximate Bayesian Computation: A minimum distance approach. Preprint. Available at arXiv:2006.14126.

FRAZIER, D. T. and DROVANDI, C. (2021). Robust approximate Bayesian inference with synthetic likelihood. *J. Comput. Graph. Statist*. **30** 958–976. MR4356598 https://doi.org/10.1080/10618600.2021.1875839

FRAZIER, D. T., DROVANDI, C. and NOTT, D. J. (2021). Synthetic likelihood in misspecified models: Consequences and corrections. Preprint. Available at arXiv:2104.03436.

FRAZIER, D. T., LOAIZA-MAYA, R. and MARTIN, G. M. (2022). Variational Bayes in state space models: Inferential and predictive accuracy. *J. Comput. Graph. Statist*. https://doi.org/10.1080/10618600.2022.2134875

FRAZIER, D. T., ROBERT, C. P. and ROUSSEAU, J. (2020). Model misspecification in approximate Bayesian computation: Consequences and diagnostics. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **82** 421–444. MR4084170 https://doi.org/10.1111/rssb.12356

FRAZIER, D. T., MARTIN, G. M., ROBERT, C. P. and ROUSSEAU, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika* **105** 593–607. MR3842887 https://doi.org/10.1093/biomet/asy027

FRAZIER, D. T., MANEESOONTHORN, W., MARTIN, G. M. and MCCABE, B. P. (2019a). Approximate Bayesian forecasting. *Int. J. Forecast*. **35** 521–539.

FRAZIER, D. T., NOTT, D. J., DROVANDI, C. and KOHN, R. (2019b). Bayesian inference using synthetic likelihood: Asymptotics and adjustments. Preprint. Available at arXiv:1902.04827.

FRAZIER, D. T., LOAIZA-MAYA, R., MARTIN, G. M. and KOO, B. (2022). Loss-based variational Bayes prediction. Preprint. Available at arXiv:2104.14054.

GALLANT, A. R. and TAUCHEN, G. (1996). Which moments to match? *Econometric Theory* **12** 657–681. MR1422547 https://doi.org/10.1017/S0266466600006976

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc*. **85** 398–409. MR1141740

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell*. **6** 721–741. https://doi.org/10.1109/tpami.1984.4767596

GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 509–579. MR3351446 https://doi.org/10.1111/rssb.12104

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57** 1317–1339. MR1035115 https://doi.org/10.2307/1913710

GEWEKE, J., KOOP, G. and VAN DIJK, H. (2011). *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press, London.

GEYER, C. J. (2011). Introduction to Markov chain Monte Carlo. 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* 3–48. CRC Press, Boca Raton, FL. MR2858443

GÓMEZ-RUBIO, V. and RUE, H. (2018). Markov chain Monte Carlo with the integrated nested Laplace approximation. *Stat. Comput*. **28** 1033–1051. MR3835632 https://doi.org/10.1007/s11222-017-9778-y

GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* 2672–2680.

GORDON, N., SALMOND, J. and SMITH, A. (1993). A novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process*. **140** 107–113.

GOURIÉROUX, C., MONFORT, A. and RENAULT, E. (1993). Indirect inference. *J. Appl. Econometrics* **8** 85–118.

GREEN, P. J., ŁATUSZYŃSKI, K., PEREYRA, M. and ROBERT, C. P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput*. **25** 835–862. MR3360496 https://doi.org/10.1007/s11222-015-9574-5

GUTMANN, M. U. and CORANDER, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *J. Mach. Learn. Res*. **17** Paper No. 125, 47 pp. MR3555016

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Methuen & Co., Ltd., London. MR0223065

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res*. **14** 1303–1347. MR3081926

HOOGERHEIDE, L. F., VAN DIJK, H. K. and VAN OEST, R. D. (2009). Simulation based Bayesian econometric inference: Principles and some recent computational advances. In *Handbook of Computational Econometrics* (H. van Dijk and R. van Oest, eds.) 215–280. Wiley, New York.

HUGGINS, J., KASPRZAK, M., CAMPBELL, T. and BRODERICK, T. (2020). Validated variational inference via practical posterior error bounds. In *International Conference on Artificial Intelligence and Statistics* 1792–1802. PMLR.

JASRA, A. (2015). Approximate Bayesian computation for a class of time series models. *Int. Stat. Rev*. **83** 405–435. MR3429284 https://doi.org/10.1111/insr.12089

JASRA, A., SINGH, S. S., MARTIN, J. S. and MCCOY, E. (2012). Filtering via approximate Bayesian computation. *Stat. Comput*. **22** 1223–1237. MR2992296 https://doi.org/10.1007/s11222-010-9185-0

JENNINGS, E. and MADIGAN, M. (2017). AstroABC: An approximate Bayesian computation sequential Monte Carlo sampler for cosmological parameter estimation. *Astron. Comput*. **19** 16–22.

JIANG, B. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. In *International Conference on Artificial Intelligence and Statistics* 1711–1721. PMLR.

JIANG, W. and TANNER, M. A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann. Statist*. **36** 2207–2231. MR2458185 https://doi.org/10.1214/07-AOS547

JOHNDROW, J. E., SMITH, A., PILLAI, N. and DUNSON, D. B. (2019). MCMC for imbalanced categorical data. *J. Amer. Statist. Assoc*. **114** 1394–1403. MR4011787 https://doi.org/10.1080/01621459.2018.1505626

JOYCE, P. and MARJORAM, P. (2008). Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mol. Biol*. **7** Art. 26, 18 pp. MR2438407 https://doi.org/10.2202/1544-6115.1389

KABISA, S., DUNSON, D. B. and MORRIS, J. S. (2016). Online variational Bayes inference for high-dimensional correlated data. *J. Comput. Graph. Statist*. **25** 426–444. MR3499688 https://doi.org/10.1080/10618600.2014.998336

KLOEK, T. and VAN DIJK, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* **46** 1–19.

KNOBLAUCH, J., JEWSON, J. and DAMOULAS, T. (2019). Generalized variational inference. Preprint. Available at arXiv:1904.02063.

Kon Kam King, G., Canale, A. and Ruggiero, M. (2019). Bayesian functional forecasting with locally-autoregressive dependent processes. *Bayesian Anal*. **14** 1121–1141. MR4044848 https://doi.org/10.1214/18-BA1140

Koop, G. and Korobilis, D. (2018). Variational Bayes inference in high-dimensional time-varying parameter models. SSRN 3246472.

Kousathanas, A., Duchen, P. and Wegmann, D. (2019). A guide to general-purpose ABC software. In *Handbook of Approximate Bayesian Computation*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 369–413. CRC Press, Boca Raton, FL. MR3889290

Kousathanas, A., Leuenberger, C., Helfer, J., Quinodoz, M., Foll, M. and Wegmann, D. (2016). Likelihood-free inference in high-dimensional models. *Genetics* **203** 893–904.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. and Blei, D. M. (2017). Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18** Paper No. 14, 45 pp. MR3634881

Lemieux, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling*. *Springer Series in Statistics*. Springer, New York. MR2723077

Li, W. and Fearnhead, P. (2018a). Convergence of regression-adjusted approximate Bayesian computation. *Biometrika* **105** 301–318. MR3804404 https://doi.org/10.1093/biomet/asx081

Li, W. and Fearnhead, P. (2018b). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika* **105** 285–299. MR3804403 https://doi.org/10.1093/biomet/asx078

Lintusaari, J., Gutmann, M. U., Dutta, R., Kaski, S. and Corander, J. (2017). Fundamentals and recent developments in approximate Bayesian computation. *Syst. Biol.* **66** e66–e82.

Llorente, F., Martino, L., Delgado, D. and Lopez-Santiago, J. (2021). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. Preprint. Available at arXiv:2005.08334.

Loaiza-Maya, R., Martin, G. M. and Frazier, D. T. (2021). Focused Bayesian prediction. *J. Appl. Econometrics* **36** 517–543. MR4309597 https://doi.org/10.1002/jae.2810

Loaiza-Maya, R., Smith, M. S., Nott, D. J. and Danaher, P. J. (2022). Fast and accurate variational inference for models with many latent variables. *J. Econometrics* **230** 339–362. MR4466728 https://doi.org/10.1016/j.jeconom.2021.05.002

Margossian, C., Vehtari, A., Simpson, D. and Agrawal, R. (2020). Hamiltonian Monte Carlo using an adjoint-differentiated Laplace approximation: Bayesian inference for latent Gaussian models and beyond. *Adv. Neural Inf. Process. Syst.* **33** 9086–9097.

Marin, J.-M., Pudlo, P., Robert, C. P. and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Stat. Comput.* **22** 1167–1180. MR2992292 https://doi.org/10.1007/s11222-011-9288-2

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* **100** 15324–15328.

Martin, G. M., Frazier, D. T. and Robert, C. P. (2020). Computing Bayes: Bayesian computation from 1763 to the 21st century. Preprint. Available at arXiv:2004.06425.

Martin, G. M., Frazier, D. T. and Robert, C. P. (2024). Computing Bayes: From then 'til now. *Statist. Sci.* **38** 3–19. https://doi.org/10.1214/22-STS876

Martin, G. M., McCabe, B. P. M., Frazier, D. T., Maneesoonthorn, W. and Robert, C. P. (2019). Auxiliary likelihood-based approximate Bayesian computation in state space models. *J. Comput. Graph. Statist.* **28** 508–522. MR4007737 https://doi.org/10.1080/10618600.2018.1552154

Martino, S. and Riebler, A. (2019). Integrated nested Laplace approximations (INLA). Preprint. Available at arXiv:1907.01248.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *J. Amer. Statist. Assoc.* **44** 335–341. MR0031341

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

Miller, J. W. and Dunson, D. B. (2019). Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* **114** 1113–1125. MR4011766 https://doi.org/10.1080/01621459.2018.1469995

Moores, M. T., Pettitt, A. N. and Mengersen, K. L. (2020). Bayesian computation with intractable likelihoods. In *Case Studies in Applied Bayesian Data Science*. *Lecture Notes in Math.* **2259** 137–151. Springer, Cham. MR4175779 https://doi.org/10.1007/978-3-030-42553-1_6

Naesseth, C. A., Lindsten, F., Schön, T. B. et al. (2019). Elements of sequential Monte Carlo. *Found. Trends Mach. Learn.* **12** 307–392.

Naylor, J. C. and Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *J. R. Stat. Soc. Ser. C* **31** 214–225. MR0694917 https://doi.org/10.2307/2347995

Nguyen, H. D., Arbel, J., Lü, H. and Forbes, F. (2020). Approximate Bayesian computation via the energy statistic. *IEEE Access* **8** 131683–131698.

Nott, D. J., Fan, Y., Marshall, L. and Sisson, S. A. (2014). Approximate Bayesian computation and Bayes' linear analysis: Toward high-dimensional ABC. *J. Comput. Graph. Statist.* **23** 65–86. MR3173761 https://doi.org/10.1080/10618600.2012.751874

Nott, D. J., Ong, V. M.-H., Fan, Y. and Sisson, S. A. (2019). High-dimensional ABC. In *Handbook of Approximate Bayesian Computation* (S. Sisson, Y. Fan and M. Beaumont, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 211–241. CRC Press, Boca Raton, FL. MR3889285

O'Hagan, A. and West, M., eds. (2010). *The Oxford Handbook of Applied Bayesian Analysis* Oxford Univ. Press, Oxford. MR2797133

Ong, V. M. H., Nott, D. J., Tran, M.-N., Sisson, S. A. and Drovandi, C. C. (2018). Variational Bayes with synthetic likelihood. *Stat. Comput.* **28** 971–988. MR3766054 https://doi.org/10.1007/s11222-017-9773-3

Ormerod, J. T. and Wand, M. P. (2010). Explaining variational approximations. *Amer. Statist.* **64** 140–153. MR2757005 https://doi.org/10.1198/tast.2010.09058

Pacchiardi, L. and Dutta, R. (2021). Generalized Bayesian likelihood-free inference using scoring rules estimators. Preprint. Available at arXiv:2104.03889.

Park, M., Jitkrittum, W. and Sejdinovic, D. (2016). K2-ABC: Approximate Bayesian computation with kernel embeddings. In *Artificial Intelligence and Statistics* 398–407. PMLR.

Peters, G. W., Sisson, S. A. and Fan, Y. (2012). Likelihood-free Bayesian inference for α-stable models. *Comput. Statist. Data Anal.* **56** 3743–3756. MR2943924 https://doi.org/10.1016/j.csda.2010.10.004

Price, L. F., Drovandi, C. C., Lee, A. and Nott, D. J. (2018). Bayesian synthetic likelihood. *J. Comput. Graph. Statist.* **27** 1–11. MR3788296 https://doi.org/10.1080/10618600.2017.1302882

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999). Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16** 1791–1798. https://doi.org/10.1093/oxfordjournals.molbev.a026091

Quiroz, M., Nott, D. J. and Kohn, R. (2022). Gaussian variational approximation for high-dimensional state space models. *Bayesian Anal.* To appear. https://doi.org/10.1214/22-BA1332

Quiroz, M., Tran, M.-N., Villani, M. and Kohn, R. (2018). Speeding up MCMC by delayed acceptance and data subsampling. *J. Comput. Graph. Statist.* **27** 12–22. MR3788297 https://doi.org/10.1080/10618600.2017.1307117

QUIROZ, M., KOHN, R., VILLANI, M. and TRAN, M.-N. (2019). Speeding up MCMC by efficient data subsampling. *J. Amer. Statist. Assoc.* **114** 831–843. MR3963184 https://doi.org/10.1080/01621459.2018.1448827

ROBERT, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1835885

ROBERT, C. and CASELLA, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statist. Sci.* **26** 102–115. MR2849912 https://doi.org/10.1214/10-STS351

ROBERT, C. P., ELVIRA, V., TAWN, N. and WU, C. (2018). Accelerating MCMC algorithms. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1435, 14 pp. MR3850448 https://doi.org/10.1002/wics.1435

RODRIGUES, G. S., NOTT, D. J. and SISSON, S. A. (2020). Likelihood-free approximate Gibbs sampling. *Stat. Comput.* **30** 1057–1073. MR4108691 https://doi.org/10.1007/s11222-020-09933-x

ROSSI, P. E. and ALLENBY, G. M. (2003). Bayesian statistics and marketing. *Mark. Sci.* **22** 304–328.

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. CRC Press/CRC, Boca Raton, FL. MR2130347 https://doi.org/10.1201/9780203492024

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

RUE, H., RIEBLER, A., SØRBYE, S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: A review. *Annu. Rev. Stat. Appl.* **4** 395–421.

SCHMON, S. M., CANNON, P. W. and KNOBLAUCH, J. (2020). Generalized posteriors in approximate Bayesian computation. Preprint. Available at arXiv:2011.08644.

SCHMON, S. M., DELIGIANNIDIS, G., DOUCET, A. and PITT, M. K. (2021). Large-sample asymptotics of the pseudo-marginal method. *Biometrika* **108** 37–51. MR4226188 https://doi.org/10.1093/biomet/asaa044

SISSON, S. A. and FAN, Y. (2011). Likelihood-free MCMC. 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* 313–335. CRC Press, Boca Raton, FL. MR2858454

SISSON, S. A. and FAN, Y. (2019). ABC samplers. In *Handbook of Approximate Bayesian Computation* (S. Sisson, Y. Fan and M. Beaumont, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 87–123. CRC Press, Boca Raton, FL. MR3889281

SISSON, S. A., FAN, Y. and BEAUMONT, A. (2019). *Handbook of Approximate Bayesian Computation*. Chapman & Hall/CRC, Boca Raton, FL. MR3930567

SISSON, S. A., FAN, Y. and TANAKA, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* **104** 1760–1765. MR2301870 https://doi.org/10.1073/pnas.0607208104

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55** 3–23. MR1210421

STIGLER, S. M. (1975). Studies in the history of probability and statistics. XXXIV. Napoleonic statistics: The work of Laplace. *Biometrika* **62** 503–517. MR0436395 https://doi.org/10.1093/biomet/62.2.503

STIGLER, S. M. (1986a). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR1057346

STIGLER, S. M. (1986b). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378. MR0858515

STOEHR, J. (2017). A review on statistical inference methods for discrete Markov random fields. Preprint. Available at arXiv:1704.03331.

STRINGER, A., BROWN, P. and STAFFORD, J. (2021). Fast, scalable approximations to posterior distributions in extended latent Gaussian models. Preprint. Available at arXiv:2103.07425.

TANG, Y. and REID, N. (2021). Laplace and saddlepoint approximations in high dimensions. Preprint. Available at arXiv:2107.10885.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. MR0898357

TAVARÉ, S., BALDING, D., GRIFFITH, R. and DONNELLY, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145** 505–518.

R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567

TIERNEY, L., KASS, R. E. and KADANE, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** 710–716. MR1132586

TOKDAR, S. and KASS, R. (2010). Importance sampling: A review. *Wiley Interdiscip. Rev.: Comput. Stat.* **2** 54–60.

TRAN, M.-N., NGUYEN, D. H. and NGUYEN, D. (2021). Variational Bayes on manifolds. *Stat. Comput.* **31** Paper No. 71, 17 pp. MR4316144 https://doi.org/10.1007/s11222-021-10047-1

TRAN, M.-N., NOTT, D. J. and KOHN, R. (2017). Variational Bayes with intractable likelihood. *J. Comput. Graph. Statist.* **26** 873–882. MR3765351 https://doi.org/10.1080/10618600.2017.1330205

TURNER, B. M. and SEDERBERG, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychon. Bull. Rev.* **21** 227–250.

VAN NIEKERK, J., BAKKA, H., RUE, H. and SCHENK, O. (2019). New frontiers in Bayesian modeling using the INLA package in R. Preprint. Available at arXiv:1907.10426.

VANHATALO, J., PIETILÄINEN, V. and VEHTARI, A. (2010). Approximate inference for disease mapping with sparse Gaussian processes. *Stat. Med.* **29** 1580–1607. MR2758849 https://doi.org/10.1002/sim.3895

VANSLETTE, K., AL ALSHEIKH, A. and YOUCEF-TOUMI, K. (2020). Why simple quadrature is just as good as Monte Carlo. *Monte Carlo Methods Appl.* **26** 1–16. MR4072613 https://doi.org/10.1515/mcma-2020-2055

WAND, M. P. (2017). Fast approximate inference for arbitrarily large semiparametric regression models via message passing. *J. Amer. Statist. Assoc.* **112** 137–156. MR3646558 https://doi.org/10.1080/01621459.2016.1197833

WANG, Y. and BLEI, D. (2019a). Variational Bayes under model misspecification. In *Advances in Neural Information Processing Systems* 13357–13367.

WANG, Y. and BLEI, D. M. (2019b). Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc.* **114** 1147–1161. MR4011769 https://doi.org/10.1080/01621459.2018.1473776

WILKINSON, R. D. (2013). Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Stat. Appl. Genet. Mol. Biol.* **12** 129–141. MR3071024 https://doi.org/10.1515/sagmb-2013-0010

WOOD, S. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466** 1102–1104.

WOOD, S. N. (2020). Simplified integrated nested Laplace approximation. *Biometrika* **107** 223–230. MR4064150 https://doi.org/10.1093/biomet/asz044

YAO, Y., VEHTARI, A., SIMPSON, D. and GELMAN, A. (2018). Yes, but did it work?: Evaluating variational inference. In *Proceedings of the* 35*th International Conference on Machine Learning* **80** 5581–5590.

YU, X., NOTT, D. J., TRAN, M.-N. and KLEIN, N. (2021). Assessment and adjustment of approximate inference algorithms using the law of total variance. *J. Comput. Graph. Statist.* **30** 977–990. MR4356599 https://doi.org/10.1080/10618600.2021.1880921

ZHANG, T. (2006a). From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Ann. Statist.* **34** 2180–2210. MR2291497 https://doi.org/10.1214/009053606000000704

ZHANG, T. (2006b). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Trans. Inf. Theory* **52** 1307–1321. MR2241190 https://doi.org/10.1109/TIT.2005.864439

ZHANG, F. and GAO, C. (2020). Convergence rates of variational posterior distributions. *Ann. Statist.* **48** 2180–2207. MR4134791 https://doi.org/10.1214/19-AOS1883

ZHANG, C., BÜTEPAGE, J., KJELLSTRÖM, H. and MANDT, S. (2018). Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 2008–2026.

Statistical Science
2024, Vol. 39, No. 1, 46–61
https://doi.org/10.1214/23-STS907

Past, Present and Future of Software for Bayesian Inference

Erik Štrumbelj, Alexandre Bouchard-Côté, Jukka Corander, Andrew Gelman, Håvard Rue, Lawrence Murray, Henri Pesonen, Martyn Plummer, and Aki Vehtari

Abstract. Software tools for Bayesian inference have undergone rapid evolution in the past three decades, following popularisation of the first generation MCMC-sampler implementations. More recently, exponential growth in the number of users has been stimulated both by the active development of new packages by the machine learning community and popularity of specialist software for particular applications. This review aims to summarize the most popular software and provide a useful map for a reader to navigate the world of Bayesian computation. We anticipate a vigorous continued development of algorithms and corresponding software in multiple research fields, such as probabilistic programming, likelihood-free inference and Bayesian neural networks, which will further broaden the possibilities for employing the Bayesian paradigm in exciting applications.

Key words and phrases: Statistics, data analysis, MCMC, computation, probabilistic programming.

REFERENCES

[1] ABADI, M. et al. (2015). *TensorFlow*: *Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.

[2] ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

[3] AN, Z., SOUTH, L. F. and DROVANDI, C. (2022). BSL: An R package for efficient parameter estimation for simulation-based models via Bayesian synthetic likelihood. *J. Stat. Softw.* **101** 1–33.

[4] BAKKA, H., RUE, H., FUGLSTAD, G.-A., RIEBLER, A., BOLIN, D., ILLIAN, J., KRAINSKI, E., SIMPSON, D. and LINDGREN, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdiscip. Rev.*: *Comput. Stat.* **10** e1443. MR3873676 https://doi.org/10.1002/wics.1443

[5] BAUDART, G., BURRONI, J., HIRZEL, M., MANDEL, L. and SHINNAR, A. (2021). Compiling Stan to generative probabilistic languages and extension to deep probabilistic programming. In *Proceedings of the* 42*nd ACM SIGPLAN International Conference on Programming Language Design and Implementation* 497–510.

[6] BAYDIN, A. G., PEARLMUTTER, B. A., RADUL, A. A. and SISKIND, J. M. (2017). Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* **18** 153. MR3800512

[7] BERAHA, M., FALCO, D. and GUGLIELMI, A. (2021). JAGS, NIMBLE, Stan: A detailed comparison among Bayesian MCMC software. Available at arXiv:2107.09357.

[8] BERGSTRA, J., BREULEUX, O., BASTIEN, F., LAMBLIN, P., PASCANU, R., DESJARDINS, G., TURIAN, J., WARDE-FARLEY, D. and BENGIO, Y. (2010). Theano: A CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference* (*SciPy*) **4** 1–7. Austin, TX.

[9] BETANCOURT, M. (2016). Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo. Available at arXiv:1604.00695.

[10] BETANCOURT, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. Available at arXiv:1701.02434.

Erik Štrumbelj is a Professor, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia (e-mail: erik.strumbelj@fri.uni-lj.si). Alexandre Bouchard-Côté is a Professor, Statistics at the University of British Columbia, Vancouver, Canada (e-mail: bouchard@stat.ubc.ca). Jukka Corander is a Professor, Faculty of Medicine, University of Oslo, Oslo, Norway, Faculty of Science, University of Helsinki, Helsinki, Finland, and Associate Faculty Member at Wellcome Sanger Institute, Hinxton, Cambridge, UK (e-mail: jukka.corander@medisin.uio.no). Andrew Gelman is a Professor, Statistics and Political Science at Columbia University, New York, New York 10027 USA (e-mail: gelman@stat.columbia.edu). Håvard Rue is a Professor, Statistics at King Abdullah University of Science and Technology, Thuwal, Saudi Arabia (e-mail: haavard.rue@kaust.edu.sa). Lawrence Murray has contributed to this work as an independent researcher (e-mail: lawrence@indii.org). Henri Pesonen is a Researcher, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway (e-mail: henri.pesonen@medisin.uio.no). Martyn Plummer is a Professor, Statistics at the University of Warwick, Warwick, UK (e-mail: martyn.plummer@warwick.ac.uk). Aki Vehtari is a Professor, Computational Bayesian Modeling at Aalto University, Aalto, Finland (e-mail: aki.vehtari@aalto.fi).

[11] BETANCOURT, M., BYRNE, S., LIVINGSTONE, S. and GIROLAMI, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli* **23** 2257–2298. MR3648031 https://doi.org/10.3150/16-BEJ810

[12] BINGHAM, E., CHEN, J. P., JANKOWIAK, M., OBERMEYER, F., PRADHAN, N., KARALETSOS, T., SINGH, R., SZERLIP, P., HORSFALL, P. et al. (2019). Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.* **20** 973–978.

[13] BISHOP, C., SPIEGELHALTER, D. and WINN, J. (2002). VIBES: A variational inference engine for Bayesian networks. *Adv. Neural Inf. Process. Syst.* **15**.

[14] BOUCHARD-CÔTÉ, A., CHERN, K., CUBRANIC, D., HOSSEINI, S., HUME, J., LEPUR, M., OUYANG, Z. and SGARBI, G. (2022). Blang: Bayesian declarative modeling of general data structures and inference via algorithms based on distribution continua. *J. Stat. Softw.* **103** 1–98.

[15] BRADBURY, J., FROSTIG, R., HAWKINS, P., JOHNSON, M. J., LEARY, C., MACLAURIN, D., NECULA, G., PASZKE, A., VANDERPLAS, J. et al. (2018). JAX: Composable transformations of Python+NumPy programs.

[16] BÜRKNER, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80** 1–28.

[17] BÜRKNER, P.-C., GABRY, J., KAY, M. and VEHTARI, A. (2022). posterior: Tools for working with posterior distributions.

[18] CAMELETTI, M. and GÓMEZ-RUBIO, V. (2021). Software for Bayesian statistics. *J. Stat. Softw.* **100** 1–7.

[19] CAPRETTO, T., PIHO, C., KUMAR, R., WESTFALL, J., YARKONI, T. and MARTIN, O. A. (2022). Bambi: A simple interface for fitting Bayesian linear models in python. *J. Stat. Softw.* **103** 1–29.

[20] CARPENTER, B. (2021). What do we need from a probabilistic programming language to support Bayesian workflow? In *International Conference on Probabilistic Programming* (*PROBPROG*) 46.

[21] CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.

[22] CARPENTER, B., HOFFMAN, M. D., BRUBAKER, M., LEE, D., LI, P. and BETANCOURT, M. (2015). The Stan math library: Reverse-mode automatic differentiation in C++. Available at arXiv:1509.07164.

[23] CEŠNOVAR, R. (2022). Parallel computation in the Stan probabilistic programming language Ph.D. thesis Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.

[24] CIGLARIČ, T., ČEŠNOVAR, R. and ŠTRUMBELJ, E. (2020). Automated OpenCL GPU kernel fusion for Stan math. In *Proceedings of the International Workshop on OpenCL* 1–6.

[25] CRANMER, K., BREHMER, J. and LOUPPE, G. (2020). The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. USA* **117** 30055–30062. MR4263287 https://doi.org/10.1073/pnas.1912789117

[26] CSÁRDI, G. (2019). cranlogs: Download logs from the 'RStudio' 'CRAN' mirror. R package version 2.1.1.

[27] CSILLÉRY, K., FRANÇOIS, O. and BLUM, M. G. (2012). abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **3** 475–479.

[28] CUSUMANO-TOWNER, M. F., SAAD, F. A., LEW, A. K. and MANSINGHKA, V. K. (2019). Gen: A general-purpose probabilistic programming system with programmable inference. In *Proceedings of the* 40*th ACM Sigplan Conference on Programming Language Design and Implementation* 221–236.

[29] DAVIS, T. A. (2006). *Direct Methods for Sparse Linear Systems*. *Fundamentals of Algorithms* **2**. SIAM, Philadelphia, PA. MR2270673 https://doi.org/10.1137/1.9780898718881

[30] DE VALPINE, P., TUREK, D., PACIOREK, C. J., ANDERSON-BERGMAN, C., TEMPLE LANG, D. and BODIK, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Statist.* **26** 403–413. MR3640196 https://doi.org/10.1080/10618600.2016.1172487

[31] DILLON, J. V., LANGMORE, I., TRAN, D., BREVDO, E., VASUDEVAN, S., MOORE, D., PATTON, B., ALEMI, A., HOFFMAN, M. et al. (2017). Tensorflow distributions. Available at arXiv:1711.10604.

[32] DRUZDZEL, M. J. (1999). SMILE: Structural modeling, inference, and learning engine and GeNIe: A development environment for graphical decision-theoretic models. In *American Association for Artificial Intelligence Proceedings* 902–903.

[33] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

[34] DUTTA, R., SCHOENGENS, M., PACCHIARDI, L., UMMADISINGU, A., WIDMER, N., KÜNZLI, P., ONNELA, J.-P. and MIRA, A. (2021). ABCpy: A high-performance computing perspective to approximate Bayesian computation. *J. Stat. Softw.* **100** 1–38.

[35] FOREMAN-MACKEY, D., FARR, W. M., SINHA, M., ARCHIBALD, A. M., HOGG, D. W., SANDERS, J. S., ZUNTZ, J., WILLIAMS, P. K., NELSON, A. R. et al. (2019). emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC. Available at arXiv:1911.07688.

[36] FOREMAN-MACKEY, D., HOGG, D. W., LANG, D. and GOODMAN, J. (2013). emcee: The MCMC hammer. *Publ. Astron. Soc. Pac.* **125** 306.

[37] FOURNIER, D. A., SKAUG, H. J., ANCHETA, J., IANELLI, J., MAGNUSSON, A., MAUNDER, M. N., NIELSEN, A. and SIBERT, J. (2012). AD model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* **27** 233–249. MR2901959 https://doi.org/10.1080/10556788.2011.597854

[38] FRÜHWIRTH-SCHNATTER, S., FRÜHWIRTH, R., HELD, L. and RUE, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Stat. Comput.* **19** 479–492. MR2565319 https://doi.org/10.1007/s11222-008-9109-4

[39] GABRY, J. and ČEŠNOVAR, R. (2022). A lightweight R interface to CmdStan.

[40] GABRY, J. and MAHR, T. (2022). bayesplot: Plotting for Bayesian models. R package version 1.10.0.

[41] GABRY, J., SIMPSON, D., VEHTARI, A., BETANCOURT, M. and GELMAN, A. (2019). Visualization in Bayesian workflow. *J. Roy. Statist. Soc. Ser. A* **182** 389–402. MR3902665 https://doi.org/10.1111/rssa.12378

[42] GABRY, J. and VEEN, D. (2022). shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. R package version 2.6.0. Available at https://CRAN.R-project.org/package=shinystan.

[43] GAEDKE-MERZHÄUSER, L., VAN NIEKERK, J., SCHENK, O. and RUE, H. (2023). Parallelized integrated nested Laplace approximations for fast Bayesian inference. *Stat. Comput.* **33** 25. MR4526361 https://doi.org/10.1007/s11222-022-10192-1

[44] GE, H., XU, K. and GHAHRAMANI, Z. (2018). Turing: A language for flexible probabilistic inference. In *International Conference on Artificial Intelligence and Statistics* 1682–1690. PMLR.

[45] GELMAN, A., VEHTARI, A., SIMPSON, D., MARGOSSIAN, C. C., CARPENTER, B., YAO, Y., KENNEDY, L., GABRY, J., BÜRKNER, P.-C. et al. (2020). Bayesian workflow. Available at arXiv:2011.01808.

[46] GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. R. Stat. Soc., Ser. C* **41** 337–348.

[47] GOLDING, N. (2019). greta: Simple and scalable statistical modelling in R. *J. Open Sour. Softw.* **4** 1601.

[48] GOODMAN, N., MANSINGHKA, V., ROY, D. M., BONAWITZ, K. and TENENBAUM, J. B. (2012). Church: A language for generative models. Available at arXiv:1206.3255.

[49] GOODMAN, J. and WEARE, J. (2010). Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **5** 65–80. MR2600822 https://doi.org/10.2140/camcos.2010.5.65

[50] GOODRICH, B., ALI, I., GABRY, J. and SAM, B. (2021). rstanarm: Bayesian applied regression modeling via Stan R package version 2.21.3. https://CRAN.R-project.org/package=rstanarm.

[51] GORINOVA, M. I. (2022). Program analysis of probabilistic programs. Available at arXiv:2204.06868.

[52] GORINOVA, M. I., GORDON, A. D. and SUTTON, C. (2019). Probabilistic programming with densities in SlicStan: Efficient, flexible, and deterministic. *Proc. ACM Program. Lang.* **3** 1–30.

[53] GORINOVA, M., MOORE, D. and HOFFMAN, M. (2020). Automatic reparameterisation of probabilistic programs. In *International Conference on Machine Learning* 3648–3657. PMLR.

[54] GOUDIE, R. J., TURNER, R. M., DE ANGELIS, D. and THOMAS, A. (2020). MultiBUGS: A parallel implementation of the BUGS modelling framework for faster Bayesian inference. *J. Stat. Softw.* **95**.

[55] GRONAU, Q. F., SINGMANN, H. and WAGENMAKERS, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *J. Stat. Softw.* **92** 1–29.

[56] GUTMANN, M. U. and CORANDER, J. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *J. Mach. Learn. Res.* **17** 125. MR3555016

[57] HIGSON, E., HANDLEY, W., HOBSON, M. and LASENBY, A. (2019). Dynamic nested sampling: An improved algorithm for parameter estimation and evidence calculation. *Stat. Comput.* **29** 891–913. MR3994608 https://doi.org/10.1007/s11222-018-9844-0

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

[59] HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

[60] HOFFMAN, M. D., RADUL, A. and SOUNTSOV, P. (2021). An adaptive MCMC scheme for setting trajectory lengths in Hamiltonian Monte Carlo. *Int. Conf. Artif. Intell. Stat.*

[61] HOFFMAN, M. and SOUNTSOV, P. (2022). Tuning-free generalized Hamiltonian Monte Carlo. *Proc. Mach. Learn. Res.* **151** 7799–7813.

[62] HOLMES, C. C. and HELD, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.* **1** 145–168. MR2227368 https://doi.org/10.1214/06-BA105

[63] HUNTER, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9** 90–95.

[64] JABOT, F., FAURE, T. and DUMOULIN, N. (2013). Easy ABC: Performing efficient approximate Bayesian computation sampling schemes using R. *Methods Ecol. Evol.* **4** 684–687.

[65] KADIE, C. M., HOVEL, D. and HORVITZ, E. (2001). *MSBNx: A Component-Centric Toolkit for Modeling and Inference with Bayesian Networks*. Microsoft Research, Richmond, WA. Technical Report MSR-TR-2001-67 28.

[66] KALLIOINEN, N., PAANANEN, T., BÜRKNER, P.-C. and VEHTARI, A. (2021). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. Available at arXiv:2107.14054.

[67] KOUSATHANAS, A., DUCHEN, P. and WEGMANN, D. (2019). A guide to general-purpose ABC software. In *Handbook of Approximate Bayesian Computation. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 369–413. CRC Press, Boca Raton, FL. MR3889290

[68] KRAINSKI, E., GÓMEZ-RUBIO, V., BAKKA, H., LENZI, A., CASTRO-CAMILO, D., SIMPSON, D., LINDGREN, F. and RUE, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. CRC Press, Boca Raton, FL.

[69] KUCUKELBIR, A., TRAN, D., RANGANATH, R., GELMAN, A. and BLEI, D. M. (2017). Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18** 14. MR3634881

[70] KUMAR, R., CARROLL, C., HARTIKAINEN, A. and MARTÍN, O. A. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *J. Open Sour. Softw.*.

[71] LINDGREN, F. and RUE, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63** 1–25.

[72] LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

[73] LINTUSAARI, J., VUOLLEKOSKI, H., KANGASRÄÄSIÖ, A., SKYTÉN, K., JÄRVENPÄÄ, M., MARTTINEN, P., GUTMANN, M. U., VEHTARI, A., CORANDER, J. et al. (2018). ELFI: Engine for likelihood-free inference. *J. Mach. Learn. Res.* **19** 16. MR3862423

[74] LUNN, D., JACKSON, C., BEST, N., THOMAS, A. and SPIEGELHALTER, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press, Boca Raton, FL.

[75] LUNN, D., SPIEGELHALTER, D., THOMAS, A. and BEST, N. (2009). The BUGS project: Evolution, critique and future directions. *Stat. Med.* **28** 3049–3067. MR2750401 https://doi.org/10.1002/sim.3680

[76] LUNN, D. J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* **10** 325–337.

[77] MADSEN, A. L., LANG, M., KJÆRULFF, U. B. and JENSEN, F. (2003). The Hugin tool for learning Bayesian networks. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Lecture Notes in Computer Science* **2711** 594–605. Springer, New York. MR2050972 https://doi.org/10.1007/978-3-540-45062-7_49

[78] MAKOWSKI, D., BEN-SHACHAR, M. S. and LÜDECKE, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Sour. Softw.* **4** 1541.

[79] MANSINGHKA, V., SELSAM, D. and PEROV, Y. (2014). Venture: A higher-order probabilistic programming platform with programmable inference. Available at arXiv:1404.0099.

[80] MARTIN, G. M., FRAZIER, D. T. and ROBERT, C. P. (2022). Computing Bayes: From then 'til now'. Available at arXiv:2208.00646.

[81] MARTIN, A. D., QUINN, K. M. and PARK, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *J. Stat. Softw.* **42** 22.

[82] MARTINS, T. G., SIMPSON, D., LINDGREN, F. and RUE, H. (2013). Bayesian computing with INLA: New features. *Comput. Statist. Data Anal.* **67** 68–83. MR3079584 https://doi.org/10.1016/j.csda.2013.04.014

[83] MINKA, T., WINN, J. M., GUIVER, J. P., ZAYKOV, Y., FABIAN, D. and BRONSKILL, J. (2018). /Infer.NET 0.3. Microsoft Research Cambridge. Available at http://dotnet.github.io/infer.

[84] MONNAHAN, C. C. and KRISTENSEN, K. (2018). No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. *PLoS ONE* **13** e0197954. https://doi.org/10.1371/journal.pone.0197954

[85] MOREY, R. D. and ROUDER, J. N. (2022). BayesFactor: Computation of Bayes factors for common designs.

[86] MURPHY, K. (2001). The Bayes Net toolbox for Matlab. *Comput. Sci. Stat.* **33** 1024–1034.

[87] MURRAY, L. M., LUNDÉN, D., KUDLICKA, J., BROMAN, D. and SCHÖN, T. B. (2018). Delayed sampling and automatic Rao-blackwellization of probabilistic programs. In *Proceedings of the* 21*st International Conference on Artificial Intelligence and Statistics* (*AISTATS*).

[88] MURRAY, L. M. and SCHÖN, T. B. (2018). Automated learning with a probabilistic programming language: Birch. *Annu. Rev. Control* **46** 29–43. MR3907522 https://doi.org/10.1016/j.arcontrol.2018.10.013

[89] NEAL, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Department of Computer Science, Univ. Toronto.

[90] NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

[91] NEAL, R. M. MCMC using Hamiltonian dynamics. *Handb. Markov Chain Monte Carlo*.

[92] NUNES, M. A. and PRANGLE, D. (2015). abctools: An R package for tuning approximate Bayesian computation analyses. *R J.* **7** 189–205.

[93] PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N. et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**.

[94] PHAN, D., PRADHAN, N. and JANKOWIAK, M. (2019). Composable effects for flexible and accelerated probabilistic programming in NumPyro. Available at arXiv:1912.11554.

[95] PIIRONEN, J., PAASINIEMI, M., CATALINA, A., WEBER, F. and VEHTARI, A. (2023). projpred: Projection predictive feature selection.

[96] PLUMMER, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the* 3*rd International Workshop on Distributed Statistical Computing* **124** 1–10.

[97] PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.

[98] POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 https://doi.org/10.1080/01621459.2013.829001

[99] PRICE, L. F., DROVANDI, C. C., LEE, A. and NOTT, D. J. (2018). Bayesian synthetic likelihood. *J. Comput. Graph. Statist.* **27** 1–11. MR3788296 https://doi.org/10.1080/10618600.2017.1302882

[100] R CORE TEAM (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[101] RAINFORTH, T. W. G. (2017). Automating inference, learning, and design using probabilistic programming Ph.D. thesis Univ. Oxford.

[102] RIDELL, A. (2022). Python interface to Stan.

[103] RONQUIST, F., KUDLICKA, J., SENDEROV, V., BORGSTRÖM, J., LARTILLOT, N., LUNDÉN, D., MURRAY, L., SCHÖN, T. B. and BROMAN, D. (2021). Universal probabilistic programming offers a powerful approach to statistical phylogenetics. *Commun. Biol.* **4**.

[104] RSTUDIO TEAM (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA.

[105] RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

[106] RUE, H., RIEBLER, A., SØRBYE, S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: A review. *Annu. Rev. Stat. Appl.* **4** 395–421.

[107] SALVATIER, J., WIECKI, T. V. and FONNESBECK, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2** e55.

[108] SCHÄLTE, Y., KLINGER, E., ALAMOUDI, E. and HASENAUER, J. (2022). pyABC: Efficient and robust easy-to-use approximate Bayesian computation. Available at arXiv:2203.13043.

[109] SCUTARI, M. (2010). Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* **35**.

[110] SISSON, S. A., FAN, Y. and BEAUMONT, M. (2018). *Handbook of Approximate Bayesian Computation*. CRC Press, Boca Raton, FL.

[111] SPEAGLE, J. S. (2020). dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493** 3132–3158.

[112] SPIEGELHALTER, D., THOMAS, A., BEST, N. and GILKS, W. (1996). BUGS 0.5: Bayesian inference using Gibbs sampling manual (version II). In *MRC Biostatistics Unit* 1–59. Institute of Public Health, Cambridge, UK.

[113] SPIEGELHALTER, D. J., THOMAS, A., BEST, N. and LUNN, D. (2003). *WinBUGS Version* 1.4 *User Manual*, MRC Biostatistics Unit, Cambridge. Available at http://www.mrc-bsu.cam.ac.uk/bugs.

[114] SPIEGELHALTER, D., THOMAS, A., BEST, N. and LUNN, D. (2014). OpenBUGS user manual. Version 3.2.3.

[115] STAN DEVELOPMENT TEAM (2022). RStan: The R interface to Stan.

[116] STAN DEVELOPMENT TEAM (2022). A lightweight Python interface to CmdStan.

[117] STURTZ, S., LIGGES, U. and GELMAN, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *J. Stat. Softw.* **12** 1–16.

[118] SYED, S., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2022). Non-reversible parallel tempering: A scalable highly parallel MCMC scheme. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 321–350. MR4412989

[119] TALTS, S., BETANCOURT, M., SIMPSON, D., VEHTARI, A. and GELMAN, A. (2018). Validating Bayesian inference algorithms with simulation-based calibration. Available at arXiv:1804.06788.

[120] TAREK, M., XU, K., TRAPP, M., GE, H. and GHAHRAMANI, Z. (2020). DynamicPPL: Stan-like speed for dynamic probabilistic models. Available at arXiv:2002.02702.

[121] TAYLOR, S. J. and LETHAM, B. (2018). Forecasting at scale. *Amer. Statist.* **72** 37–45. MR3790566 https://doi.org/10.1080/00031305.2017.1380080

[122] TAYLOR, S. J. and LETHAM, B. (2021). prophet: Automatic Forecasting Procedure.

[123] TAYLOR, S. J. and LETHAM, B. (2022). Prophet: Automatic Forecasting Procedure.

[124] TEHRANI, N., ARORA, N. S., LI, Y. L., SHAH, K. D., NOURSI, D., TINGLEY, M., TORABI, N., LIPPERT, E., MEIJER, E. et al. (2020). Bean machine: A declarative probabilistic programming language for efficient programmable inference. In *International Conference on Probabilistic Graphical Models* 485–496. PMLR.

[125] TEJERO-CANTERO, A., BOELTS, J., DEISTLER, M., LUECKMANN, J.-M., DURKAN, C., GONÇALVES, P. J., GREENBERG, D. S. and MACKE, J. H. (2020). sbi: A toolkit for simulation-based inference. *J. Open Sour. Softw.* **5** 2505.

[126] THOMAS, N. (2020). R2OpenBUGS: Running OpenBUGS from R.

[127] THORNTON, K. R. (2009). Automating approximate Bayesian computation by local linear regression. *BMC Genet*. **10** 1–5.

[128] TOLPIN, D., VAN DE MEENT, J.-W., YANG, H. and WOOD, F. (2016). Design and implementation of probabilistic programming language anglican. In *Proceedings of the* 28*th Symposium on the Implementation and Application of Functional Programming Languages* 1–12.

[129] TRAN, D., HOFFMAN, M. W., MOORE, D., SUTER, C., VASUDEVAN, S. and RADUL, A. (2018). Simple, distributed, and accelerated probabilistic programming. *Adv. Neural Inf. Process. Syst*. **31**.

[130] VAN NIEKERK, J., BAKKA, H., RUE, H. and SCHENK, O. (2021). New frontiers in Bayesian modeling using the INLA package in R. *J. Stat. Softw.* **100** 1–28.

[131] VAN NIEKERK, J., KRAINSKI, E., RUSTAND, D. and RUE, H. (2023). A new avenue for Bayesian inference with INLA. *Comput. Statist. Data Anal.* **181** 107692. MR4540934 https://doi.org/10.1016/j.csda.2023.107692

[132] VAN NIEKERK, J. and RUE, H. (2021). Correcting the Laplace method with variational Bayes. Available at arXiv:2111.12945.

[133] VEHTARI, A., GABRY, J., MAGNUSSON, M., YAO, Y., BÜRKNER, P.-C., PAANANEN, T. and GELMAN, A. (2022). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.5.1.

[134] VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

[135] WICKHAM, H. (2016). *Ggplot*2: *Elegant Graphics for Data Analysis*. Springer, Berlin.

[136] WIGREN, A., RISULEO, R. S., MURRAY, L. M. and LINDSTEN, F. (2019). Parameter elimination in particle Gibbs sampling. *Advances in Neural Information Processing Systems* 32 (*NeurIPS* 2019).

[137] WOOD, S. N. (2017). *Generalized Additive Models*: *An Introduction with R*, Second ed. CRC Press, Boca Raton, FL.

Statistical Science
2024, Vol. 39, No. 1, 62–89
https://doi.org/10.1214/23-STS919

Emerging Directions in Bayesian Computation

Steven Winter, Trevor Campbell, Lizhen Lin, Sanvesh Srivastava and David B. Dunson

Abstract. Bayesian models are powerful tools for studying complex data, allowing the analyst to encode rich hierarchical dependencies and leverage prior information. Most importantly, they facilitate a complete characterization of uncertainty through the posterior distribution. Practical posterior computation is commonly performed via MCMC, which can be computationally infeasible for high-dimensional models with many observations. In this article, we discuss the potential to improve posterior computation using ideas from machine learning. Concrete directions are explored in vignettes on normalizing flows, statistical properties of variational approximations, Bayesian coresets and distributed Bayesian inference.

Key words and phrases: Coresets, federated learning, machine learning, normalizing flows, posterior computation, variational Bayes.

REFERENCES

[1] AGARWAL, P. K., HAR-PELED, S. and VARADARAJAN, K. R. (2005). Geometric approximation via coresets. In *Combinatorial and Computational Geometry*. *Math*. *Sci*. *Res*. *Inst*. *Publ*. **52** 1–30. Cambridge Univ. Press, Cambridge. MR2178310 https://doi.org/10.4171/PRIMS/172

[2] AGRAWAL, R. (1995). The continuum-armed bandit problem. *SIAM J*. *Control Optim*. **33** 1926–1951. MR1358102 https://doi.org/10.1137/S0363012992237273

[3] AHN, S., CHEN, Y. and WELLING, M. (2013). Distributed and adaptive darting Monte Carlo through regenerations. In *Artificial Intelligence and Statistics* 108–116. PMLR.

[4] AHN, S., SHAHBABA, B. and WELLING, M. (2014). Distributed stochastic gradient MCMC. *Int*. *Conf*. *Mach*. *Learn*. **32** 1044–1052.

[5] AICHER, C., PUTCHA, S., NEMETH, C., FEARNHEAD, P. and FOX, E. B. (2023). Stochastic gradient MCMC for nonlinear state space models. *Bayesian Anal*. 1–23.

[6] ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann*. *Statist*. **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

[7] ANDERSON, W. (2012). *Continuous-Time Markov Chains*: *An Applications Oriented Approach*. Springer, Berlin.

[8] ANDRICIOAEI, I., STRAUB, J. E. and VOTER, A. F. (2001). Smart darting Monte Carlo. *J*. *Chem*. *Phys*. **114** 6994–7000.

[9] ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat*. *Comput*. **18** 343–373. MR2461882 https://doi.org/10.1007/s11222-008-9110-y

[10] ARBEL, M., MATTHEWS, A. and DOUCET, A. (2021). Annealed flow transport Monte Carlo. In *Int*. *Conf*. *Mach*. *Learn*. **38** 318–330. PMLR.

[11] BACHEM, O., LUCIC, M. and KRAUSE, A. (2017). Practical coreset constructions for machine learning. Available at arXiv:1703.06476.

[12] BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton, FL.

[13] BARDENET, R., DOUCET, A. and HOLMES, C. (2017). On Markov chain Monte Carlo methods for tall data. *J*. *Mach*. *Learn*. *Res*. **18** 47. MR3670492

[14] BETANCOURT, M. (2015). The fundamental incompatibility of Hamiltonian Monte Carlo and data subsampling. *Int*. *Conf*. *Mach*. *Learn*. **37** 533–540.

[15] BLEI, D., GRIFFITHS, T., JORDAN, M. and TENENBAUM, J. (2003). Hierarchical topic models and the nested Chinese restaurant process. *Adv*. *Neural Inf*. *Process*. *Syst*. **16**.

[16] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J*. *Amer*. *Statist*. *Assoc*. **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

[17] BOYD, S., PARIKH, N., CHU, E., PELEATO, B., ECKSTEIN, J. et al. (2011). Distributed optimization and statistical learn-

Steven Winter is Ph.D. Student, Department of Statistical Science, Duke University, Durham, North Carolina 27710, USA (e-mail: steven.winter@duke.edu). Trevor Campbell is Associate Professor, Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4 (e-mail: trevor@stat.ubc.ca). Lizhen Lin is Robert and Sara Lumpkins Associate Professor, Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana 46556, USA (e-mail: lizhen.lin@nd.edu). Sanvesh Srivastava is Associate Professor, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, USA (e-mail: sanvesh-srivastava@uiowa.edu). David B. Dunson is Arts and Sciences Distinguished Professor, Departments of Statistical Science and Mathematics, Duke University, Durham, North Carolina 27710, USA (e-mail: dunson@duke.edu).

ing via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*.
[18] BRODERICK, T., BOYD, N., WIBISONO, A., WILSON, A. and JORDAN, M. (2013). Streaming variational Bayes. *Adv. Neural Inf. Process. Syst.* **26**.
[19] BROFOS, J., GABRIÉ, M., BRUBAKER, M. A. and LEDERMAN, R. R. (2022). Adaptation of the independent Metropolis-Hastings sampler with normalizing flow proposals. *Int. Conf. Artif. Intell. Stat.* **151** 5949–5986.
[20] BROSSE, N., DURMUS, A. and MOULINES, E. (2018). The promises and pitfalls of stochastic gradient Langevin dynamics. *Adv. Neural Inf. Process. Syst.* **31**.
[21] BUBECK, S., ELDAN, R. and LEHEC, J. (2018). Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* **59** 757–783. MR3802303 https://doi.org/10.1007/s00454-018-9992-1
[22] CAMPBELL, T. and BERONOV, B. (2019). Sparse variational inference: Bayesian coresets from scratch. *Adv. Neural Inf. Process. Syst.* **32**.
[23] CAMPBELL, T. and BRODERICK, T. (2018). Bayesian coreset construction via greedy iterative geodesic ascent. *Int. Conf. Mach. Learn.* **80** 698–706.
[24] CAMPBELL, T. and BRODERICK, T. (2019). Automated scalable Bayesian inference via Hilbert coresets. *J. Mach. Learn. Res.* **20** 15. MR3911422
[25] CAMUTO, A. and WILLETTS, M. (2022). Variational autoencoders: A harmonic perspective. In *Proceedings of the* 25*th International Conference on Artificial Intelligence and Statistics* **151** 4595–4611. PMLR.
[26] CAMUTO, A., WILLETTS, M., ROBERTS, S., HOLMES, C. and RAINFORTH, T. (2021). Towards a theoretical understanding of the robustness of variational autoencoders. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* **24** 3565–3573. PMLR.
[27] CARPENTER, B., GELMAN, A., HOFFMAN, M., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.
[28] CASELLA, G. and ROBERT, C. P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika* **83** 81–94. MR1399157 https://doi.org/10.1093/biomet/83.1.81
[29] CHAE, M., KIM, D., KIM, Y. and LIN, L. (2023). A likelihood approach to nonparametric estimation of a singular distribution using deep generative models. *J. Mach. Learn. Res.* **24** 77. MR4582499
[30] CHAN, R., POLLOCK, M., JOHANSEN, A. and ROBERTS, G. (2021). Divide-and-conquer Monte Carlo fusion. Available at arXiv:2110.07265.
[31] CHATTERJEE, S. and DIACONIS, P. (2018). The sample size required in importance sampling. *Ann. Appl. Probab.* **28** 1099–1135. MR3784496 https://doi.org/10.1214/17-AAP1326
[32] CHEN, C., DING, N., LI, C., ZHANG, Y. and CARIN, L. (2016). Stochastic gradient MCMC with stale gradients. *Adv. Neural Inf. Process. Syst.* **29**.
[33] CHEN, N. and CAMPBELL, T. (2023). Coreset Markov chain Monte Carlo. Available at arXiv:2310.17063.
[34] CHEN, N., XU, Z. and CAMPBELL, T. (2022). Bayesian inference via sparse Hamiltonian flows. *Adv. Neural Inf. Process. Syst.* **35**.
[35] CHEN, R. T., RUBANOVA, Y., BETTENCOURT, J. and DUVENAUD, D. K. (2018). Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.* **31**.
[36] CHEN, X., WEN, H. and LI, Y. (2021). Differentiable particle filters through conditional normalizing flow. In 2021 *IEEE* 24*th International Conference on Information Fusion* **24** 1–6.
[37] CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2018). Consistency of variational Bayes inference for estimation and model selection in mixtures. *Electron. J. Stat.* **12** 2995–3035. MR3855643 https://doi.org/10.1214/18-EJS1475
[38] DAI, H., POLLOCK, M. and ROBERTS, G. (2019). Monte Carlo fusion. *J. Appl. Probab.* **56** 174–191. MR3981152 https://doi.org/10.1017/jpr.2019.12
[39] DEISENROTH, M. and NG, J. W. (2015). Distributed Gaussian processes. In *Int. Conf. Mach. Learn.* **37** 1481–1490. PMLR.
[40] DE SOUZA, D. A., MESQUITA, D., KASKI, S. and ACERBI, L. (2022). Parallel MCMC without embarrassing failures. *Int. Conf. Artif. Intell. Stat.* **151** 1786–1804.
[41] DIACONIS, P. (1988). Sufficiency as statistical symmetry. *Proc. AMS Centen. Symp.* 15–26.
[42] DIENG, A., TRAN, D., RANGANATH, R., PAISLEY, J. and BLEI, D. (2017). Variational inference via χ-upper bound minimization. *Adv. Neural Inf. Process. Syst.* **30**.
[43] DINH, L., SOHL-DICKSTEIN, J. and BENGIO, S. (2016). Density estimation using real NVP. ArXiv preprint. Available at arXiv:1605.08803.
[44] DINH, V., BILGE, A., ZHANG, C. and MATSEN IV, F. A. (2017). Probabilistic path Hamiltonian Monte Carlo. *Int. Conf. Mach. Learn.* **70** 1009–1018.
[45] DOMKE, J., GARRIGOS, G. and GOWER, R. (2019). Provable convergence guarantees for black-box variational inference. *Adv. Neural Inf. Process. Syst.* **32**.
[46] DUNSON, D. B. and JOHNDROW, J. E. (2020). The Hastings algorithm at fifty. *Biometrika* **107** 1–23. MR4064137 https://doi.org/10.1093/biomet/asz066
[47] DURKAN, C., BEKASOV, A., MURRAY, I. and PAPAMAKARIOS, G. (2019). Neural spline flows. *Adv. Neural Inf. Process. Syst.* **32**.
[48] EL MEKKAOUI, K., MESQUITA, D., BLOMSTEDT, P. and KASKI, S. (2021). Federated stochastic gradient Langevin dynamics. *Uncertainty Artif. Intell.* **161** 1703–1712.
[49] FELDMAN, D. (2020). Introduction to Core-sets: An updated survey. Available at arXiv:2011.09384.
[50] FELDMAN, D., FIAT, A., KAPLAN, H. and NISSIM, K. (2009). Private coresets. In *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 361–370. ACM, New York. MR2780082
[51] FELDMAN, D. and LANGBERG, M. (2011). A unified framework for approximating and clustering data. In *STOC'*11*—Proceedings of the* 43*rd ACM Symposium on Theory of Computing* 569–578. ACM, New York. MR2932007 https://doi.org/10.1145/1993636.1993712
[52] FOSTER, D. J., SEKHARI, A. and SRIDHARAN, K. (2018). Uniform convergence of gradients for non-convex learning and optimization. *Adv. Neural Inf. Process. Syst.* **32**.
[53] GABRIÉ, M., ROTSKOFF, G. M. and VANDEN-EIJNDEN, E. (2022). Adaptive Monte Carlo augmented with normalizing flows. *Proc. Natl. Acad. Sci. USA* **119** e2109420119. MR4417577 https://doi.org/10.1073/pnas.2109420119
[54] GAO, C., ISAACSON, J. and KRAUSE, C. (2020). I-flow: High-dimensional integration and sampling with normalizing flows. *Mach. Learn.: Sci. Technol.* **1**.
[55] GE, H., XU, K. and GHAHRAMANI, Z. (2018). Turing: A language for flexible probabilistic inference. *Artif. Intell. Stat.* **84** 1682–1690.
[56] GEFFNER, T. and DOMKE, J. (2021). MCMC variational inference via uncorrected Hamiltonian annealing. *Adv. Neural Inf. Process. Syst.* **34**.
[57] GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

[58] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228
[59] GHOSAL, S., LEMBER, J. and VAN DER VAART, A. (2008). Nonparametric Bayesian model selection and averaging. *Electron. J. Stat.* **2** 63–89. MR2386086 https://doi.org/10.1214/07-EJS090
[60] GHOSH, S., DELLE FAVE, F. and YEDIDIA, J. (2016). Assumed density filtering methods for learning Bayesian neural networks. *Proc. AAAI Conf. Artif. Intell.* **30**.
[61] GIORDANO, R., BRODERICK, T. and JORDAN, M. I. (2018). Covariances, robustness, and variational Bayes. *J. Mach. Learn. Res.* **19** 51. MR3874159
[62] GONG, W., LI, Y. and HERNÁNDEZ-LOBATO, J. M. (2019). Meta-learning for stochastic gradient MCMC.
[63] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2020). Generative adversarial networks. *Commun. ACM* **63**.
[64] GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* **24** 561–578. MR3357395 https://doi.org/10.1080/10618600.2014.914442
[65] GREGOR, K., DANIHELKA, I., GRAVES, A., REZENDE, D. and WIERSTRA, D. (2015). DRAW: A recurrent neural network for image generation. In *Proceedings of the* 32*nd Int. Conf. Mach. Learn.* **37** 1462–1471. PMLR.
[66] GUHANIYOGI, R. and BANERJEE, S. (2018). Meta-Kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics* **60** 430–444. MR3878099 https://doi.org/10.1080/00401706.2018.1437474
[67] GUHANIYOGI, R., LI, C., SAVITSKY, T. and SRIVASTAVA, S. (2023). Distributed Bayesian inference in massive spatial data. *Statist. Sci.* **38** 262–284. MR4597336 https://doi.org/10.1214/22-sts868
[68] GUHANIYOGI, R., LI, C., SAVITSKY, T. D. and SRIVASTAVA, S. (2022). Distributed Bayesian varying coefficient modeling using a Gaussian process prior. *J. Mach. Learn. Res.* **23** 84. MR4576669
[69] GUI, J., SUN, Z., WEN, Y., TAO, D. and YE, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* **35** 3313–3332.
[70] GULRAJANI, I., AHMED, F., ARJOVSKY, M., DUMOULIN, V. and COURVILLE, A. C. (2017). Improved training of Wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **30**.
[71] GUO, F., WANG, X., FAN, K., BRODERICK, T. and DUNSON, D. (2016). Boosting variational inference **29**.
[72] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737
[73] HALL, P., PHAM, T., WAND, M. P. and WANG, S. S. J. (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *Ann. Statist.* **39** 2502–2532. MR2906876 https://doi.org/10.1214/11-AOS908
[74] HARSHVARDHAN, G. M., GOURISARIA, M. K., PANDEY, M. and RAUTARAY, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Comput. Sci. Rev.* **38** 100285. MR4131875 https://doi.org/10.1016/j.cosrev.2020.100285
[75] HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926
[76] HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779
[77] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7
[78] HUGGINS, J., CAMPBELL, T. and BRODERICK, T. (2016). Coresets for scalable Bayesian logistic regression. *Adv. Neural Inf. Process. Syst.* **29**.
[79] JANKOWIAK, M. and PHAN, D. (2022). Surrogate likelihoods for variational annealed importance sampling. *Int. Conf. Mach. Learn.* **162** 9881–9901.
[80] JEONG, K., CHAE, M. and KIM, Y. (2023). Online learning for the Dirichlet process mixture model via weakly conjugate approximation. *Comput. Statist. Data Anal.* **179** 107626. MR4495773 https://doi.org/10.1016/j.csda.2022.107626
[81] JOHNDROW, J., PILLAI, N. and SMITH, A. (2020). No free lunch for approximate MCMC. Available at arXiv:2010.12514.
[82] JOLICOEUR-MARTINEAU, A. (2019). The relativistic discriminator: A key element missing from standard GAN.
[83] JORDAN, M., GHAHRAMANI, Z., JAAKKOLA, T. and SAUL, L. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.
[84] JORDAN, M. I., LEE, J. D. and YANG, Y. (2019). Communication-efficient distributed statistical inference. *J. Amer. Statist. Assoc.* **114** 668–681. MR3963171 https://doi.org/10.1080/01621459.2018.1429274
[85] KATSEVICH, A. and RIGOLLET, P. (2023). On the approximation accuracy of Gaussian variational Inference.
[86] KIDD, B., WANG, K., XU, Y. and NI, Y. (2022). Federated learning for sparse Bayesian models with applications to electronic health records and genomics. *Pac. Symp. BioComputing* **28** 484–495.
[87] KIM, K., WU, K., OH, J., MA, Y. and GARDNER, J. (2023). On the convergence of black-box variational inference. *Adv. Neural Inf. Process. Syst.* **37**.
[88] KIM, Y., CHAE, M., JEONG, K., KANG, B. and CHUNG, H. (2016). An online Gibbs sampler algorithm for hierarchical Dirichlet processes prior. *Mach. Learn. Knowl. Discov. Databases* 509–523.
[89] KINGMA, D. P. and BA, J. (2017). Adam: A method for stochastic optimization.
[90] KINGMA, D. P., SALIMANS, T., JOZEFOWICZ, R., CHEN, X., SUTSKEVER, I. and WELLING, M. (2016). Improved variational inference with inverse autoregressive flow. *Adv. Neural Inf. Process. Syst.* **30**.
[91] KINGMA, D. P. and WELLING, M. (2014). Auto-encoding variational Bayes. In 2*nd International Conference on Learning Representations*, *Banff*, *AB*, *Canada*, *April* 14-16, 2014, *Conference Track Proceedings*.
[92] KINGMA, D. P. and WELLING, M. (2019). An introduction to variational autoencoders. *Found. Trends Mach. Learn.* **12** 307–392.
[93] KOBYZEV, I., PRINCE, S. J. and BRUBAKER, M. A. (2021). Normalizing flows: An introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 3964–3979.
[94] KORATTIKARA, A., CHEN, Y. and WELLING, M. (2014). Austerity in MCMC land: Cutting the Metropolis-Hastings budget. *Int. Conf. Mach. Learn.* **32** 181–189.
[95] LAN, S., STREETS, J. and SHAHBABA, B. (2014). Wormhole Hamiltonian Monte Carlo. *Proc. AAAI Conf. Artif. Intell.* 1953–1959.
[96] LAURITZEN, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. *Lecture Notes in Statistics* **49**. Springer, New York. MR0971253 https://doi.org/10.1007/978-1-4612-1023-8

[97] LE, T. A., BAYDIN, A. G. and WOOD, F. (2017). Inference compilation and universal probabilistic programming. *Artif. Intell. Stat.* **54** 1338–1348.

[98] LEVY, D., HOFFMAN, M. D. and SOHL-DICKSTEIN, J. (2018). Generalizing Hamiltonian Monte Carlo with neural networks. *Int. Conf. Learn. Represent.*

[99] LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. MR3694589 https://doi.org/10.1093/biomet/asx033

[100] LI, J., LUO, X. and QIAO, M. (2020). On generalization error bounds of noisy gradient methods for non-convex learning.

[101] LI, W., AHN, S. and WELLING, M. (2016). Scalable MCMC for mixed membership stochastic blockmodels. In *Artificial Intelligence and Statistics* **51** 723–731. PMLR.

[102] LI, Y. and TURNER, R. (2016). Rényi divergence variational inference. *Adv. Neural Inf. Process. Syst.* **30**.

[103] LIANG, F., MAHONEY, M. and HODGKINSON, L. (2022). Fat–tailed variational inference with anisotropic tail adaptive flows. *Int. Conf. Mach. Learn.* **162** 13257–13270.

[104] LIN, D. (2013). Online learning of nonparametric mixture models via sequential variational approximation. *Adv. Neural Inf. Process. Syst.* **26**.

[105] LOH, P.-L. and WAINWRIGHT, M. J. (2013). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Adv. Neural Inf. Process. Syst.* **26**.

[106] LOO, N., SWAROOP, S. and TURNER, R. E. (2021). Generalized variational continual learning. In *International Conference on Learning Representations*.

[107] LU, X., PERRONE, V., HASENCLEVER, L., TEH, Y. W. and VOLLMER, S. (2017). Relativistic Monte Carlo. *Artif. Intell. Stat.* **54** 1236–1245.

[108] MA, Y.-A., CHEN, T. and FOX, E. (2015). A complete recipe for stochastic gradient MCMC. *Adv. Neural Inf. Process. Syst.* **28**.

[109] MA, Y.-A., FOTI, N. J. and FOX, E. B. (2017). Stochastic gradient MCMC methods for hidden Markov models. In *International Conference on Machine Learning* 2265–2274. PMLR.

[110] MACLAURIN, D. and ADAMS, R. (2014). Firefly Monte Carlo: Exact MCMC with subsets of data. *Conf. Uncertain. Artif. Intell.* **30** 543–552.

[111] MAHONEY, M. (2011). Randomized algorithms for matrices and data. *Found. Trends Mach. Learn.* **3** 123–224.

[112] MANGOUBI, O., PILLAI, N. S. and SMITH, A. (2018). Does Hamiltonian Monte Carlo mix faster than a random walk on multimodal densities? ArXiv preprint. Available at arXiv:1808.03230.

[113] MANOUSAKAS, D., RITTER, H. and KARALETSOS, T. (2022). Black-box coreset variational inference. *Adv. Neural Inf. Process. Syst.* **36**.

[114] MANOUSAKAS, D., XU, Z., MASCOLO, C. and CAMPBELL, T. (2020). Bayesian pseudocoresets. *Adv. Neural Inf. Process. Syst.* **33**.

[115] MAO, X., LI, Q., XIE, H., LAU, R. Y., WANG, Z. and PAUL SMOLLEY, S. (2017). Least-squares generative adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*.

[116] MATHIEU, E. and NICKEL, M. (2020). Riemannian continuous normalizing flows. *Adv. Neural Inf. Process. Syst.* **33**.

[117] MATTHEWS, A., ARBEL, M., REZENDE, D. J. and DOUCET, A. (2022). Continual repeated annealed flow transport Monte Carlo. In *Int. Conf. Mach. Learn.* **162** 15196–15219. PMLR.

[118] MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

[119] MESQUITA, D., BLOMSTEDT, P. and KASKI, S. (2020). Embarrassingly parallel MCMC using deep invertible transformations. *Uncertain. Artif. Intell.* **115** 1244–1252.

[120] MILLER, A. C., FOTI, N. J. and ADAMS, R. P. (2017). Variational boosting: Iteratively refining posterior approximations. *Proc.* 34*th Int. Conf. Mach. Learn.* **70** 2420–2429.

[121] MINKA, T. and LAFFERTY, J. (2002). Expectation-propagation for the generative aspect model. *Proc. Eighteen. Conf. Uncertain. Artif. Intell.* **18** 352–359.

[122] MINSKER, S. (2019). Distributed statistical estimation and rates of convergence in normal approximation. *Electron. J. Stat.* **13** 5213–5252. MR4043072 https://doi.org/10.1214/19-EJS1647

[123] MINSKER, S., SRIVASTAVA, S., LIN, L. and DUNSON, D. (2017). Scalable and robust Bayesian inference via the median posterior. *Int. Conf. Mach. Learn.* **32** 1656–1664.

[124] MIRZA, M. and OSINDERO, S. (2014). Conditional generative adversarial nets. ArXiv preprint. Available at arXiv:1411.1784.

[125] MIYATO, T., KATAOKA, T., KOYAMA, M. and YOSHIDA, Y. (2018). Spectral normalization for generative adversarial networks.

[126] MOHASEL AFSHAR, H. and DOMKE, J. (2015). Reflection, refraction, and Hamiltonian Monte Carlo. *Adv. Neural Inf. Process. Syst.* **28**.

[127] MÜLLER, T., MCWILLIAMS, B., ROUSSELLE, F., GROSS, M. and NOVÁK, J. (2019). Neural importance sampling. *ACM Trans. Graph.* **38** 1–19.

[128] NAGAPETYAN, T., DUNCAN, A., HASENCLEVER, L., VOLLMER, S., SZPRUCH, L. and ZYGALAKIS, K. (2017). The true cost of stochastic gradient Langevin dynamics. Available at arXiv:1706.02692.

[129] NAIK, C., ROUSSEAU, J. and CAMPBELL, T. (2022). Fast Bayesian coresets via subsampling and quasi-Newton refinement. *Adv. Neural Inf. Process. Syst.* **35**.

[130] NEAL, R. M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132 https://doi.org/10.1023/A:1008923215028

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

[132] NEISWANGER, W., WANG, C. and XING, E. P. (2014). Asymptotically exact, embarrassingly parallel MCMC. *Proc. Thirtieth Conf. Uncertain. Artif. Intell.*

[133] NEMETH, C. and FEARNHEAD, P. (2021). Stochastic gradient Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **116** 433–450. MR4227705 https://doi.org/10.1080/01621459.2020.1847120

[134] NEMETH, C. and SHERLOCK, C. (2018). Merging MCMC subposteriors through Gaussian-process approximations. *Bayesian Anal.* **13** 507–530. MR3780433 https://doi.org/10.1214/17-BA1063

[135] NGUYEN, C. V., LI, Y., BUI, T. D. and TURNER, R. E. (2018). Variational continual learning. *Int. Conf. Learn. Represent.*.

[136] NI, Y., JI, Y. and MÜLLER, P. (2020). Consensus Monte Carlo for random subsets using shared anchors. *J. Comput. Graph. Statist.* **29** 703–714. MR4191237 https://doi.org/10.1080/10618600.2020.1737085

[137] NIEMAN, D., SZABO, B. and VAN ZANTEN, H. (2023). Uncertainty quantification for sparse spectral variational approximations in Gaussian process regression. *Electron. J. Stat.* **17** 2250–2288. MR4649981 https://doi.org/10.1214/23-ejs2155

[138] NIKOLAKAKIS, K. E., HADDADPOUR, F., KARBASI, A. and KALOGERIAS, D. S. (2022). Beyond Lipschitz: Sharp generalization and excess risk bounds for full-batch GD. ArXiv preprint. Available at arXiv:2204.12446.

[139] NING, B. (2021). Spike and slab Bayesian sparse principal component analysis. ArXiv preprint. Available at arXiv:2102.00305.
[140] NISHIMURA, A. and DUNSON, D. (2016). Geometrically tempered Hamiltonian Monte Carlo. ArXiv preprint. Available at arXiv:1604.00872.
[141] NISHIMURA, A., DUNSON, D. and LU, J. (2017). Discontinuous Hamiltonian Monte Carlo for sampling discrete parameters. ArXiv preprint. Available at arXiv:1705.08510.
[142] OHN, I. and LIN, L. (2021). Adaptive variational Bayes: Optimality, computation and applications. ArXiv preprint. Available at arXiv:2109.03204.
[143] ONEILL, J. (2020). An overview of neural network compression. Available at arXiv:2006.03669.
[144] ORBANZ, P. (2017). Subsampling large graphs and invariance in networks. Available at arXiv:1710.04217.
[145] ORBANZ, P. and ROY, D. (2015). Bayesian models of graphs, arrays, and other exchangeable structures. *IEEE Trans. Pattern Anal. Mach. Intell*. **37** 437–461.
[146] OU, R., SEN, D. and DUNSON, D. (2021). Scalable Bayesian inference for time series via divide-and-conquer. ArXiv preprint. Available at arXiv:2106.11043.
[147] PAISLEY, J., BLEI, D. M. and JORDAN, M. I. (2012). Variational Bayesian inference with stochastic search. In *Proceedings of the* 29*th International Coference on Int. Conf. Mach. Learn*. 1363–1370. Omnipress.
[148] PAKMAN, A. and PANINSKI, L. (2013). Auxiliary-variable exact Hamiltonian Monte Carlo samplers for binary distributions. *Adv. Neural Inf. Process. Syst*. **26**.
[149] PAPAMAKARIOS, G., NALISNICK, E., REZENDE, D. J., MOHAMED, S. and LAKSHMINARAYANAN, B. (2021). Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res*. **22** 57. MR4253750
[150] PARIKH, N., BOYD, S. et al. (2014). Proximal algorithms. *Found. Trends Optim*. **1**.
[151] PASARICA, C. and GELMAN, A. (2010). Adaptively scaling the Metropolis algorithm using expected squared jumped distance. *Statist. Sinica* **20** 343–364. MR2640698
[152] PATI, D., BHATTACHARYA, A. and YANG, Y. (2018). On statistical optimality of variational Bayes. *Proc.* 21*st Int. Conf. Artif. Intell. Stat*. **84** 1579–1588.
[153] PATTERSON, S. and TEH, Y. W. (2013). Stochastic gradient Riemannian Langevin dynamics on the probability simplex. *Adv. Neural Inf. Process. Syst*. **26**.
[154] PETZKA, H., FISCHER, A. and LUKOVNICOV, D. (2017). On the regularization of Wasserstein GANs. ArXiv preprint. Available at arXiv:1709.08894.
[155] PHILLIPS, J. M. and TAI, W. M. (2020). Near-optimal coresets of kernel density estimates. *Discrete Comput. Geom*. **63** 867–887. MR4110524 https://doi.org/10.1007/s00454-019-00134-6
[156] PILLAI, N. S., STUART, A. M. and THIÉRY, A. H. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22** 2320–2356. MR3024970 https://doi.org/10.1214/11-AAP828
[157] PLASSIER, V., VONO, M., DURMUS, A. and MOULINES, E. (2021). DG-LMC: A turn-key and scalable synchronous distributed MCMC algorithm via Langevin Monte Carlo within Gibbs. *Int. Conf. Mach. Learn*. **139** 8577–8587.
[158] PLUMMER, M. (2015). Cuts in Bayesian graphical models. *Stat. Comput*. **25** 37–43. MR3304902 https://doi.org/10.1007/s11222-014-9503-z
[159] POMPE, E., HOLMES, C. and ŁATUSZYŃSKI, K. (2020). A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist*. **48** 2930–2952. MR4152629 https://doi.org/10.1214/19-AOS1916
[160] QUIROZ, M., KOHN, R., DANG, K.-D., VILLANI, M. and TRAN, M.-N. (2018). Subsampling MCMC—an introduction for the survey statistician. *Sankhya A* **80** S33–S69. MR3968357 https://doi.org/10.1007/s13171-018-0153-7
[161] QUIROZ, M., KOHN, R., VILLANI, M. and TRAN, M.-N. (2019). Speeding up MCMC by efficient data subsampling. *J. Amer. Statist. Assoc*. **114** 831–843. MR3963184 https://doi.org/10.1080/01621459.2018.1448827
[162] RANGANATH, R., GERRISH, S. and BLEI, D. (2014). Black box variational inference. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* **33** 814–822. PMLR.
[163] RASMUSSEN, C. and GHAHRAMANI, Z. (2001). Infinite mixtures of Gaussian process experts. *Adv. Neural Inf. Process. Syst*. **14**.
[164] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435
[165] RAY, K. and SZABÓ, B. (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc*. **117** 1270–1281. MR4480711 https://doi.org/10.1080/01621459.2020.1847121
[166] RAY, K., SZABÓ, B. and CLARA, G. (2020). Spike and slab variational Bayes for high dimensional logistic regression. *Proc.* 34*th Int. Conf. Neural Inf. Proc. Syst*. **34**.
[167] RENDELL, L. J., JOHANSEN, A. M., LEE, A. and WHITELEY, N. (2021). Global consensus Monte Carlo. *J. Comput. Graph. Statist*. **30** 249–259. MR4270501 https://doi.org/10.1080/10618600.2020.1811105
[168] REZENDE, D. and MOHAMED, S. (2015). Variational inference with normalizing flows. *Int. Conf. Mach. Learn.*.
[169] REZENDE, D. J., PAPAMAKARIOS, G., RACANIERE, S., ALBERGO, M., KANWAR, G., SHANAHAN, P. and CRANMER, K. (2020). Normalizing flows on tori and spheres. *Int. Conf. Mach. Learn*. **119** 8083–8092.
[170] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123
[171] ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist*. **18** 349–367. MR2749836 https://doi.org/10.1198/jcgs.2009.06134
[172] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418
[173] ROSS, S. M. (2002). *Simulation*. Academic Press, Amsterdam.
[174] ROSSKY, P. J., DOLL, J. D. and FRIEDMAN, H. L. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys*. **69** 4628–4633.
[175] ROTH, K., LUCCHI, A., NOWOZIN, S. and HOFMANN, T. (2017). Stabilizing training of generative adversarial networks through regularization. *Adv. Neural Inf. Process. Syst*. **31**.
[176] SALIMANS, T., KINGMA, D. and WELLING, M. (2015). Markov chain Monte Carlo and variational inference: Bridging the gap. *Int. Conf. Mach. Learn*. **37** 1218–1226.
[177] SAVITSKY, T. D. and SRIVASTAVA, S. (2018). Scalable Bayes under informative sampling. *Scand. J. Stat*. **45** 534–556. MR3858945 https://doi.org/10.1111/sjos.12312
[178] SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *Int. J. Manag. Sci. Eng. Manag*. **11** 78–88.

[179] SEN, D., SACHS, M., LU, J. and DUNSON, D. B. (2020). Efficient posterior sampling for high-dimensional imbalanced logistic regression. *Biometrika* **107** 1005–1012. MR4186502 https://doi.org/10.1093/biomet/asaa035

[180] SHANG, Z., HAO, B. and CHENG, G. (2019). Nonparametric Bayesian aggregation for massive data. *J. Mach. Learn. Res.* **20** 140. MR4030154

[181] SHUN, Z. and MCCULLAGH, P. (2018). Laplace approximation of high dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** 749–760.

[182] SHYAMALKUMAR, N. D. and SRIVASTAVA, S. (2022). An algorithm for distributed Bayesian inference. *Stat* **11** e432. MR4449228 https://doi.org/10.1002/sta4.432

[183] SNELSON, E. and GHAHRAMANI, Z. (2005). Sparse Gaussian processes using pseudo-inputs. *Adv. Neural Inf. Process. Syst.* **18**.

[184] SONG, J., ZHAO, S. and ERMON, S. (2017). A-NICE-MC: Adversarial training for MCMC. *Adv. Neural Inf. Process. Syst.* **30**.

[185] SRIVASTAVA, S., CEVHER, V., DINH, Q. and DUNSON, D. (2015). WASP: Scalable Bayes via barycenters of subset posteriors. *Artif. Intell. Stat.* **38** 912–920.

[186] SZABÓ, B. and VAN ZANTEN, H. (2019). An asymptotic analysis of distributed nonparametric methods. *J. Mach. Learn. Res.* **20** 87. MR3960941

[187] SZABÓ, B. and VAN ZANTEN, H. (2022). Distributed function estimation: Adaptation using minimal communication. *Math. Stat. Learn.* **5** 159–199. MR4526299 https://doi.org/10.4171/msl/33

[188] TANG, R. and YANG, Y. (2021). On empirical Bayes variational autoencoder: An excess risk bound. In *Proceedings of Thirty Fourth Conference on Learning Theory* **134** 4068–4125. PMLR.

[189] TANG, R. and YANG, Y. (2023). Minimax rate of distribution estimation on unknown submanifolds under adversarial losses. *Ann. Statist.* **51** 1282–1308. MR4630949 https://doi.org/10.1214/23-aos2291

[190] R CORE TEAM (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[191] THIN, A., KOTELEVSKII, N., DOUCET, A., DURMUS, A., MOULINES, E. and PANOV, M. (2021). Monte Carlo variational auto-encoders. *Int. Conf. Mach. Learn.* **139** 10247–10257.

[192] TOKDAR, S. T. and KASS, R. E. (2010). Importance sampling: A review. *Wiley Interdiscip. Rev.: Comput. Stat.* **2**.

[193] TRAN, D., RANGANATH, R. and BLEI, D. (2017). Hierarchical implicit models and likelihood-free variational inference. *Adv. Neural Inf. Process. Syst.* **31**.

[194] TRAN, D., VAFA, K., AGRAWAL, K., DINH, L. and POOLE, B. (2019). Discrete flows: Invertible generative models of discrete data. *Adv. Neural Inf. Process. Syst.* **32**.

[195] TRESP, V. (2000). A Bayesian committee machine. *Neural Comput.* **12** 2719–2741. https://doi.org/10.1162/089976600300014908

[196] UEHARA, M., SATO, I., SUZUKI, M., NAKAYAMA, K. and MATSUO, Y. (2016). Generative adversarial nets from a density ratio estimation perspective. ArXiv preprint. Available at arXiv:1610.02920.

[197] VAKILI, S., SCARLETT, J., SHAN SHIU, D. and BERNACCHIA, A. (2022). Improved convergence rates for sparse approximation methods in kernel-based learning. In *Int. Conf. Mach. Learn.* **162** 21960–21983. PMLR.

[198] VEHTARI, A., GELMAN, A., SIVULA, T., JYLÄNKI, P., TRAN, D., SAHAI, S., BLOMSTEDT, P., CUNNINGHAM, J. P., SCHIMINOVICH, D. et al. (2020). Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *J. Mach. Learn. Res.* **21** 17. MR4071200

[199] VIHOLA, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Stat. Comput.* **22** 997–1008. MR2950080 https://doi.org/10.1007/s11222-011-9269-5

[200] VONO, M., DOBIGEON, N. and CHAINAIS, P. (2019). Split-and-augmented Gibbs sampler—application to large-scale inference problems. *IEEE Trans. Signal Process.* **67** 1648–1661. MR3938771 https://doi.org/10.1109/TSP.2019.2894825

[201] VONO, M., DOBIGEON, N. and CHAINAIS, P. (2021). Asymptotically exact data augmentation: Models, properties, and algorithms. *J. Comput. Graph. Statist.* **30** 335–348. MR4270508 https://doi.org/10.1080/10618600.2020.1826954

[202] VONO, M., PAULIN, D. and DOUCET, A. (2022). Efficient MCMC sampling with dimension-free convergence rate using ADMM-type splitting. *J. Mach. Learn. Res.* **23** 25. MR4420750

[203] VYNER, C., NEMETH, C. and SHERLOCK, C. (2023). SwISS: A scalable Markov chain Monte Carlo divide-and-conquer strategy. *Stat* **12** e523. MR4538659 https://doi.org/10.1002/sta4.523

[204] WAINWRIGHT, M. and JORDAN, M. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

[205] WALKER, J., DOERSCH, C., GUPTA, A. K. and HEBERT, M. (2016). An uncertain future: Forecasting from static images using variational autoencoders. In *European Conference on Computer Vision* **9911** 835–851.

[206] WAN, N., LI, D. and HOVAKIMYAN, N. (2020). f-divergence variational inference. In *Advances in Neural Information Processing Systems* **33**.

[207] WANG, C. and SRIVASTAVA, S. (2023). Divide-and-conquer Bayesian inference in hidden Markov models. *Electron. J. Stat.* **17** 895–947. MR4563529 https://doi.org/10.1214/23-ejs2118

[208] WANG, H. and MA, Y. (2021). Optimal subsampling for quantile regression in big data. *Biometrika* **108** 99–112. MR4226192 https://doi.org/10.1093/biomet/asaa043

[209] WANG, H., YANG, M. and STUFKEN, J. (2019). Information-based optimal subdata selection for big data linear regression. *J. Amer. Statist. Assoc.* **114** 393–405. MR3941263 https://doi.org/10.1080/01621459.2017.1408468

[210] WANG, H., ZHU, R. and MA, P. (2018). Optimal subsampling for large sample logistic regression. *J. Amer. Statist. Assoc.* **113** 829–844. MR3832230 https://doi.org/10.1080/01621459.2017.1292914

[211] WANG, T., ZHU, J.-Y., TORRALBA, A. and EFROS, A. (2018). Dataset distillation. Available at arXiv:1811.10959.

[212] WANG, W., SUN, Y. and HALGAMUGE, S. (2018). Improving MMD-GAN training with repulsive loss function. ArXiv preprint. Available at arXiv:1812.09916.

[213] WANG, X. and DUNSON, D. B. (2013). Parallelizing MCMC via Weierstrass sampler. ArXiv preprint. Available at arXiv:1312.4605.

[214] WANG, X., GUO, F., HELLER, K. A. and DUNSON, D. B. (2015). Parallelizing MCMC with random partition trees. *Adv. Neural Inf. Process. Syst.* **28**.

[215] WANG, Y., AUDIBERT, J.-Y. and MUNOS, R. (2008). Algorithms for infinitely many-armed bandits. *Adv. Neural Inf. Process. Syst.* **21**.

[216] WANG, Y. and BLEI, D. M. (2019). Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc.* **114** 1147–1161. MR4011769 https://doi.org/10.1080/01621459.2018.1473776

[217] WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. *Int. Conf. Mach. Learn.*.

[218] WILLIAMS, C. and SEEGER, M. (2001). Using the Nyström method to speed up kernel machines. *Adv. Neural Inf. Process. Syst.* **13**.

[219] WINN, J. and BISHOP, C. M. (2005). Variational message passing. *J. Mach. Learn. Res.* **6** 661–694. MR2249835

[220] WU, C. and ROBERT, C. P. (2017). Average of recentered parallel MCMC for big data. ArXiv preprint. Available at arXiv:1706.04780.

[221] XU, Z., CHEN, N. and CAMPBELL, T. (2023). MixFlows: Principled variational inference via mixed flows. In *Int. Conf. Mach. Learn.* **202** 38342–38376. PMLR.

[222] XUE, J. and LIANG, F. (2019). Double-parallel Monte Carlo for Bayesian analysis of big data. *Stat. Comput.* **29** 23–32. MR3905537 https://doi.org/10.1007/s11222-017-9791-1

[223] YANG, Y. and MARTIN, R. (2020). Variational approximations of empirical Bayes posteriors in high-dimensional linear models. ArXiv preprint. Available at arXiv:2007.15930.

[224] YANG, Y., PATI, D. and BHATTACHARYA, A. (2020). α-variational inference with statistical guarantees. *Ann. Statist.* **48** 886–905. MR4102680 https://doi.org/10.1214/19-AOS1827

[225] YIN, M. and ZHOU, M. (2018). Semi-implicit variational inference. In *Int. Conf. Mach. Learn.* **80** 5660–5669. PMLR.

[226] YU, J., WANG, H., AI, M. and ZHANG, H. (2022). Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data. *J. Amer. Statist. Assoc.* **117** 265–276. MR4399084 https://doi.org/10.1080/01621459.2020.1773832

[227] ZHANG, F. and GAO, C. (2020). Convergence rates of variational posterior distributions. *Ann. Statist.* **48** 2180–2207. MR4134791 https://doi.org/10.1214/19-AOS1883

[228] ZHANG, G., HSU, K., LI, J., FINN, C. and GROSSE, R. (2021). Differentiable annealed importance sampling and the perils of gradient noise. *Adv. Neural Inf. Process. Syst.* **34**.

[229] ZHANG, J., KHANNA, R., KYRILLIDIS, A. and KOYEJO, O. (2021). Bayesian coresets: Revisiting the nonconvex optimization perspective. *Artif. Intell. Stat.* **130** 2782–2790.

[230] ZHANG, M. M. and WILLIAMSON, S. A. (2019). Embarrassingly parallel inference for Gaussian processes. *J. Mach. Learn. Res.* **20** 169. MR4048980

[231] ZHOU, J., KHARE, K. and SRIVASTAVA, S. (2023). Asynchronous and distributed data augmentation for massive data settings. *J. Comput. Graph. Statist.* **32** 895–907. MR4641467 https://doi.org/10.1080/10618600.2022.2130928

[232] ZHU, Y. and LAFFERTY, J. (2018). Distributed nonparametric regression under communication constraints. In *Int. Conf. Mach. Learn.* **35** 6009–6017. PMLR.

[233] ZIEGLER, Z. and RUSH, A. (2019). Latent normalizing flows for discrete sequences. *Int. Conf. Mach. Learn.* **97** 7673–7682.

Statistical Science
2024, Vol. 39, No. 1, 90–99
https://doi.org/10.1214/23-STS908

Diffusion Schrödinger Bridges for Bayesian Computation

Jeremy Heng, Valentin De Bortoli and Arnaud Doucet

Abstract. Denoising diffusion models are a novel class of generative models that have recently become extremely popular in machine learning. In this paper, we describe how such ideas can also be used to sample from posterior distributions and, more generally, any target distribution whose density is known up to a normalizing constant. The key idea is to consider a forward "noising" diffusion initialized at the target distribution, which "transports" this latter to a normal distribution for long diffusion times. The time reversal of this process, the "denoising" diffusion, thus "transports" the normal distribution to the target distribution and can be approximated so as to sample from the target. To accelerate simulation, we show how one can introduce and approximate a Schrödinger bridge between these two distributions, that is, a diffusion which transports the normal to the target in finite time.

Key words and phrases: Optimal transport, Schrödinger bridge, score matching, stochastic differential equation, time reversal.

REFERENCES

ALBERGO, M. S. and VANDEN-EIJNDEN, E. (2023). Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

ANDERSON, B. D. O. (1982). Reverse-time diffusion equation models. *Stochastic Process*. *Appl*. **12** 313–326. MR0656280 https://doi.org/10.1016/0304-4149(82)90051-5

AUSTIN, J., JOHNSON, D. D., HO, J., TARLOW, D. and VAN DEN BERG, R. (2021). Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, *December* 6–14, 2021, *Virtual* (M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang and J. Wortman Vaughan, eds.) 17981–17993.

BARR, A., GISPEN, W. and LAMACRAFT, A. (2020). Quantum ground states from reinforcement learning. In *Mathematical and Scientific Machine Learning*.

BEAUMONT, M. A. (2019). Approximate Bayesian computation. *Annu*. *Rev*. *Stat*. *Appl*. **6** 379–403. MR3939526 https://doi.org/10.1146/annurev-statistics-030718-105212

BENTON, J., DE BORTOLI, V., DOUCET, A. and DELIGIANNIDIS, G. (2023). Linear convergence bounds for diffusion models via stochastic localization. arXiv preprint. Available at arXiv:2308.03686.

BENTON, J., SHI, Y., DE BORTOLI, V., DELIGIANNIDIS, G. and DOUCET, A. (2022). From denoising diffusions to denoising Markov models. arXiv preprint. Available at arXiv:2211.03595.

CAMPBELL, A., BENTON, J., DE BORTOLI, V., RAINFORTH, T., DELIGIANNIDIS, G. and DOUCET, A. (2022). A continuous time framework for discrete denoising models. In *Advances in Neural Information Processing Systems* 35: *Annual Conference on Neural Information Processing Systems* 2022, *NeurIPS* 2022 (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.).

CHEN, S., CHEWI, S., LI, J., LI, Y., SALIM, A. and ZHANG, A. (2023). Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

CHEN, T., LIU, G. and THEODOROU, E. A. (2022). Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *The Tenth International Conference on Learning Representations*, *ICLR* 2022, *Virtual Event*, *April* 25–29, 2022. OpenReview.net.

CHEN, T., RUBANOVA, Y., BETTENCOURT, J. and DUVENAUD, D. (2018). Neural ordinary differential equations. In *Advances in Neural Information Processing Systems* 31: *Annual Conference on Neural Information Processing Systems* 2018, *NeurIPS* 2018, *December* 3–8, 2018, *Montréal*, *Canada* (S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) 6572–6583.

CHEN, Y., GEORGIOU, T. and PAVON, M. (2016). Entropic and displacement interpolation: A computational approach using the Hilbert metric. *SIAM J*. *Appl*. *Math*. **76** 2375–2396. MR3579702 https://doi.org/10.1137/16M1061382

CHUNG, H., KIM, J., MCCANN, M. T., KLASKY, M. L. and YE, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

Jeremy Heng is Assistant Professor, ESSEC Business School, Singapore 139408, Singapore (e-mail: heng@essec.edu). Valentin De Bortoli is Research Scientist, Center for Sciences of Data, ENS Ulm, Paris, France (e-mail: valentin.debortoli@gmail.com). Arnaud Doucet is Professor, Department of Statistics, University of Oxford, Oxford OX1 3LB, UK (e-mail: doucet@stats.ox.ac.uk).

DAI PRA, P. (1991). A stochastic control approach to reciprocal diffusion processes. *Appl. Math. Optim.* **23** 313–329. MR1095665 https://doi.org/10.1007/BF01442404

DE BORTOLI, V. (2022). Convergence of denoising diffusion models under the manifold hypothesis. *Trans. Mach. Learn. Res.* **2022**.

DE BORTOLI, V., MATHIEU, E., HUTCHINSON, M., THORNTON, J., TEH, Y. W. and DOUCET, A. (2022). Riemannian score-based generative modeling. In *Advances in Neural Information Processing Systems* 35: *Annual Conference on Neural Information Processing Systems* 2022, *NeurIPS* 2022 (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.).

DE BORTOLI, V., THORNTON, J., HENG, J. and DOUCET, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, *December* 6–14, 2021, *Virtual* (M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang and J. Wortman Vaughan, eds.) 17695–17709.

DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11** 427–444. MR0003527 https://doi.org/10.1214/aoms/1177731829

DOCKHORN, T., VAHDAT, A. and KREIS, K. (2022). Score-based generative modeling with critically-damped Langevin diffusion. In *The Tenth International Conference on Learning Representations*, *ICLR* 2022, *Virtual Event*, *April* 25–29, 2022. OpenReview.net.

FÖLLMER, H. (1985). An entropy approach to the time reversal of diffusion processes. In *Stochastic Differential Systems* (*Marseille-Luminy*, 1984). *Lect. Notes Control Inf. Sci.* **69** 156–163. Springer, Berlin. MR0798318 https://doi.org/10.1007/BFb0005070

FORTET, R. (1940). Résolution d'un système d'équations de M. Schrödinger. *J. Math. Pures Appl.* (9) **19** 83–105. MR0002692

GRENANDER, U. and MILLER, M. I. (1994). Representations of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B* **56** 549–603. MR1293234

HAUSSMANN, U. G. and PARDOUX, É. (1986). Time reversal of diffusions. *Ann. Probab.* **14** 1188–1205. MR0866342

HENG, J., BISHOP, A. N., DELIGIANNIDIS, G. and DOUCET, A. (2020). Controlled sequential Monte Carlo. *Ann. Statist.* **48** 2904–2929. MR4152628 https://doi.org/10.1214/19-AOS1914

HO, J., JAIN, A. and ABBEEL, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems* 33: *Annual Conference on Neural Information Processing Systems* 2020, *NeurIPS* 2020, *December* 6–12, 2020, *Virtual* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin, eds.).

HOOGEBOOM, E., NIELSEN, D., JAINI, P., FORRÉ, P. and WELLING, M. (2021). Argmax flows and multinomial diffusion: Learning categorical distributions. In *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, *December* 6–14, 2021, *Virtual* (M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang and J. Wortman Vaughan, eds.) 12454–12465.

HUANG, C.-W., AGHAJOHARI, M., BOSE, J., PANANGADEN, P. and COURVILLE, A. C. (2022). Riemannian diffusion models. In *Advances in Neural Information Processing Systems* 35: *Annual Conference on Neural Information Processing Systems* 2022, *NeurIPS* 2022 (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.).

IACUS, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations*: *With R Examples*. *Springer Series in Statistics*. Springer, New York. MR2410254 https://doi.org/10.1007/978-0-387-75839-8

KAPPEN, H. J., GÓMEZ, V. and OPPER, M. (2012). Optimal control as a graphical model inference problem. *Mach. Learn.* **87** 159–182. MR2914010 https://doi.org/10.1007/s10994-012-5278-7

KAPPEN, H. J. and RUIZ, H. C. (2016). Adaptive importance sampling for control and inference. *J. Stat. Phys.* **162** 1244–1266. MR3462019 https://doi.org/10.1007/s10955-016-1446-7

KARRAS, T., AITTALA, M., AILA, T. and LAINE, S. (2022). Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems* 35: *Annual Conference on Neural Information Processing Systems* 2022, *NeurIPS* 2022 (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.).

KESSLER, M., LINDNER, A. and SØRENSEN, M., eds. (2012) *Statistical Methods for Stochastic Differential Equations*. *Monographs on Statistics and Applied Probability* **124**. CRC Press, Boca Raton, FL. MR2975799

KINGMA, D. P. and BA, J. (2014). Adam: A method for stochastic optimization. arXiv preprint. Available at arXiv:1412.6980.

KINGMA, D. P. and WELLING, M. (2014). Auto-encoding variational Bayes. In 2*nd International Conference on Learning Representations*, *ICLR* 2014, *Banff*, *AB*, *Canada*, *April* 14–16, 2014, *Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.).

KLEBANER, F. C. (2012). *Introduction to Stochastic Calculus with Applications*, 3rd ed. Imperial College Press, London. MR2933773 https://doi.org/10.1142/p821

KULLBACK, S. (1968). Probability densities with given marginals. *Ann. Math. Stat.* **39** 1236–1243. MR0229330 https://doi.org/10.1214/aoms/1177698249

LÉGER, F. (2021). A gradient descent perspective on Sinkhorn. *Appl. Math. Optim.* **84** 1843–1855. MR4304891 https://doi.org/10.1007/s00245-020-09697-w

LÉONARD, C. (2012). From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** 1879–1920. MR2873864 https://doi.org/10.1016/j.jfa.2011.11.026

LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631 https://doi.org/10.3934/dcds.2014.34.1533

LIPMAN, Y., CHEN, R. T. Q., BEN-HAMU, H., NICKEL, M. and LE, M. (2023). Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

LIU, X., GONG, C. and LIU, Q. (2023). Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

MARIN, J.-M., PUDLO, P., ROBERT, C. P. and RYDER, R. J. (2012). Approximate Bayesian computational methods. *Stat. Comput.* **22** 1167–1180. MR2992292 https://doi.org/10.1007/s11222-011-9288-2

MIKAMI, T. (2004). Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields* **129** 245–260. MR2063377 https://doi.org/10.1007/s00440-004-0340-4

MNIH, A. and GREGOR, K. (2014). Neural variational inference and learning in belief networks. In *Proceedings of the* 31*th International Conference on Machine Learning*, *ICML* 2014, *Beijing*, *China*, 21–26 *June* 2014. *JMLR Workshop and Conference Proceedings* **32** 1791–1799. JMLR.org.

MONTANARI, A. (2023). Sampling, diffusions, and stochastic localization. arXiv preprint. Available at arXiv:2305.10690.

MONTANARI, A. and WU, Y. (2023). Posterior sampling from the spiked models via diffusion processes. arXiv preprint. Available at arXiv:2304.11449.

ØKSENDAL, B. (2003). *Stochastic Differential Equations*: *An Introduction with Applications*, 6th ed. *Universitext*. Springer, Berlin. MR2001996 https://doi.org/10.1007/978-3-642-14394-6

PELUCHETTI, S. (2023). Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. arXiv preprint. Available at arXiv:2304.00917.

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions*, *Markov Processes*, *and Martingales*, *Vol*. 2. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. Itô calculus, Reprint of the second (1994) edition. MR1780932 https://doi.org/10.1017/CBO9781107590120

RÜSCHENDORF, L. (1995). Convergence of the iterative proportional fitting procedure. *Ann*. *Statist*. **23** 1160–1174. MR1353500 https://doi.org/10.1214/aos/1176324703

SALIMANS, T. and HO, J. (2022). Progressive distillation for fast sampling of diffusion models. In *The Tenth International Conference on Learning Representations*, *ICLR* 2022, *Virtual Event*, *April* 25–29, 2022. OpenReview.net.

SHARROCK, L., SIMONS, J., LIU, S. and BEAUMONT, M. (2022). Sequential neural score estimation: Likelihood-free inference with conditional score based diffusion models. arXiv preprint. Available at arXiv:2210.04872.

SHI, Y., BORTOLI, V. D., DELIGIANNIDIS, G. and DOUCET, A. (2022). Conditional simulation using diffusion Schrödinger bridges. In *Uncertainty in Artificial Intelligence*, *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, *UAI* 2022, 1–5 *August* 2022, *Eindhoven*, *The Netherlands* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 1792–1802. PMLR.

SHI, Y., DE BORTOLI, V., CAMPBELL, A. and DOUCET, A. (2023). Diffusion Schrödinger bridge matching. arXiv preprint. Available at arXiv:2303.16852.

SISSON, S. A., FAN, Y. and BEAUMONT, M. (2018). *Handbook of Approximate Bayesian Computation*. CRC Press, Boca Raton, FL.

SOHL-DICKSTEIN, J., WEISS, E. A., MAHESWARANATHAN, N. and GANGULI, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the* 32*nd International Conference on Machine Learning*, *ICML* 2015, *Lille*, *France*, 6–11 *July* 2015 (F. R. Bach and D. M. Blei, eds.). *JMLR Workshop and Conference Proceedings* **37** 2256–2265. JMLR.org.

SONG, J., ZHANG, Q., YIN, H., MARDANI, M., LIU, M., KAUTZ, J., CHEN, Y. and VAHDAT, A. (2023). Loss-guided diffusion models for plug-and-play controllable generation. In *Proceedings of the* 40*th International Conference on Machine Learning* (A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato and J. Scarlett, eds.). *Proceedings of Machine Learning Research* **202** 32483–32498. PMLR.

SONG, Y., DURKAN, C., MURRAY, I. and ERMON, S. (2021). Maximum likelihood training of score-based diffusion models. In *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, *December* 6–14, 2021, *Virtual* (M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang and J. Wortman Vaughan, eds.) 1415–1428.

SONG, Y., SOHL-DICKSTEIN, J., KINGMA, D. P., KUMAR, A., ERMON, S. and POOLE, B. (2021). Score-based generative modeling through stochastic differential equations. In 9*th International Conference on Learning Representations*, *ICLR* 2021, *Virtual Event*, *Austria*, *May* 3–7, 2021. OpenReview.net.

TZEN, B. and RAGINSKY, M. (2019). Theoretical guarantees for sampling and inference in generative models with latent diffusions. In *Conference on Learning Theory*, *COLT* 2019, 25–28 *June* 2019, *Phoenix*, *AZ*, *USA* (A. Beygelzimer and D. Hsu, eds.). *Proceedings of Machine Learning Research* **99** 3084–3114. PMLR.

VARGAS, F., GRATHWOHL, W. S. and DOUCET, A. (2023). Denoising diffusion samplers. In *The Eleventh International Conference on Learning Representations*, *ICLR* 2023, *Kigali*, *Rwanda*, *May* 1–5, 2023. OpenReview.net.

VARGAS, F., OVSIANAS, A., FERNANDES, D., GIROLAMI, M., LAWRENCE, N. D. and NÜSKEN, N. (2023). Bayesian learning via neural Schrödinger–Föllmer flows. *Stat*. *Comput*. **33** Paper No. 3, 22. MR4514992 https://doi.org/10.1007/s11222-022-10172-5

VARGAS, F., THODOROFF, P., LAMACRAFT, A. and LAWRENCE, N. (2021). Solving Schrödinger bridges via maximum likelihood. *Entropy* **23** Paper No. 1134, 30. MR4319496 https://doi.org/10.3390/e23091134

VINCENT, P. (2011). A connection between score matching and denoising autoencoders. *Neural Comput*. **23** 1661–1674. MR2839543 https://doi.org/10.1162/NECO_a_00142

ZHANG, B., SAHAI, T. and MARZOUK, Y. (2021). Sampling via controlled stochastic dynamical systems. In *I* (*Still*) *Can't Believe It's Not Better! NeurIPS* 2021 *Workshop*.

ZHANG, Q. and CHEN, Y. (2022). Path integral sampler: A stochastic control approach for sampling. In *The Tenth International Conference on Learning Representations*, *ICLR* 2022, *Virtual Event*, *April* 25–29, 2022. OpenReview.net.

Statistical Science
2024, Vol. 39, No. 1, 100–114
https://doi.org/10.1214/23-STS915
This research was funded, in whole or in part, by [EPSRC, EP/SO23151/1]. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions.

Modern Bayesian Experimental Design

Tom Rainforth, Adam Foster, Desi R. Ivanova and Freddie Bickford Smith

Abstract. Bayesian experimental design (BED) provides a powerful and general framework for optimizing the design of experiments. However, its deployment often poses substantial computational challenges that can undermine its practical use. In this review, we outline how recent advances have transformed our ability to overcome these challenges and thus utilize BED effectively, before discussing some areas for future development in the field.

Key words and phrases: Bayesian optimal design, Bayesian adaptive design, active learning, information maximization.

REFERENCES

[1] AMZAL, B., BOIS, F. Y., PARENT, E. and ROBERT, C. P. (2006). Bayesian-optimal design via interacting particle systems. *J. Amer. Statist. Assoc.* **101** 773–785. MR2281248 https://doi.org/10.1198/016214505000001159

[2] ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS. Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

[3] ATKINSON, A. C. and FEDOROV, V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika* **62** 57–70. MR0370955 https://doi.org/10.1093/biomet/62.1.57

[4] BARBER, D. and AGAKOV, F. (2003). The IM algorithm: A variational approach to information maximization. In *Conference on Neural Information Processing Systems*.

[5] BARNARD, G. A., JENKINS, G. M. and WINSTEN, C. B. (1962). Likelihood inference and time series. *J. R. Stat. Soc., A*.

[6] BECK, J., DIA, B. M., ESPATH, L. F. R., LONG, Q. and TEMPONE, R. (2018). Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Comput. Methods Appl. Mech. Engrg.* **334** 523–553. MR3775005 https://doi.org/10.1016/j.cma.2018.01.053

[7] BELGHAZI, M. I., BARATIN, A., RAJESHWAR, S., OZAIR, S., BENGIO, Y., COURVILLE, A. and HJELM, D. (2018). Mutual information neural estimation. In *International Conference on Machine Learning*.

[8] BELUCH, W. H., GENEWEIN, T., NÜRNBERGER, A. and KÖHLER, J. M. (2018). The power of ensembles for active learning in image classification. In *Conference on Computer Vision and Pattern Recognition*.

[9] BERGER, J. O. and WOLPERT, R. L. (1984). *The Likelihood Principle. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **6**. IMS, Hayward. MR0773665

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

[11] BHANOT, G. and KENNEDY, A. D. (1984). Lattices, demons and the microcanonical ensemble. In *Progress in Gauge Field Theory* (*Cargèse*, 1983). *NATO Adv. Sci. Inst. Ser. B: Phys.* **115** 155–167. Plenum, New York. MR0782330

[12] BICKFORD SMITH, F., KIRSCH, A., FARQUHAR, S., GAL, Y., FOSTER, A. and RAINFORTH, T. (2023). Prediction-oriented Bayesian active learning. In *International Conference on Artificial Intelligence and Statistics*.

[13] BLAU, T., BONILLA, E. V., CHADES, I. and DEZFOULI, A. (2022). Optimizing sequential experimental design with deep reinforcement learning. In *International Conference on Machine Learning*.

[14] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

[15] BORTH, D. M. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. *J. Roy. Statist. Soc. Ser. B* **37** 77–87. MR0375681

[16] BOX, G. E. P. (1982). Choice of response surface design and alphabetic optimality. Technical Report, Univ. Wisconsin—Madison. MR0683914

[17] BRYNJARSDÓTTIR, J. and O'HAGAN, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30** 114007, 24. MR3274591 https://doi.org/10.1088/0266-5611/30/11/114007

[18] BURDA, Y., GROSSE, R. and SALAKHUTDINOV, R. (2015). Importance weighted autoencoders. arXiv preprint. Available at arXiv:1509.00519.

[19] CARLIN, B. P., KADANE, J. B. and GELFAND, A. E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics*.

[20] CARLON, A. G., DIA, B. M., ESPATH, L., LOPEZ, R. H. and TEMPONE, R. (2020). Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. *Comput. Methods Appl. Mech. Engrg.* **363** 112909, 27. MR4067945 https://doi.org/10.1016/j.cma.2020.112909

[21] CAVAGNARO, D. R., MYUNG, J. I., PITT, M. A. and KUJALA, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in

Tom Rainforth is a Senior Researcher, Department of Statistics, University of Oxford, Oxford, UK (e-mail: rainforth@stats.ox.ac.uk). Adam Foster is a Senior Researcher, Microsoft Research AI4Science, Cambridge, UK (e-mail: adam.e.foster@microsoft.com). Desi R. Ivanova is a DPhil Student, University of Oxford, Oxford, UK (e-mail: desi.ivanova@stats.ox.ac.uk). Freddie Bickford Smith is a DPhil Student, University of Oxford, Oxford, UK (e-mail: freddie@robots.ox.ac.uk).

cognitive science. *Neural Comput*. **22** 887–905. MR2654272 https://doi.org/10.1162/neco.2009.02-09-959

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

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

[24] CHEN, R.-B., CHANG, S.-P., WANG, W., TUNG, H.-C. and WONG, W. K. (2015). Minimax optimal designs via particle swarm optimization methods. *Stat. Comput*. **25** 975–988. MR3375629 https://doi.org/10.1007/s11222-014-9466-0

[25] CHENG, Y. and SHEN, Y. (2005). Bayesian adaptive designs for clinical trials. *Biometrika* **92** 633–646. MR2202651 https://doi.org/10.1093/biomet/92.3.633

[26] CLYDE, M. A., MÜLLER, P. and PARMIGIANI, G. (1996). Exploring expected utility surfaces by Markov chains. Technical Report.

[27] COOK, A. R., GIBSON, G. J. and GILLIGAN, C. A. (2008). Optimal observation times in experimental epidemic processes. *Biometrics* **64** 860–868. MR2526637 https://doi.org/10.1111/j.1541-0420.2007.00931.x

[28] COVEY-CRUMP, P. A. K. and SILVEY, S. D. (1970). Optimal regression designs with previous observations. *Biometrika* **57** 551–566. MR0275590 https://doi.org/10.1093/biomet/57.3.551

[29] DEHIDENIYA, M. B., DROVANDI, C. C. and MCGREE, J. M. (2018). Optimal Bayesian design for discriminating between models with intractable likelihoods in epidemiology. *Comput. Statist. Data Anal*. **124** 277–297. MR3787626 https://doi.org/10.1016/j.csda.2018.03.004

[30] DROVANDI, C. C. and PETTITT, A. N. (2013). Bayesian experimental design for models with intractable likelihoods. *Biometrics* **69** 937–948. MR3146789 https://doi.org/10.1111/biom.12081

[31] DUFF, M. O. (2002). Optimal learning: Computational procedures for Bayes-adaptive Markov decision processes. Ph.D. thesis, Univ. Massachusetts Amherst. MR2703036

[32] DUSHENKO, S., AMBAL, K. and MCMICHAEL, R. D. (2020). Sequential Bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers. *Appl. Phys. Rev.* **14**. https://doi.org/10.1103/PhysRevApplied.14.054036

[33] ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Stat*. **23** 255–262. MR0047998 https://doi.org/10.1214/aoms/1177729442

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

[35] FANOURGAKIS, G. S., KANAKIDOU, M., NENES, A., BAUER, S. E., BERGMAN, T., CARSLAW, K. S., GRINI, A., HAMILTON, D. S., JOHNSON, J. S. et al. (2019). Evaluation of global simulations of aerosol particle and cloud condensation nuclei number, with implications for cloud droplet formation. *Atmos. Chem. Phys*.

[36] FARQUHAR, S., GAL, Y. and RAINFORTH, T. (2021). On statistical bias in active learning: How and when to fix it. In *International Conference on Learning Representations*.

[37] FENG, C. (2015). Optimal Bayesian experimental design in the presence of model error. Master's thesis, Massachusetts Institute of Technology.

[38] FIRTH, D. and HINDE, J. P. (1997). Parameter neutral optimum design for non-linear models. *J. Roy. Statist. Soc. Ser. B* **59** 799–811. MR1483215 https://doi.org/10.1111/1467-9868.00097

[39] FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric*.

[40] FISHER, R. A. (1936). The design of experiments. *Br. Med. J*. MR1523616 https://doi.org/10.2307/2300364

[41] FOSTER, A., IVANOVA, D. R., MALIK, I. and RAINFORTH, T. (2021). Deep adaptive design: Amortizing sequential Bayesian experimental design. In *International Conference on Machine Learning*.

[42] FOSTER, A., JANKOWIAK, M., BINGHAM, E., HORSFALL, P., TEH, Y. W., RAINFORTH, T. and GOODMAN, N. (2019). Variational Bayesian optimal experimental design. In *Conference on Neural Information Processing Systems*.

[43] FOSTER, A., JANKOWIAK, M., O'MEARA, M., TEH, Y. W. and RAINFORTH, T. (2020). A unified stochastic gradient approach to designing Bayesian-optimal experiments. In *International Conference on Artificial Intelligence and Statistics*.

[44] FOSTER, A. E. (2021). Variational, Monte Carlo and policy-based approaches to Bayesian experimental design. Ph.D. thesis, Univ. Oxford.

[45] FUDENBERG, D., ROMANYUK, G. and STRACK, P. (2017). Active learning with a misspecified prior. *Theor. Econ*. **12** 1155–1189. MR3713677 https://doi.org/10.3982/te2480

[46] GAL, Y., ISLAM, R. and GHAHRAMANI, Z. (2017). Deep Bayesian active learning with image data. In *International Conference on Machine Learning*.

[47] GELMAN, A. (2008). Objections to Bayesian statistics. *Bayesian Anal*. **3** 445–449. MR2434394 https://doi.org/10.1214/08-BA318

[48] GHAVAMZADEH, M., MANNOR, S., PINEAU, J. and TAMAR, A. (2015). Bayesian reinforcement learning: A survey. *Found. Trends Mach. Learn*.

[49] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Oper. Res*. **56** 607–617. MR2436856 https://doi.org/10.1287/opre.1070.0496

[50] GILES, M. B. and SZPRUCH, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab*. **24** 1585–1620. MR3211005 https://doi.org/10.1214/13-AAP957

[51] GILLESPIE, C. S. and BOYS, R. J. (2019). Efficient construction of Bayes optimal designs for stochastic process models. *Stat. Comput*. **29** 697–706. MR3955280 https://doi.org/10.1007/s11222-018-9833-3

[52] GO, J. and ISAAC, T. (2022). Robust expected information gain for optimal Bayesian experimental design using ambiguity sets. In *Conference on Uncertainty in Artificial Intelligence*.

[53] GODA, T., HIRONAKA, T. and IWAMOTO, T. (2020). Multilevel Monte Carlo estimation of expected information gains. *Stoch. Anal. Appl*. **38** 581–600. MR4112737 https://doi.org/10.1080/07362994.2019.1705168

[54] GODA, T., HIRONAKA, T., KITADE, W. and FOSTER, A. (2022). Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. *SIAM J. Sci. Comput*. **44** A286–A311. MR4372651 https://doi.org/10.1137/20M1338848

[55] GOLOVIN, D., KRAUSE, A. and RAY, D. (2010). Near-optimal Bayesian active learning with noisy observations. In *Conference on Neural Information Processing Systems*.

[56] GONZÁLEZ, J., OSBORNE, M. and LAWRENCE, N. (2016). GLASSES: Relieving the myopia of Bayesian optimisation. In *International Conference on Artificial Intelligence and Statistics*.

[57] GRÜNWALD, P. and VAN OMMEN, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal*. **12** 1069–1103. MR3724979 https://doi.org/10.1214/17-BA1085

[58] GUEZ, A., SILVER, D. and DAYAN, P. (2012). Efficient Bayes-adaptive reinforcement learning using sample-based search. In *Conference on Neural Information Processing Systems*.

[59] GUO, Q., CHEN, J., WANG, D., YANG, Y., DENG, X., HUANG, J., CARIN, L., LI, F. and TAO, C. (2022). Tight mutual information estimation with contrastive Fenchel–Legendre optimization. In *Conference on Neural Information Processing Systems*.

[60] HAINY, M., DROVANDI, C. C. and MCGREE, J. M. (2016). Likelihood-free extensions for Bayesian sequentially designed experiments. In *mODa* 11*—Advances in Model-Oriented Design and Analysis*.

[61] HAMADA, M., MARTZ, H. F., REESE, C. S. and WILSON, A. G. (2001). Finding near-optimal Bayesian experimental designs via genetic algorithms. *Amer. Statist.* **55** 175–181. MR1947263 https://doi.org/10.1198/000313001317098121

[62] HAN, C. and CHALONER, K. (2004). Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics* **60** 25–33. MR2043615 https://doi.org/10.1111/j.0006-341X.2004.00148.x

[63] HENNIG, P. and SCHULER, C. J. (2012). Entropy search for information-efficient global optimization. *J. Mach. Learn. Res.* **13** 1809–1837. MR2956343

[64] HERNÁNDEZ-LOBATO, D., HERNÁNDEZ-LOBATO, J., SHAH, A. and ADAMS, R. (2016). Predictive entropy search for multi-objective Bayesian optimization. In *International Conference on Machine Learning*.

[65] HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

[66] HOULSBY, N., HUSZÁR, F. and GHAHRAMANI, Z. (2011). Bayesian active learning for classification and preference learning. arXiv preprint. Available at arXiv:1112.5745.

[67] HUAN, X. and MARZOUK, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* **232** 288–317. MR2994301 https://doi.org/10.1016/j.jcp.2012.08.013

[68] HUAN, X. and MARZOUK, Y. M. (2014). Gradient-based stochastic optimization methods in Bayesian experimental design. *Int. J. Uncertain. Quantif.* **4** 479–510. MR3275924 https://doi.org/10.1615/Int.J.UncertaintyQuantification.2014006730

[69] HUAN, X. and MARZOUK, Y. M. (2016). Sequential Bayesian optimal experimental design via approximate dynamic programming. arXiv preprint. Available at arXiv:1604.08320.

[70] HUGHES-OLIVER, J. M. (1998). Optimal designs for nonlinear models with correlated errors. *IMS Lecture Notes Monogr. Ser.*

[71] IGL, M., ZINTGRAF, L., LE, T. A., WOOD, F. and WHITESON, S. (2018). Deep variational reinforcement learning for POMDPs. In *International Conference on Machine Learning*.

[72] IVANOVA, D. R., FOSTER, A., KLEINEGESSE, S., GUTMANN, M. U. and RAINFORTH, T. (2021). Implicit deep adaptive design: Policy-based experimental design without likelihoods. In *Conference on Neural Information Processing Systems*.

[73] IVANOVA, D. R., JENNINGS, J., RAINFORTH, T., ZHANG, C. and FOSTER, A. (2023). CO-BED: Information-theoretic contextual optimization via Bayesian experimental design. In *International Conference on Machine Learning*.

[74] JIANG, S., CHAI, H., GONZALEZ, J. and GARNETT, R. (2020). BINOCULARS for efficient, nonmyopic sequential experimental design. In *International Conference on Machine Learning*.

[75] JÖRKE, M., LEE, J. and BRUNSKILL, E. (2022). Simple regret minimization for contextual bandits using Bayesian optimal experimental design. In *ICML Workshop on "Adaptive Experimental Design and Active Learning in the Real World"*.

[76] KAPOOR, A., GRAUMAN, K., URTASUN, R. and DARRELL, T. (2007). Active learning with Gaussian processes for object categorization. In *International Conference on Computer Vision*.

[77] KASY, M. and SAUTMANN, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica* **89** 113–132. MR4220384 https://doi.org/10.3982/ecta17527

[78] KENNAMER, N., WALTON, S. and IHLER, A. (2023). Design amortization for Bayesian optimal experimental design. In *AAAI Conference on Artificial Intelligence*.

[79] KENNEDY, A. D. and KUTI, J. (1985). Noise without noise: A new Monte Carlo method. *Phys. Rev. Lett.*

[80] KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 https://doi.org/10.1111/1467-9868.00294

[81] KINGMA, D. P. and BA, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

[82] KINGMA, D. P. and WELLING, M. (2014). Auto-encoding variational Bayes. In *International Conference on Learning Representations*.

[83] KIRSCH, A., VAN AMERSFOORT, J. and GAL, Y. (2019). BatchBALD: Efficient and diverse batch acquisition for deep Bayesian active learning. In *Conference on Neural Information Processing Systems*.

[84] KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein–Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675

[85] KLEINEGESSE, S., DROVANDI, C. and GUTMANN, M. U. (2021). Sequential Bayesian experimental design for implicit models via mutual information. *Bayesian Anal.* **16** 773–802. MR4303868 https://doi.org/10.1214/20-BA1225

[86] KLEINEGESSE, S. and GUTMANN, M. (2020). Bayesian experimental design for implicit models by mutual information neural estimation. In *International Conference on Machine Learning*.

[87] KLEINEGESSE, S. and GUTMANN, M. U. (2019). Efficient Bayesian experimental design for implicit models. In *International Conference on Artificial Intelligence and Statistics*.

[88] KOSSEN, J., FARQUHAR, S., GAL, Y. and RAINFORTH, T. (2021). Active testing: Sample-efficient model evaluation. In *International Conference on Machine Learning*.

[89] KOSSEN, J., FARQUHAR, S., GAL, Y. and RAINFORTH, T. (2022). Active surrogate estimators: An active learning approach to label-efficient model evaluation. In *Conference on Neural Information Processing Systems*.

[90] KUCK, H., DE FREITAS, N. and DOUCET, A. (2006). SMC samplers for Bayesian optimal nonlinear design. In *IEEE Nonlinear Statistical Signal Processing Workshop*.

[91] KUHFELD, W. F., TOBIAS, R. D. and GARRATT, M. (1994). Efficient experimental design with marketing research applications. *J. Mark. Res.*

[92] LEWI, J., BUTERA, R. and PANINSKI, L. (2009). Sequential optimal design of neurophysiology experiments. *Neural Comput.* **21** 619–687. MR2478314 https://doi.org/10.1162/neco.2008.08-07-594

[93] LIM, V., NOVOSELLER, E., ICHNOWSKI, J., HUANG, H. and GOLDBERG, K. (2022). Policy-based Bayesian experimental design for non-differentiable implicit models. arXiv preprint. Available at arXiv:2203.04272.

[94] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Stat.* **27** 986–1005. MR0083936 https://doi.org/10.1214/aoms/1177728069

[95] LINDLEY, D. V. (1971). *Bayesian Statistics, a Review. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics* **2**. SIAM, Philadelphia. MR0329081

[96] LONG, Q. (2022). Multimodal information gain in Bayesian design of experiments. *Comput. Statist.* **37** 865–885. MR4400226 https://doi.org/10.1007/s00180-021-01145-9

[97] LONG, Q., SCAVINO, M., TEMPONE, R. and WANG, S. (2013). Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Comput. Methods Appl. Mech. Engrg.* **259** 24–39. MR3063259 https://doi.org/10.1016/j.cma.2013.02.017

[98] LOREDO, T. J. (2004). Bayesian adaptive exploration. *AIP Conf. Proc.* **707** 330–346. MR2085769 https://doi.org/10.1063/1.1751377

[99] LYNE, A.-M., GIROLAMI, M., ATCHADÉ, Y., STRATHMANN, H. and SIMPSON, D. (2015). On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Statist. Sci.* **30** 443–467. MR3432836 https://doi.org/10.1214/15-STS523

[100] LYU, J., WANG, S., BALIUS, T. E., SINGH, I., LEVIT, A., MOROZ, Y. S., O'MEARA, M. J., CHE, T., ALGAA, E. et al. (2019). Ultra-large library docking for discovering new chemotypes. *Nature* **566** 224–229. https://doi.org/10.1038/s41586-019-0917-9

[101] MACKAY, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Comput.*

[102] MARGATINA, K., BARRAULT, L. and ALETRAS, N. (2021). Bayesian active learning with pretrained language models. arXiv preprint. Available at arXiv:2104.08320.

[103] MCLEISH, D. (2011). A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.* **17** 301–315. MR2890424 https://doi.org/10.1515/mcma.2011.013

[104] MEDICAL OUTREACH SUBTEAM OF THE DRUG INFORMATION ASSOCIATION BAYESIAN SCIENTIFIC WORKING GROUP, CLARK, J., MUHLEMANN, N., NATANEGARA, F., HARTLEY, A., WENKERT, D., WANG, F., HARRELL, F. E. and BRAY, R. (2022). Why are not there more Bayesian clinical trials? Perceived barriers and educational preferences among medical researchers involved in drug development. *Ther. Innov. Regul. Sci.*

[105] MELENDEZ, J. A., FURNSTAHL, R. J., GRIESSHAMMER, H. W., MCGOVERN, J. A., PHILLIPS, D. R. and PRATOLA, M. T. (2021). Designing optimal experiments: An application to proton Compton scattering. *Eur. Phys. J. A.*

[106] MEYER, R. K. and NACHTSHEIM, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* **37** 60–69. MR1322047 https://doi.org/10.2307/1269153

[107] MÜLLER, P. (2005). Simulation based optimal design. In *Bayesian Thinking: Modeling and Computation. Handbook of Statist.* **25** 509–518. Elsevier, Amsterdam. MR2490537 https://doi.org/10.1016/S0169-7161(05)25017-4

[108] MÜLLER, P., SANSÓ, B. and DE IORIO, M. (2004). Optimal Bayesian design by inhomogeneous Markov chain simulation. *J. Amer. Statist. Assoc.* **99** 788–798. MR2090911 https://doi.org/10.1198/016214504000001123

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

[110] NEISWANGER, W., WANG, K. A. and ERMON, S. (2021). Bayesian algorithm execution: Estimating computable properties of black-box functions using mutual information. In *International Conference on Machine Learning*.

[111] NGUYEN, X., WAINWRIGHT, M. J. and JORDAN, M. I. (2010). Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Trans. Inf. Theory* **56** 5847–5861. MR2808937 https://doi.org/10.1109/TIT.2010.2068870

[112] OSBAND, I., ASGHARI, S. M., VAN ROY, B., MCALEESE, N., ASLANIDES, J. and IRVING, G. (2023). Fine-tuning language models via epistemic neural networks. arXiv preprint. Available at arXiv:2211.01568.

[113] OUYANG, L., TESSLER, M. H., LY, D. and GOODMAN, N. (2016). Practical optimal experiment design with probabilistic programs. In *Conference on Neural Information Processing Systems*.

[114] OVERSTALL, A. and MCGREE, J. (2020). Bayesian design of experiments for intractable likelihood models using coupled auxiliary models and multivariate emulation. *Bayesian Anal.* **15** 103–131. MR4050879 https://doi.org/10.1214/19-BA1144

[115] OVERSTALL, A. and MCGREE, J. (2022). Bayesian decision-theoretic design of experiments under an alternative model. *Bayesian Anal.* **17** 1021–1041. MR4506020 https://doi.org/10.1214/21-ba1286

[116] OVERSTALL, A. M. (2022). Properties of Fisher information gain for Bayesian design of experiments. *J. Statist. Plann. Inference* **218** 138–146. MR4341631 https://doi.org/10.1016/j.jspi.2021.10.006

[117] OVERSTALL, A. M. and WOODS, D. C. (2017). Bayesian design of experiments using approximate coordinate exchange. *Technometrics* **59** 458–470. MR3740963 https://doi.org/10.1080/00401706.2016.1251495

[118] PACHECO, J. and FISHER, J. (2019). Variational information planning for sequential decision making. In *International Conference on Artificial Intelligence and Statistics*.

[119] PALLMANN, P., BEDDING, A. W., CHOODARI-OSKOOEI, B., DIMAIRO, M., FLIGHT, L., HAMPSON, L. V., HOLMES, J., MANDER, A. P., ODONDI, L. et al. (2018). Adaptive designs in clinical trials: Why use them, and how to run and report them. *BMC Med.* **16** 29. https://doi.org/10.1186/s12916-018-1017-7

[120] PANINSKI, L. (2005). Asymptotic theory of information-theoretic experimental design. *Neural Comput.* **17** 1480–1507. https://doi.org/10.1162/0899766053723032

[121] PAPADIMITRIOU, C. (2004). Optimal sensor placement methodology for parametric identification of structural systems. *J. Sound Vib.* **278** 923–947. MR2101893 https://doi.org/10.1016/j.jsv.2003.10.063

[122] PINSLER, GORDON, NALISNICK and HERNÁNDEZ-LOBATO (2019). Bayesian batch active learning as sparse subset approximation. In *Conference on Neural Information Processing Systems*.

[123] POOLE, B., OZAIR, S., VAN DEN OORD, A. and ALEMI, A. (2019). On variational bounds of mutual information. In *International Conference on Machine Learning*.

[124] PRANGLE, D., HARBISHER, S. and GILLESPIE, C. S. (2023). Bayesian experimental design without posterior calculations: An adversarial approach. *Bayesian Anal.* **18** 133–163. MR4515728 https://doi.org/10.1214/22-ba1306

[125] PRICE, D. J., BEAN, N. G., ROSS, J. V. and TUKE, J. (2016). On the efficient determination of optimal Bayesian experimental designs using ABC: A case study in optimal observation of epidemics. *J. Statist. Plann. Inference* **172** 1–15. MR3461907 https://doi.org/10.1016/j.jspi.2015.12.008

[126] PRICE, D. J., BEAN, N. G., ROSS, J. V. and TUKE, J. (2018). An induced natural selection heuristic for finding optimal Bayesian experimental designs. *Comput. Statist. Data Anal.* **126** 112–124. MR3808393 https://doi.org/10.1016/j.csda.2018.04.011

[127] RAINFORTH, T. (2017). Automating inference, learning, and design using probabilistic programming. Ph.D. thesis, Univ. Oxford.

[128] RAINFORTH, T., CORNISH, R., YANG, H. and WARRINGTON, A. (2018). On nesting Monte Carlo estimators. In *International Conference on Machine Learning*.

[129] REZENDE, D. J., MOHAMED, S. and WIERSTRA, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning*.

[130] RHEE, C.-H. and GLYNN, P. W. (2015). Unbiased estimation with square root convergence for SDE models. *Oper. Res.* **63** 1026–1043. MR3422533 https://doi.org/10.1287/opre.2015.1404

[131] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

[132] ROBERTSON, D. S., LEE, K. M., LÓPEZ-KOLKOVSKA, B. C. and VILLAR, S. S. (2023). Response-adaptive randomization in clinical trials: From myths to practical considerations. *Statist. Sci.* **38** 185–208. MR4597328 https://doi.org/10.1214/22-sts865

[133] ROSS, S., CHAIB-DRAA, B. and PINEAU, J. (2007). Bayes-adaptive POMDPs. In *Conference on Neural Information Processing Systems*.

[134] RUBERG, S. J., BECKERS, F., HEMMINGS, R., HONIG, P., IRONY, T., LAVANGE, L., LIEBERMAN, G., MAYNE, J. and MOSCICKI, R. (2023). Application of Bayesian approaches in drug development: Starting a virtuous cycle. *Nat. Rev. Drug Discov.* **22** 235–250. https://doi.org/10.1038/s41573-023-00638-0

[135] RYAN, E. G., DROVANDI, C. C., MCGREE, J. M. and PETTITT, A. N. (2016). A review of modern computational algorithms for Bayesian optimal design. *Int. Stat. Rev.* **84** 128–154. MR3491282 https://doi.org/10.1111/insr.12107

[136] RYAN, E. G., DROVANDI, C. C. and PETTITT, A. N. (2015). Fully Bayesian experimental design for pharmacokinetic studies. *Entropy*.

[137] RYAN, E. G., DROVANDI, C. C., THOMPSON, M. H. and PETTITT, A. N. (2014). Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Comput. Statist. Data Anal.* **70** 45–60. MR3125477 https://doi.org/10.1016/j.csda.2013.08.017

[138] RYAN, K. J. (2003). Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *J. Comput. Graph. Statist.* **12** 585–603. MR2002637 https://doi.org/10.1198/1061860032012

[139] RYAN, T. P. and MORGAN, J. P. (2007). Modern experimental design. *J. Stat. Theory Pract.* MR2285428 https://doi.org/10.1002/0470074353

[140] SCHMIDHUBER, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans. Auton. Ment. Dev.*

[141] SEBASTIANI, P. and WYNN, H. P. (2000). Maximum entropy sampling and optimal Bayesian experimental design. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 145–157. MR1747401 https://doi.org/10.1111/1467-9868.00225

[142] SETTLES, B. (2009). Active learning literature survey. Technical Report, Univ. Wisconsin—Madison.

[143] SHABABO, B., PAIGE, B., PAKMAN, A. and PANINSKI, L. (2013). Bayesian inference and online experimental design for mapping neural microcircuits. In *Conference on Neural Information Processing Systems*.

[144] SHAHRIARI, B., SWERSKY, K., WANG, Z., ADAMS, R. P. and DE FREITAS, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE.*

[145] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423, 623–656. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[146] SHEN, W. and HUAN, X. (2023). Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning. *Comput. Methods Appl. Mech. Engrg.* **416** Paper No. 116304, 28. MR4626387 https://doi.org/10.1016/j.cma.2023.116304

[147] SHEN, Y., YUN, H., LIPTON, Z. C., KRONROD, Y. and ANANDKUMAR, A. (2018). Deep active learning for named entity recognition. In *International Conference on Learning Representations*.

[148] SIDDHANT, A. and LIPTON, Z. C. (2018). Deep Bayesian active learning for natural language processing: Results of a large-scale empirical study. arXiv preprint. Available at arXiv:1808.05697.

[149] SLOMAN, S. J., OPPENHEIMER, D. M., BROOMELL, S. B. and SHALIZI, C. R. (2022). Characterizing the robustness of Bayesian adaptive experimental designs to active learning bias. arXiv preprint. Available at arXiv:2205.13698.

[150] SMITH, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*.

[151] STROUD, J. R., MÜLLER, P. and ROSNER, G. L. (2001). Optimal sampling times in population pharmacokinetic studies. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 345–359. MR1856330 https://doi.org/10.1111/1467-9876.00239

[152] STUHLMÜLLER, A., TAYLOR, J. and GOODMAN, N. (2013). Learning stochastic inverses. In *Conference on Neural Information Processing Systems*.

[153] SUN, Y., GOMEZ, F. and SCHMIDHUBER, J. (2011). Planning to be surprised: Optimal Bayesian exploration in dynamic environments. In *International Conference on Artificial General Intelligence*.

[154] SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge. MR3889951

[155] SZEPESVÁRI, C. (2010). *Algorithms for Reinforcement Learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Springer, Cham. MR4647095 https://doi.org/10.1007/978-3-031-01551-9

[156] THOMAS, O., DUTTA, R., CORANDER, J., KASKI, S. and GUTMANN, M. U. (2022). Likelihood-free inference by ratio estimation. *Bayesian Anal.* **17** 1–31. MR4377135 https://doi.org/10.1214/20-ba1238

[157] TSUTAKAWA, R. K. (1972). Design of experiment for bioassay. *J. Amer. Statist. Assoc.*

[158] US FOOD AND DRUG ADMINISTRATION (2019). Adaptive designs for clinical trials of drugs and biologics: Guidance for industry.

[159] VAN DEN BERG, J., CURTIS, A. and TRAMPERT, J. (2003). Optimal nonlinear Bayesian experimental design: An application to amplitude versus offset experiments. *Geophys. J. Int.*

[160] VAN DEN OORD, A., LI, Y. and VINYALS, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint. Available at arXiv:1807.03748.

[161] VANLIER, J., TIEMANN, C. A., HILBERS, P. A. J. and VAN RIEL, N. A. W. (2012). A Bayesian approach to targeted experiment design. *Bioinformatics*.

[162] VINCENT, B. T. and RAINFORTH, T. (2017). The DARC Toolbox: Automated, flexible, and efficient delayed and risky choice experiments using Bayesian adaptive design. PsyArXiv preprint.

[163] VON KÜGELGEN, J., RUBENSTEIN, P. K., SCHÖLKOPF, B. and WELLER, A. (2019). Optimal experimental design via Bayesian optimization: Active causal structure learning for Gaussian process networks. In *NeurIPS Workshop on "Do the Right Thing: Machine Learning and Causal Inference for Improved Decision Making"*.

[164] WALKER, S. G. (2016). Bayesian information in an experiment and the Fisher information distance. *Statist. Probab. Lett.* **112** 5–9. MR3475481 https://doi.org/10.1016/j.spl.2016.01.014

[165] WANG, Z. and JEGELKA, S. (2017). Max-value entropy search for efficient Bayesian optimization. In *International Conference on Machine Learning*.

[166] WATSON, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Int. J. Comput. Vis.*

[167] WITTLER, N., ROY, F., PACK, K., WERNINGHAUS, M., ROY, A. S., EGGER, D. J., FILIPP, S., WILHELM, F. K. and MACHNES, S. (2021). Integrated tool set for control, calibration, and characterization of quantum devices applied to superconducting qubits. *Phys. Rev., A*.

[168] ZHANG, J., BI, S. and ZHANG, G. (2021). A scalable gradient-free method for Bayesian experimental design with implicit models. In *International Conference on Artificial Intelligence and Statistics*.

[169] ZHENG, S., HAYDEN, D., PACHECO, J. and FISHER, J. W. III (2020). Sequential Bayesian experimental design with variable cost structure. In *Conference on Neural Information Processing Systems*.

[170] ZHENG, S., PACHECO, J. and FISHER, J. (2018). A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*.

[171] ZHOU, X., LIU, S., KIM, E. S., HERBST, R. S. and LEE, J. J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clin. Trials*.

Statistical Science
2024, Vol. 39, No. 1, 115–136
https://doi.org/10.1214/23-STS888

Living on the Edge: An Unified Approach to Antithetic Sampling

Roberto Casarin, Radu V. Craiu, Lorenzo Frattarolo and Christian P. Robert

Abstract. We identify recurrent ingredients in the antithetic sampling literature leading to a unified sampling framework. We introduce a new class of antithetic schemes that includes the most used antithetic proposals. This perspective enables the derivation of new properties of the sampling schemes: (i) optimality in the Kullback–Leibler sense; (ii) closed-form multivariate Kendall's τ and Spearman's ρ; (iii) ranking in concordance order and (iv) a central limit theorem that characterizes stochastic behaviour of Monte Carlo estimators when the sample size tends to infinity. The proposed simulation framework inherits the simplicity of the standard antithetic sampling method, requiring the definition of a set of reference points in the sampling space and the generation of uniform numbers on the segments joining the points. We provide applications to Monte Carlo integration and Markov Chain Monte Carlo Bayesian estimation.

Key words and phrases: Antithetic variables, countermonotonicity, Monte Carlo, negative dependence, variance reduction.

REFERENCES

AHN, J. Y. and FUCHS, S. (2020). On minimal copulas under the concordance order. *J. Optim. Theory Appl.* **184** 762–780. MR4061667 https://doi.org/10.1007/s10957-019-01618-4

ALBONICO, A., CALÉS, L., CARDANI, R., CROITOROV, O., FERRONI, F., GIOVANNINI, M., HOHBERGER, S., PATARACCHIA, B., PERICOLI, F. M. et al. (2019). Comparing post-crisis dynamics across euro area countries with the global multi-country model. *Econ. Model.* **81** 242–273.

ALIPRANTIS, C. D. and BORDER, K. C. (2007). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin.

ANDRÉASSON, I. J. (1972). Combinations of antithetic methods in simulation Technical Report No. 72.49 Kungliga Tekniska Högskolan. Department of Information Processing. Computer Science.

ANDRÉASSON, I. J. and DAHLQUIST, G. (1972). Groups of antithetic transformations in simulation Technical Report No. 72.57 Kungliga Tekniska Högskolan. Department of Information Processing. Computer Science.

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

ARVIDSEN, N. I. and JOHNSSON, T. (1982). Variance reduction through negative correlation, a simulation study. *J. Stat. Comput. Simul.* **15** 119–127.

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models. International Series in Decision Processes*. Holt, Rinehart and Winston, New York. MR0438625

BIRKHOFF, G. (1946). Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A.* **5** 147–151. MR0020547

BROWN, J. R. (1966). Approximation theorems for Markov operators. *Pacific J. Math.* **16** 13–23. MR0192552

BUBENIK, P. and HOLBROOK, J. (2007). Densities for random balanced sampling. *J. Multivariate Anal.* **98** 350–369. MR2301757 https://doi.org/10.1016/j.jmva.2006.01.007

CASARIN, R., CRAIU, R. V, FRATTAROLO, L. and ROBERT, C. P (2024). Supplement to "Living on the Edge: An Unified Approach to Antithetic Sampling." https://doi.org/10.1214/23-STS888SUPP

CHAN, D., KOHN, R. and KIRBY, C. (2006). Multivariate stochastic volatility models with correlated errors. *Econometric Rev.* **25** 245–274. MR2256288 https://doi.org/10.1080/07474930600713309

CRAIU, R. V. and MENG, X.-L. (2005). Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Ann. Statist.* **33** 661–697. MR2163156 https://doi.org/10.1214/009053604000001075

CRAIU, R. V. and MENG, X.-L. (2006). Meeting Hausdorff in Monte Carlo: A surprising tour with antihype fractals. *Statist. Sinica* **16** 77–91. MR2256080

DADUNA, H. and SZEKLI, R. (2006). Dependence ordering for Markov processes on partially ordered spaces. *J. Appl. Probab.* **43** 793–814. MR2274801 https://doi.org/10.1239/jap/1158784947

DURANTE, F., FERNÁNDEZ SÁNCHEZ, J. and TRUTSCHNIG, W. (2014). Multivariate copulas with hairpin support. *J. Multivariate*

Roberto Casarin is Professor, Ca' Foscari University of Venice, Venice, Italy (e-mail: r.casarin@unive.it). Radu V. Craiu is Professor, University of Toronto, Toronto, Ontario M5G 1Z5 Canada (e-mail: radu.craiu@utoronto.ca). Lorenzo Frattarolo is Assistant Professor, University of Verona, Verona, Italy and European Commission Joint Research Centre, Ispra, Italy (e-mail: lorenzo.frattarolo@univr.it). Christian P. Robert is Professor, CEREMADE, Université Paris-Dauphine PSL, Paris, France, and University of Warwick, Coventry, UK (e-mail: xian@ceremade.dauphine.fr).

Anal. **130** 323–334. MR3229541 https://doi.org/10.1016/j.jmva.2014.06.009

DURANTE, F. and SÁNCHEZ, J. F. (2012). On the approximation of copulas via shuffles of Min. *Statist*. *Probab*. *Lett*. **82** 1761–1767. MR2956614 https://doi.org/10.1016/j.spl.2012.06.008

FISHMAN, G. S. and HUANG, B. S. D. (1983). Antithetic variates revisited. *Commun*. *ACM* **26** 964–971. MR0784123 https://doi.org/10.1145/182.358462

FRÉCHET, M. (1935). Généralisation du théorème des probabilités totales. *Fund*. *Math*. **25** 379–387.

FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann*. *Univ*. *Lyon Sect*. *A* (3) **14** 53–77. MR0049518

FRIGESSI, A., GÅSEMYR, J. and RUE, H. (2000). Antithetic coupling of two Gibbs sampler chains. *Ann*. *Statist*. **28** 1128–1149. MR1810922 https://doi.org/10.1214/aos/1015956710

FUCHS, S., MCCORD, Y. and SCHMIDT, K. D. (2018). Characterizations of copulas attaining the bounds of multivariate Kendall's tau. *J*. *Optim*. *Theory Appl*. **178** 424–438. MR3825632 https://doi.org/10.1007/s10957-018-1285-6

GAFFKE, N. and RÜSCHENDORF, L. (1981). On a class of extremal problems in statistics. *Math*. *Operationsforsch*. *Statist*. *Ser*. *Optim*. **12** 123–135. MR0609133 https://doi.org/10.1080/02331938108842712

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J*. *Amer*. *Statist*. *Assoc*. **85** 398–409. MR1141740

GERBER, M. and CHOPIN, N. (2015). Sequential quasi Monte Carlo. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **77** 509–579. MR3351446 https://doi.org/10.1111/rssb.12104

GEROW, K. and HOLBROOK, J. (1996). Statistical sampling and fractal distributions. *Math*. *Intelligencer* **18** 12–22. MR1395086 https://doi.org/10.1007/BF03027288

HAMMERSLEY, J. M. and MAULDON, J. G. (1956). General principles of antithetic variates. *Proc*. *Camb*. *Philos*. *Soc*. **52** 476–481. MR0081038 https://doi.org/10.1017/s0305004100031467

HAMMERSLEY, J. M. and MORTON, K. W. (1956). A new Monte Carlo technique: Antithetic variates. *Proc*. *Camb*. *Philos*. *Soc*. **52** 449–475. MR0080984 https://doi.org/10.1017/s0305004100031455

HANDSCOMB, D. C. (1958). Proof of the antithetic variates theorem for n > 2. *Math*. *Proc*. *Cambridge Philos*. *Soc*. **54** 300–301.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, (1934). *Inequalities*. Cambridge Univ. Press, Cambridge.

HÖFFDING, W. (1940). Maszstabinvariante Korrelationstheorie. *Schr*. *Math*. *Inst*. *U*. *Inst*. *Angew*. *Math*. *Univ*. *Berlin* **5** 181–233. MR0004426

ISBELL, J. R. (1955). Birkhoff's problem 111. *Proc*. *Amer*. *Math*. *Soc*. **6** 217–218. MR0068507 https://doi.org/10.2307/2032343

JOE, H. (1990). Multivariate concordance. *J*. *Multivariate Anal*. **35** 12–30. MR1084939 https://doi.org/10.1016/0047-259X(90)90013-8

KNOTT, M. and SMITH, C. (2006). Choosing joint distributions so that the variance of the sum is small. *J*. *Multivariate Anal*. **97** 1757–1765. MR2298887 https://doi.org/10.1016/j.jmva.2006.01.004

LEE, W. and AHN, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance Math*. *Econom*. **56** 68–79. MR3208102 https://doi.org/10.1016/j.insmatheco.2014.03.002

LEE, W., CHEUNG, K. C. and AHN, J. Y. (2017). Multivariate countermonotonicity and the minimal copulas. *J*. *Comput*. *Appl*. *Math*. **317** 589–602. MR3606098 https://doi.org/10.1016/j.cam.2016.12.032

LÉVY, P. (1962). Extensions d'un théorème de D. Dugué et M. Girault. *Z*. *Wahrsch*. *Verw*. *Gebiete* **1** 159–173. MR0145565 https://doi.org/10.1007/BF01844419

LINDENSTRAUSS, J. (1965). A remark on extreme doubly stochastic measures. *Amer*. *Math*. *Monthly* **72** 379–382. MR0181728 https://doi.org/10.2307/2313497

LOH, W.-L. (1996). On Latin hypercube sampling. *Ann*. *Statist*. **24** 2058–2080. MR1421161 https://doi.org/10.1214/aos/1069362310

LU, T.-Y. and YI, Z. (2004). Generalized correlation order and stop-loss order. *Insurance Math*. *Econom*. **35** 69–76. MR2080980 https://doi.org/10.1016/j.insmatheco.2004.04.003

MCKAY, M. D., BECKMAN, R. J. and CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245. MR0533252 https://doi.org/10.2307/1268522

MIRA, A. (2002). Ordering and improving the performance of Monte Carlo Markov chains. *Statist*. *Sci*. **16** 340–350.

MOAMENI, A. (2016). Supports of extremal doubly stochastic measures. *Canad*. *Math*. *Bull*. **59** 381–391. MR3492649 https://doi.org/10.4153/CMB-2015-075-3

MÜLLER, A. and SCARSINI, M. (2000). Some remarks on the supermodular order. *J*. *Multivariate Anal*. **73** 107–119. MR1766123 https://doi.org/10.1006/jmva.1999.1867

NELSEN, R. B. and ÚBEDA-FLORES, M. (2012). Directional dependence in multivariate distributions. *Ann*. *Inst*. *Statist*. *Math*. **64** 677–685. MR2880874 https://doi.org/10.1007/s10463-011-0329-6

OWEN, A. B. (1992). A central limit theorem for Latin hypercube sampling. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **54** 541–551. MR1160481

OWEN, A. B. (2003). The dimension distribution and quadrature test functions. *Statist*. *Sinica* **13** 1–17. MR1963917

PUCCETTI, G. and WANG, R. (2015). Extremal dependence concepts. *Statist*. *Sci*. **30** 485–517. MR3432838 https://doi.org/10.1214/15-STS525

ROACH, W. and WRIGHT, R. (1977). Optimal antithetic sampling plans. *J*. *Stat*. *Comput*. *Simul*. **5** 99–114.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series*, *No*. 28. Princeton Univ. Press, Princeton, NJ. MR0274683

RUBINSTEIN, Y. R. and SAMORODNITSKY, G. (1987). A modified version of Handscomb's antithetic variates theorem. *SIAM J*. *Sci*. *Statist*. *Comput*. **8** 82–98. MR0873925 https://doi.org/10.1137/0908008

RÜSCHENDORF, L. (2013). *Mathematical Risk Analysis*: *Dependence*, *Risk Bounds*, *Optimal Allocations and Portfolios*. *Springer Series in Operations Research and Financial Engineering*. Springer, Heidelberg. MR3051756 https://doi.org/10.1007/978-3-642-33590-7

RÜSCHENDORF, L. and UCKELMANN, L. (2002). Variance minimization and random variables with constant sum. In *Distributions with Given Marginals and Statistical Modelling* 211–222. Kluwer Academic, Dordrecht. MR2058994

SHEPP, L. A. (1962). Symmetric random walk. *Trans*. *Amer*. *Math*. *Soc*. **104** 144–153. MR0139212 https://doi.org/10.2307/1993938

STEIN, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29** 143–151. MR0887702 https://doi.org/10.2307/1269769

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J*. *Amer*. *Statist*. *Assoc*. **88** 1392–1397. MR1245375

TUKEY, J. W. (1957). Antithesis or regression? *Proc*. *Camb*. *Philos*. *Soc*. **53** 923–924. MR0091586

VAN DYK, D. A. and MENG, X. L. (2001). The art of data augmentation. *J*. *Comput*. *Graph*. *Statist*. **10** 1–50.

VITALE, R. A. (1990). On stochastic dependence and a class of degenerate distributions. *Top*. *Stat*. *Depend*. **16** 459.

VON NEUMANN, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. In *Contributions to the Theory of Games*, *Vol*. 2. *Annals of Mathematics Studies*, *No*. 28 5–12. Princeton Univ. Press, Princeton, NJ. MR0054920

WANG, B. and WANG, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal.* **102** 1344–1360. MR2819953 https://doi.org/10.1016/j.jmva.2011.05.002

WANG, X. and FANG, K.-T. (2003). The effective dimension and quasi-Monte Carlo integration. *J. Complexity* **19** 101–124. MR1966664 https://doi.org/10.1016/S0885-064X(03)00003-7

WANG, X. and SLOAN, I. H. (2005). Why are high-dimensional finance problems often of low effective dimension? *SIAM J. Sci. Comput.* **27** 159–183. MR2201179 https://doi.org/10.1137/S1064827503429429

WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4** 1280–1289. MR0426099

WILSON, J. R. (1979). Proof of the antithetic-variates theorem for unbounded functions. *Math. Proc. Cambridge Philos. Soc.* **86** 477–479. MR0542694 https://doi.org/10.1017/S0305004100056334

WILSON, J. R. (1983). Antithetic sampling with multivariate inputs. *Amer. J. Math. Management Sci.* **3** 121–144. MR0723937 https://doi.org/10.1080/01966324.1983.10737119

Statistical Science
2024, Vol. 39, No. 1, 137–164
https://doi.org/10.1214/23-STS893

Sampling Algorithms in Statistical Physics: A Guide for Statistics and Machine Learning

Michael F. Faulkner and Samuel Livingstone

Abstract. We discuss several algorithms for sampling from unnormalized probability distributions in statistical physics, but using the language of statistics and machine learning. We provide a self-contained introduction to some key ideas and concepts of the field, before discussing three well-known problems: phase transitions in the Ising model, the melting transition on a two-dimensional plane and simulation of an all-atom model for liquid water. We review the classical Metropolis, Glauber and molecular dynamics sampling algorithms before discussing several more recent approaches, including cluster algorithms, novel variations of hybrid Monte Carlo and Langevin dynamics and piece-wise deterministic processes such as event chain Monte Carlo. We highlight cross-over with statistics and machine learning throughout and present some results on event chain Monte Carlo and sampling from the Ising model using tools from the statistics literature. We provide a simulation study on the Ising and XY models, with reproducible code freely available online, and following this we discuss several open areas for interaction between the disciplines that have not yet been explored and suggest avenues for doing so.

Key words and phrases: Statistical physics, sampling algorithms, Markov chain Monte Carlo, Ising model, Potts model, XY model, hard-disk model, molecular simulation, Metropolis, Glauber dynamics, molecular dynamics, hybrid Monte Carlo, Langevin dynamics, event chain Monte Carlo.

REFERENCES

ALDER, B. J. and WAINWRIGHT, T. E. (1957). Phase transition for a hard sphere system. *J. Chem. Phys.* **27** 1208. https://doi.org/10.1063/1.1743957

ALDER, B. J. and WAINWRIGHT, T. E. (1959). Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **31** 459–466. MR0115313 https://doi.org/10.1063/1.1730376

ALDER, B. J. and WAINWRIGHT, T. E. (1960). Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. *J. Chem. Phys.* **33** 1439–1451. MR0115314 https://doi.org/10.1063/1.1731425

ALDER, B. J. and WAINWRIGHT, T. E. (1962). Phase transition in elastic disks. *Phys. Rev.* **127** 359. https://doi.org/10.1103/PhysRev.127.359

ANDRECUT, M. (2009). Molecular dynamics multidimensional scaling. *Phys. Lett. A* **373** 2001–2006.

ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2021). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31** 2478–2517. MR4332703 https://doi.org/10.1214/20-aap1653

ANDRIEU, C., LEE, A. and LIVINGSTONE, S. (2020). A general perspective on the Metropolis–Hastings kernel. Preprint. Available at arXiv:2012.14881.

ANDRIEU, C. and LIVINGSTONE, S. (2021). Peskun–Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario. *Ann. Statist.* **49** 1958–1981. MR4319237 https://doi.org/10.1214/20-aos2008

ARCHAMBAULT, P., BRAMWELL, S. T. and HOLDSWORTH, P. C. W. (1997). Magnetic fluctuations in a finite two-dimensional XY model. *J. Phys. A* **30** 8363.

ARTEMOVA, S. and REDON, S. (2012). Adaptively restrained particle simulations. *Phys. Rev. Lett.* **109** 190201. https://doi.org/10.1103/PhysRevLett.109.190201

BARDUCCI, A., BONOMI, M. and PARRINELLO, M. (2011). Metadynamics. *WIREs, Comput. Mol. Sci.* **1** 826. https://doi.org/10.1002/wcms.31

BARKER, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton–electron plasma. *Aust. J. Phys.* **18** 119–134.

BAXTER, R. J. (2008). *Exactly Solved Models in Statistical Mechanics*. Dover Publications Inc., New York.

BEREZINSKII, V. L. (1971). Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. Classical systems. *Sov. Phys.—JETP* **32** 493.

Michael F. Faulkner is EPSRC Postdoctoral Fellow, HH Wills Physics Laboratory, University of Bristol, Bristol, UK (e-mail: michael.faulkner@bristol.ac.uk). Samuel Livingstone is Associate Professor, Department of Statistical Science, University College London, London, UK (e-mail: samuel.livingstone@ucl.ac.uk).

BERG, B. A. and NEUHAUS, T. (1991). Multicanonical algorithms for first order phase transitions. *Phys. Lett. B* **267** 249. https://doi.org/10.1016/0370-2693(91)91256-U

BERNARD, E. P. (2011). Algorithms and applications of the Monte Carlo method: Two-dimensional melting and perfect sampling. Ph.D. thesis, Univ. Pierre-et-Marie-Curie—Paris VI.

BERNARD, E. P. and KRAUTH, W. (2011). Two-step melting in two dimensions: First-order liquid-hexatic transition. *Phys. Rev. Lett.* **107** 155704. https://doi.org/10.1103/PhysRevLett.107.155704

BERNARD, E. P., KRAUTH, W. and WILSON, D. B. (2009). Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E* **80** 056704. https://doi.org/10.1103/PhysRevE.80.056704

BETANCOURT, M., BYRNE, S., LIVINGSTONE, S. and GIROLAMI, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli* **23** 2257–2298. MR3648031 https://doi.org/10.3150/16-BEJ810

BETANCOURT, M. and GIROLAMI, M. (2015). Hamiltonian Monte Carlo for hierarchical models. In *Current Trends in Bayesian Methodology with Applications* 79–101. CRC Press, Boca Raton, FL. MR3644666

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

BIERKENS, J., KAMATANI, K. and ROBERTS, G. O. (2022). High-dimensional scaling limits of piecewise deterministic sampling algorithms. *Ann. Appl. Probab.* **32** 3361–3407. MR4497848 https://doi.org/10.1214/21-aap1762

BIERKENS, J., ROBERTS, G. O. and ZITT, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1103–1130. MR3557191 https://doi.org/10.1111/rssb.12158

BOU-RABEE, N. and SANZ-SERNA, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.* **27** 113–206. MR3826507 https://doi.org/10.1017/s0962492917000101

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

BRAMWELL, S. T. and HOLDSWORTH, P. C. W. (1993). Magnetization and universal sub-critical behaviour in two-dimensional XY magnets. *J. Phys., Condens. Matter* **5** L53.

BRAMWELL, S. T., HOLDSWORTH, P. C. W. and PINTON, J. F. (1998). Universality of rare fluctuations in turbulence and critical phenomena. *Nature* **396** 552. https://doi.org/10.1038/25083

CAMPOS, C. M. and SANZ-SERNA, J. M. (2015). Extra chance generalized hybrid Monte Carlo. *J. Comput. Phys.* **281** 365–374. MR3281978 https://doi.org/10.1016/j.jcp.2014.09.037

CANCÈS, E., LEGOLL, F. and STOLTZ, G. (2007). Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM Math. Model. Numer. Anal.* **41** 351–389. MR2339633 https://doi.org/10.1051/m2an:2007014

CORBELLA, A., SPENCER, S. E. F. and ROBERTS, G. O. (2022). Automatic Zig-Zag sampling in practice. *Stat. Comput.* **32** Paper No. 107, 16 pp. MR4507158 https://doi.org/10.1007/s11222-022-10142-x

COULOMB, C. A. (1785). Second mémoire sur l'électricité et le magnétisme. *Hist. Acad. R. Sci.* 578.

DALALYAN, A. S. and RIOU-DURAND, L. (2020). On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli* **26** 1956–1988. MR4091098 https://doi.org/10.3150/19-BEJ1178

DARVE, E., RODRÍGUEZ-GÓMEZ, D. and POHORILLE, A. (2008). Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **128** 144120.

DE LEEUW, S. W., PERRAM, J. W. and SMITH, E. R. (1980). Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. Lond. Ser. A* **373** 27–56. MR0592747 https://doi.org/10.1098/rspa.1980.0135

DELIGIANNIDIS, G., PAULIN, D., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2021). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *Ann. Appl. Probab.* **31** 2612–2662. MR4350970 https://doi.org/10.1214/20-aap1659

DRESS, C. and KRAUTH, W. (1995). Cluster algorithm for hard spheres and related systems. *J. Phys. A* **28** L597–L601. MR1381129

DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

DURMUS, A., ENFROY, A., MOULINES, É. and STOLTZ, G. (2021). Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics. Preprint. Available at arXiv:2107.14542.

EDWARDS, R. G. and SOKAL, A. D. (1988). Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Phys. Rev. D* (3) **38** 2009–2012. MR0965465 https://doi.org/10.1103/PhysRevD.38.2009

ENGEL, M., ANDERSON, J. A., GLOTZER, S. C., ISOBE, M., BERNARD, E. P. and KRAUTH, W. (2013). Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods. *Phys. Rev. E* **87** 042134. https://doi.org/10.1103/PhysRevE.87.042134

EVANS, D. J. and HOLIAN, B. L. (1985). The Nosé–Hoover thermostat. *J. Chem. Phys.* **83** 4069–4074.

FAULKNER, M. F. (2022). General symmetry breaking at a topological phase transition. Preprint. Available at arXiv:2209.03699.

FAULKNER, M. F., BRAMWELL, S. T. and HOLDSWORTH, P. C. W. (2015). Topological-sector fluctuations and ergodicity breaking at the Berezinskii–Kosterlitz–Thouless transition. *Phys. Rev. B* **91** 155412. https://doi.org/10.1103/PhysRevB.91.155412

FAULKNER, M. F. and LIVINGSTONE, S. (2024). Supplement to "Sampling algorithms in statistical physics: A guide for statistics and machine learning." https://doi.org/10.1214/23-STS893SUPP

FAULKNER, M. F., QIN, L., MAGGS, A. C. and KRAUTH, W. (2018). All-atom computations with irreversible Markov chains. *J. Chem. Phys.* **149** 064113. https://doi.org/10.1063/1.5036638

FEARNHEAD, P., BIERKENS, J., POLLOCK, M. and ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. MR3843382 https://doi.org/10.1214/18-STS648

FRANKS, J. and VIHOLA, M. (2020). Importance sampling correction versus standard averages of reversible MCMCs in terms of the asymptotic variance. *Stochastic Process. Appl.* **130** 6157–6183. MR4140030 https://doi.org/10.1016/j.spa.2020.05.006

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741. https://doi.org/10.1109/tpami.1984.4767596

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

GLAUBER, R. J. (1963). Time-dependent statistics of the Ising model. *J. Math. Phys.* **4** 294–307. MR0148410 https://doi.org/10.1063/1.1703954

GORE, V. K. and JERRUM, M. R. (1999). The Swendsen–Wang process does not always mix rapidly. *J. Stat. Phys.* **97** 67–86. MR1733467 https://doi.org/10.1023/A:1004610900745

GRÜNWALD, P. and VAN OMMEN, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* **12** 1069–1103. MR3724979 https://doi.org/10.1214/17-BA1085

HADDOUCHE, M., GUEDJ, B., RIVASPLATA, O. and SHAWE-TAYLOR, J. (2021). PAC-Bayes unleashed: Generalisation bounds with unbounded losses. *Entropy* **23** Paper No. 1330, 20 pp. MR4330257 https://doi.org/10.3390/e23101330

HAIRER, E., LUBICH, C. and WANNER, G. (2003). Geometric numerical integration illustrated by the Störmer–Verlet method. *Acta Numer.* **12** 399–450. MR2249159 https://doi.org/10.1017/S0962492902000144

HAIRER, E., LUBICH, C. and WANNER, G. (2006). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed. *Springer Series in Computational Mathematics* **31**. Springer, Berlin. MR2221614

HALPERIN, B. I. and NELSON, D. R. (1978). Theory of two-dimensional melting. *Phys. Rev. Lett.* **41** 121–124. MR0495791 https://doi.org/10.1103/PhysRevLett.41.121

HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

HOHENBERG, P. C. (1967). Existence of long-range order in one and two dimensions. *Phys. Rev.* **158** 383.

HÖLLMER, P., MAGGS, A. C. and KRAUTH, W. (2022). Hard-disk dipoles and non-reversible Markov chains. *J. Chem. Phys.* **156** 084108. https://doi.org/10.1063/5.0080101

HÖLLMER, P., QIN, L., FAULKNER, M. F., MAGGS, A. C. and KRAUTH, W. (2020). JeLLyFysh-Version1.0—A Python application for all-atom event-chain Monte Carlo. *Comput. Phys. Commun.* **253** 107168, 20 pp. MR4105638 https://doi.org/10.1016/j.cpc.2020.107168

HOROWITZ, A. M. (1991). A generalized guided Monte Carlo algorithm. *Phys. Lett. B* **268** 247–252.

HUBER, M. (2003). A bounding chain for Swendsen–Wang. *Random Structures Algorithms* **22** 43–59. MR1943858 https://doi.org/10.1002/rsa.10071

ISING, E. (1925). Beitrag zur theorie des ferromagnetismus. *Z. Phys.* **31** 253. https://doi.org/10.1007/BF02980577

IZAGUIRRE, J. A. and HAMPTON, S. S. (2004). Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules. *J. Comput. Phys.* **200** 581–604.

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

JASRA, A., HOLMES, C. C. and STEPHENS, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.* **20** 50–67. MR2182987 https://doi.org/10.1214/088342305000000016

JASTER, A. (1999a). Computer simulations of the two-dimensional melting transition using hard disks. *Phys. Rev. E* **59** 2594.

JASTER, A. (1999b). An improved Metropolis algorithm for hard core systems. *Phys. A, Stat. Mech. Appl.* **264** 134–141.

JEWSON, J., SMITH, J. Q. and HOLMES, C. (2018). Principles of Bayesian inference using general divergence criteria. *Entropy* **20** Paper No. 442, 25 pp. MR3879894 https://doi.org/10.3390/e20060442

JOSÉ, J. V., KADANOFF, L. P., KIRKPATRICK, S. and NELSON, D. R. (1977). Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model. *Phys. Rev. B* **16** 1217.

JOSÉ, J. V., KOSTERLITZ, J. M., THOULESS, D. J., ORTIZ, G., COBANERA, E., NUSSINOV, Z., GOLDMAN, A. M., BENFATTO, L., CASTELLANI, C., GIAMARCHI, T., TEITEL, S., FAZIO, R., SCHÖN, G., VINOKUR, V. M., BATURINA, T. I., HADZIBABIC, Z., DALIBARD, J., FERTIG, H. A. and MURTHY, G. (2013). 40 *Years of Berezinskii–Kosterlitz–Thouless Theory*. World Scientific, Singapore. https://doi.org/10.1142/8572

KAISER, V., BRAMWELL, S. T., HOLDSWORTH, P. C. W. and MOESSNER, R. (2013). Onsager's Wien effect on a lattice. *Nat. Mater.* **12** 1033–1037. https://doi.org/10.1038/nmat3729

KAISER, V., BRAMWELL, S. T., HOLDSWORTH, P. C. W. and MOESSNER, R. (2015). ac Wien effect in spin ice, manifest in nonlinear, nonequilibrium susceptibility. *Phys. Rev. Lett.* **115** 037201, 6 pp. MR3466024 https://doi.org/10.1103/PhysRevLett.115.037201

KAMPMANN, T. A., BOLTZ, H.-H. and KIERFELD, J. (2015a). Monte Carlo simulation of dense polymer melts using event chain algorithms. *J. Chem. Phys.* **143** 044105. https://doi.org/10.1063/1.4927084

KAMPMANN, T. A., BOLTZ, H.-H. and KIERFELD, J. (2015b). Parallelized event chain algorithm for dense hard sphere and polymer systems. *J. Comput. Phys.* **281** 864–875. MR3281999 https://doi.org/10.1016/j.jcp.2014.10.059

KAPFER, S. C. and KRAUTH, W. (2015). Two-dimensional melting: From liquid-hexatic coexistence to continuous transitions. *Phys. Rev. Lett.* **114** 035702. https://doi.org/10.1103/PhysRevLett.114.035702

KAPFER, S. C. and KRAUTH, W. (2016). Cell-veto Monte Carlo algorithm for long-range systems. *Phys. Rev. E* **94** 031302. https://doi.org/10.1103/PhysRevE.94.031302

KLEMENTA, M. and ENGEL, M. (2019). Efficient equilibration of hard spheres with Newtonian event chains. *J. Chem. Phys.* **150** 174108. https://doi.org/10.1063/1.5090882

KNOBLAUCH, J., JEWSON, J. and DAMOULAS, T. (2022). An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *J. Mach. Learn. Res.* **23** Paper No. [132], 109 pp. MR4577084

KOSTERLITZ, J. M. (1974). The critical properties of the two-dimensional XY model. *J. Phys. C, Solid State Phys.* **7** 1046.

KOSTERLITZ, J. M. and THOULESS, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C, Solid State Phys.* **6** 1181.

KRAMERS, H. A. and WANNIER, G. H. (1941a). Statistics of the two-dimensional ferromagnet. I. *Phys. Rev.* (2) **60** 252–262. MR0004803

KRAMERS, H. A. and WANNIER, G. H. (1941b). Statistics of the two-dimensional ferromagnet. II. *Phys. Rev.* (2) **60** 263–276. MR0004804

KRAUTH, W. (2006). *Statistical Mechanics: Algorithms and Computations*. *Oxford Master Series in Physics* **13**. Oxford Univ. Press, Oxford. MR2370557

ŁATUSZYŃSKI, K. and ROBERTS, G. O. (2013). CLTs and asymptotic variance of time-sampled Markov chains. *Methodol. Comput. Appl. Probab.* **15** 237–247. MR3030220 https://doi.org/10.1007/s11009-011-9237-8

LECUN, Y., CHOPRA, S., HADSELL, R., RANZATO, M. and HUANG, F. (2006). A tutorial on energy-based learning. *Predict. Struct. Data* **1**.

LEI, Z. and KRAUTH, W. (2018). Irreversible Markov chains in spin models: Topological excitations. *Europhys. Lett.* **121** 10008. https://doi.org/10.1209/0295-5075/121/10008

Leimkuhler, B. and Matthews, C. (2016). *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. *Interdisciplinary Applied Mathematics* **39**. Springer, Cham. MR3362507

Leimkuhler, B., Noorizadeh, E. and Theil, F. (2009). A gentle stochastic thermostat for molecular dynamics. *J. Stat. Phys.* **135** 261–277. MR2505736 https://doi.org/10.1007/s10955-009-9734-0

Leimkuhler, B. and Reich, S. (2004). *Simulating Hamiltonian Dynamics*. *Cambridge Monographs on Applied and Computational Mathematics* **14**. Cambridge Univ. Press, Cambridge. MR2132573

Lelièvre, T., Rousset, M. and Stoltz, G. (2010). *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, London. MR2681239 https://doi.org/10.1142/9781848162488

Livingstone, S., Faulkner, M. F. and Roberts, G. O. (2019). Kinetic energy choice in Hamiltonian/hybrid Monte Carlo. *Biometrika* **106** 303–319. MR3949305 https://doi.org/10.1093/biomet/asz013

Lu, X., Perrone, V., Hasenclever, L., Teh, Y. W. and Vollmer, S. (2017). Relativistic Monte Carlo. In *Artificial Intelligence and Statistics* 1236–1245. PMLR.

Martyna, G. J., Klein, M. L. and Tuckerman, M. (1992). Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97** 2635–2643.

Mattingly, J. C., Stuart, A. M. and Higham, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. MR1931266 https://doi.org/10.1016/S0304-4149(02)00150-3

Mermin, N. D. and Wagner, H. (1966). Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.* **17** 1133.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087. https://doi.org/10.1063/1.1699114

Michel, M., Durmus, A. and Sénécal, S. (2020). Forward event-chain Monte Carlo: Fast sampling by randomness control in irreversible Markov chains. *J. Comput. Graph. Statist.* **29** 689–702. MR4191236 https://doi.org/10.1080/10618600.2020.1750417

Michel, M., Kapfer, S. C. and Krauth, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.* **140** 054116. https://doi.org/10.1063/1.4863991

Michel, M., Mayer, J. and Krauth, W. (2015). Event-chain Monte Carlo for classical continuous spin models. *Europhys. Lett.* **112** 20003.

Mira, A. (2001a). On Metropolis–Hastings algorithms with delayed rejection. *Metron* **59** 231–241. MR1889712

Mira, A. (2001b). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16** 340–350. MR1888449 https://doi.org/10.1214/ss/1015346319

Monemvassitis, A., Guillin, A. and Michel, M. (2023). PDMP characterisation of event-chain Monte Carlo algorithms for particle systems. *J. Stat. Phys.* **190** Paper No. 66, 29 pp. MR4549874 https://doi.org/10.1007/s10955-023-03069-8

Neal, R. M. (1993). Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems* 475–482.

Neal, R. M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132 https://doi.org/10.1023/A:1008923215028

Neal, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

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

Neal, R. M. (2012). *Bayesian Learning for Neural Networks* **118**. Springer, Berlin.

Nishimura, A., Dunson, D. B. and Lu, J. (2020). Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods. *Biometrika* **107** 365–380. MR4126286 https://doi.org/10.1093/biomet/asz083

Nott, D. J. and Green, P. J. (2004). Bayesian variable selection and the Swendsen–Wang algorithm. *J. Comput. Graph. Statist.* **13** 141–157. MR2044875 https://doi.org/10.1198/1061860042958

Onsager, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* (2) **65** 117–149. MR0010315

Onsager, L. (1949). Discussion (comment on the spontaneous magnetisation of the two-dimensional Ising model). *Nuovo Cimento* **6** 251. https://doi.org/10.1007/BF02780989

Palmer, R. G. (1982). Broken ergodicity. *Adv. Phys.* **31** 669.

Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2007). A general framework for the parametrization of hierarchical models. *Statist. Sci.* **22** 59–73. MR2408661 https://doi.org/10.1214/088342307000000014

Park, J. and Atchadé, Y. (2020). Markov chain Monte Carlo algorithms with sequential proposals. *Stat. Comput.* **30** 1325–1345. MR4137254 https://doi.org/10.1007/s11222-020-09948-4

Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*. *Texts in Applied Mathematics* **60**. Springer, New York. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

Peng, Y., Wang, Z., Alsayed, A. M., Yodh, A. G. and Han, Y. (2010). Melting of colloidal crystal films. *Phys. Rev. Lett.* **104** 205703. https://doi.org/10.1103/PhysRevLett.104.205703

Peters, E. A. J. F. and de With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85** 026703. https://doi.org/10.1103/PhysRevE.85.026703

Potts, R. B. (1952). Some generalized order-disorder transformations. *Proc. Camb. Philos. Soc.* **48** 106–109. MR0047571 https://doi.org/10.1017/S0305004100027419

Power, S. and Goldman, J. V. (2019). Accelerated sampling on discrete spaces with non-reversible Markov processes. Preprint. Available at arXiv:1912.04681.

Radivojević, T. and Akhmatskaya, E. (2020). Modified Hamiltonian Monte Carlo for Bayesian inference. *Stat. Comput.* **30** 377–404. MR4064627 https://doi.org/10.1007/s11222-019-09885-x

Radivojević, T., Fernández-Pendás, M., Sanz-Serna, J. M. and Akhmatskaya, E. (2018). Multi-stage splitting integrators for sampling with modified Hamiltonian Monte Carlo methods. *J. Comput. Phys.* **373** 900–916. MR3854175 https://doi.org/10.1016/j.jcp.2018.07.023

Redon, S., Stoltz, G. and Trstanova, Z. (2016). Error analysis of modified Langevin dynamics. *J. Stat. Phys.* **164** 735–771. MR3529154 https://doi.org/10.1007/s10955-016-1544-6

Ren, B., Bacallado, S., Favaro, S., Holmes, S. and Trippa, L. (2017). Bayesian nonparametric ordination for the analysis of microbial communities. *J. Amer. Statist. Assoc.* **112** 1430–1442. MR3750866 https://doi.org/10.1080/01621459.2017.1288631

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

Roscilde, T., Faulkner, M. F., Bramwell, S. T. and Holdsworth, P. C. W. (2016). From quantum to thermal topological-sector fluctuations of strongly interacting bosons in a ring lattice. *New J. Phys.* **18** 075003.

ROSENBLUTH, M. N. and ROSENBLUTH, A. W. (1954). Further results on Monte Carlo equations of state. *J. Chem. Phys.* **22** 881. https://doi.org/10.1063/1.1740207

ROSSKY, P. J., DOLL, J. and FRIEDMAN, H. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.* **69** 4628–4633.

SALZBERG, A. M. and PRAGER, S. (1963). Equation of state for a two-dimensional electrolyte. *J. Chem. Phys.* **38** 2587.

SHERLOCK, C., URBAS, S. and LUDKIN, M. (2021). Apogee to Apogee path sampler. Preprint. Available at arXiv:2112.08187.

SIMÁNYI, N. (2003). Proof of the Boltzmann–Sinai ergodic hypothesis for typical hard disk systems. *Invent. Math.* **154** 123–178. MR2004458 https://doi.org/10.1007/s00222-003-0304-9

SINAI, Y. G. (1970). Dynamical systems with elastic reflections. *Russian Math. Surveys* **25** 137.

SINGH, K. K., MARIN, D. F. and REDON, S. (2017). Parallel adaptively restrained molecular dynamics. In 2017 *International Conference on High Performance Computing & Simulation* (*HPCS*) 308–314. IEEE, New York.

SMITH, M. and KOHN, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75** 317–343.

STOLTZ, G. (2021). Computational statistical physics and hypocoercivity. Preprint. Available at arXiv:2112.08221.

STOLTZ, G. and TRSTANOVA, Z. (2018). Langevin dynamics with general kinetic energies. *Multiscale Model. Simul.* **16** 777–806. MR3799045 https://doi.org/10.1137/16M110575X

STORATH, M., WEINMANN, A., FRIKEL, J. and UNSER, M. (2015). Joint image reconstruction and segmentation using the Potts model. *Inverse Probl.* **31** 025003, 29 pp. MR3303168 https://doi.org/10.1088/0266-5611/31/2/025003

STÖRMER, C. (1907). Sur les trajectoires des corpuscules électrisés dans l'espace. Applications à l'aurore boréale et aux perturbations magnétiques. *Radium* **4** 2–5.

SWENDSEN, R. H. and WANG, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58** 86.

SYRING, N. and MARTIN, R. (2020). Gibbs posterior concentration rates under sub-exponential type losses. Preprint. Available at arXiv:2012.04505.

TABACHNIKOV, S. (2005). *Geometry and Billiards. Student Mathematical Library* **30**. Amer. Math. Soc., Providence, RI. MR2168892 https://doi.org/10.1090/stml/030

TALAY, D. (2002). Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields* **8** 163–198. MR1924934

TAMAYO, P., BROWER, R. C. and KLEIN, W. (1990). Single-cluster Monte Carlo dynamics for the Ising model. *J. Stat. Phys.* **58** 1083–1094. MR1049057 https://doi.org/10.1007/BF01026564

THORNEYWORK, A. L., ABBOTT, J. L., AARTS, D. G. A. L. and DULLENS, R. P. A. (2017). Two-dimensional melting of colloidal hard spheres. *Phys. Rev. Lett.* **118** 158001. https://doi.org/10.1103/PhysRevLett.118.158001

TORGERSON, W. S. (1952). Multidimensional scaling. I. Theory and method. *Psychometrika* **17** 401–419. MR0054219 https://doi.org/10.1007/BF02288916

TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D, Nonlinear Phenom.* **240** 410–414.

VALLAT, A. and BECK, H. (1994). Coulomb-gas representation of the two-dimensional XY model on a torus. *Phys. Rev. B* **50** 4015.

VERLET, L. (1967). Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard–Jones molecules. *Phys. Rev.* **159** 98.

VIHOLA, M., HELSKE, J. and FRANKS, J. (2020). Importance sampling type estimators based on approximate marginal Markov chain Monte Carlo. *Scand. J. Stat.* **47** 1339–1376. MR4178196 https://doi.org/10.1111/sjos.12492

VUCELJA, M. (2016). Lifting—A nonreversible Markov chain Monte Carlo algorithm. *Am. J. Phys.* **84** 958. https://doi.org/10.1119/1.4961596

WANG, F. and LANDAU, D. P. (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86** 2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050

WEI, D., SONG, Y. and WANG, F. (2011). A simple molecular mechanics potential for μm scale graphene simulations from the adaptive force matching method. *J. Chem. Phys.* **134** 184704.

WOLFF, U. (1989a). Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.* **62** 361.

WOLFF, U. (1989b). Comparison between cluster Monte Carlo algorithms in the Ising model. *Phys. Lett. B* **228** 379–382.

WOOD, W. W. and PARKER, F. R. (1957). Monte Carlo equation of state of molecules interacting with the Lennard–Jones potential. I. A supercritical isotherm at about twice the critical temperature. *J. Chem. Phys.* **27** 720. https://doi.org/10.1063/1.1743822

WORRALL, J., BROWNING, R., WU, P. and MENGERSEN, K. (2022). Fifty years later: New directions in Hawkes processes. *SORT* **46** 3–37. MR4443331

WU, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.* **91** 205–238. MR1807683 https://doi.org/10.1016/S0304-4149(00)00061-2

WU, P.-S. and MARTIN, R. (2023). A comparison of learning rate selection methods in generalized Bayesian inference. *Bayesian Anal.* **18** 105–132. MR4515727 https://doi.org/10.1214/21-ba1302

WU, Y., TEPPER, H. L. and VOTH, G. A. (2006). Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **124** 024503. https://doi.org/10.1063/1.2136877

YANG, C. N. (1952). The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.* (2) **85** 808–816. MR0051740 https://doi.org/10.1103/PhysRev.85.808

YOUNG, A. P. (1979). Melting and the vector Coulomb gas in two dimensions. *Phys. Rev. B* **19** 1855. https://doi.org/10.1103/PhysRevB.19.1855

ZANELLA, G. (2020). Informed proposals for local MCMC in discrete spaces. *J. Amer. Statist. Assoc.* **115** 852–865. MR4107684 https://doi.org/10.1080/01621459.2019.1585255

ZHANG, Y., WANG, X., CHEN, C., HENAO, R., FAN, K. and CARIN, L. (2016). Towards unifying Hamiltonian Monte Carlo and slice sampling. *Adv. Neural Inf. Process. Syst.* **29**.

Statistical Science
2024, Vol. 39, No. 1, 165–191
https://doi.org/10.1214/23-STS889

Causal Inference Methods for Combining Randomized Trials and Observational Studies: A Review

Bénédicte Colnet, Imke Mayer, Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse and Shu Yang

Abstract. With increasing data availability, causal effects can be evaluated across different data sets, both randomized controlled trials (RCTs) and observational studies. RCTs isolate the effect of the treatment from that of unwanted (confounding) co-occurring effects but they may suffer from unrepresentativeness, and thus lack external validity. On the other hand, large observational samples are often more representative of the target population but can conflate confounding effects with the treatment of interest. In this paper, we review the growing literature on methods for causal inference on combined RCTs and observational studies, striving for the best of both worlds. We first discuss identification and estimation methods that improve generalizability of RCTs using the representativeness of observational data. Classical estimators include weighting, difference between conditional outcome models and doubly robust estimators. We then discuss methods that combine RCTs and observational data to either ensure unconfoundedness of the observational analysis or to improve (conditional) average treatment effect estimation. We also connect and contrast works developed in both the potential outcomes literature and the structural causal model literature. Finally, we compare the main methods using a simulation study and real world data to analyze the effect of tranexamic acid on the mortality rate in major trauma patients. A review of available codes and new implementations is also provided.

Key words and phrases: Causal effect generalization, transportability, double robustness, data integration, heterogeneous data, S-admissibility.

REFERENCES

ACKERMAN, B., LESKO, C. R., SIDDIQUE, J., SUSUKIDA, R. and STUART, E. A. (2021). Generalizing randomized trial findings to a target population using complex survey population data. *Stat. Med.* **40** 1101–1120. MR4384364 https://doi.org/10.1002/sim.8822

ANDREWS, I. and OSTER, E. (2019). A simple approximation for evaluating external validity bias. *Econom. Lett.* **178** 58–62. MR3921042 https://doi.org/10.1016/j.econlet.2019.02.020

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

ATHEY, S., CHETTY, R. and IMBENS, G. (2020). Combining experimental and observational data to estimate treatment effects on long term outcomes. Preprint. Available at arXiv:2006.09676.

ATHEY, S., CHETTY, R., IMBENS, G. and KANG, H. (2020). Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index.

ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963

Bénédicte Colnet is Ph.D. candidate, Soda project-team, INRIA Saclay, Palaiseau, France (e-mail: benedicte.colnet@inria.fr). Imke Mayer is Research Scientist, Owkin, London, UK (e-mail: Imke.mayer@owkin.com). Guanhua Chen is Associate Professor, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53726, USA (e-mail: gchen25@wisc.edu). Awa Dieng is Research Associate, Google DeepMind, Montreal, Canada (e-mail: awadieng@google.com). Ruohong Li is Senior Machine Learning Scientist, Microsoft, Kirkland, WA 98033 , USA (e-mail: hannahli@microsoft.com). Gaël Varoquaux is Research Director, Soda project-team, INRIA Saclay, Paries, France (e-mail: gael.varoquaux@inria.fr). Jean-Philippe Vert is Chief R&D Officer, Owkin, Paris, France (e-mail: jean-philippe.vert@owkin.com). Julie Josse is Head, Premedical project Inria team, University of Montpellier, France (e-mail: julie.josse@inria.fr). Shu Yang is Associate Professor, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695, USA (e-mail: syang24@ncsu.edu).

https://doi.org/10.1214/18-AOS1709

BACH, P., CHERNOZHUKOV, V., KURZ, M. S. and SPINDLER, M. (2021). DoubleML—An object-oriented implementation of double machine learning in R. Available at arXiv:2103.09603 [stat.ML].

BACH, P., CHERNOZHUKOV, V., KURZ, M. S. and SPINDLER, M. (2022). DoubleML—An object-oriented implementation of double machine learning in Python. *J. Mach. Learn. Res.* **23** 1–6.

BAREINBOIM, E., LEE, S., HONAVAR, V. and PEARL, J. (2013). Transportability from multiple environments with limited experiments.

BAREINBOIM, E. and PEARL, J. (2012a). Controlling selection bias in causal inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (N. D. Lawrence and M. Girolami, eds.). *Proceedings of Machine Learning Research* **22** 100–108. La Palma, Canary Islands. PMLR.

BAREINBOIM, E. and PEARL, J. (2012b). Transportability of causal effects: Completeness results. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, *AAAI'*12 698–704. AAAI Press, Menlo Park.

BAREINBOIM, E. and PEARL, J. (2013). A general algorithm for deciding transportability of experimental results. *J. Causal Inference* **1** 107–133. MR4289403 https://doi.org/10.1515/jci-2012-0004

BAREINBOIM, E. and PEARL, J. (2016). Causal inference and the data-fusion problem. *Proc. Natl. Acad. Sci. USA* **113** 7345–7352.

BAREINBOIM, E., TIAN, J. and PEARL, J. (2014). Recovering from selection bias in causal and statistical inference. In *Proceedings of the AAAI Conference on Artificial Intelligence* **28**.

BEGG, C. B. and LEUNG, D. H. Y. (2000). On the use of surrogate end points in randomized trials. *J. Roy. Statist. Soc. Ser. A* **163** 15–28.

BUCHANAN, A. L., HUDGENS, M. G., COLE, S. R., MOLLAN, K. R., SAX, P. E., DAAR, E. S., ADIMORA, A. A., ERON, J. J. and MUGAVERO, M. J. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. *J. Roy. Statist. Soc. Ser. A* **181** 1193–1209. MR3876388 https://doi.org/10.1111/rssa.12357

CAMPBELL, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychol. Bull.* **54** 297–312.

CAP, A. P. (2019). Crash-3: A win for patients with traumatic brain injury. *Lancet* **394** 1687–1688.

CARPENTER, J. R. and KENWARD, M. G. (2007). Missing data in randomised controlled trials: a practical guide.

CHEN, R., CHEN, G. and YU, M. (2023). A generalizability score for aggregate causal effect. *Biostatistics* **24** 309–326. MR4597229 https://doi.org/10.1093/biostatistics/kxab029

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

CHU, J., LU, W. and YANG, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*. https://doi.org/10.1093/biomet/asad020

CINELLI, C. and PEARL, J. (2021). Generalizing experimental results by leveraging knowledge of mechanisms. *Eur. J. Epidemiol.* **36** 149–164. https://doi.org/10.1007/s10654-020-00687-4

COCHRAN, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24** 295–313. MR0228136 https://doi.org/10.2307/2528036

COLE, S. R. and STUART, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Amer. J. Epidemiol.* **172** 107–115. https://doi.org/10.1093/aje/kwq084

COLNET, B., JOSSE, J., VAROQUAUX, G. and SCORNET, E. (2022a). Causal effect on a target population: A sensitivity analysis to handle missing covariates. *J. Causal Inference* **10** 372–414. MR4512969 https://doi.org/10.1515/jci-2021-0059

COLNET, B., JOSSE, J., VAROQUAUX, G. and SCORNET, E. (2022b). Reweighting the RCT for generalization: finite sample analysis and variable selection.

COLNET, B., MAYER, I., CHEN, G., DIENG, A., LI, R., VAROQUAUX, G., VERT, J.-P, JOSSE, J. and YANG, S. (2024). Supplement to "Causal inference methods for combining randomized trials and observational studies: A review." https://doi.org/10.1214/23-STS889SUPP

CONCATO, J., SHAH, N. and HORWITZ, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N. Engl. J. Med.* **342** 1887–1892. https://doi.org/10.1056/NEJM200006223422507

COOPER, G. (1995). Causal discovery from data in the presence of selection bias. In *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics* 140–150.

CORNFIELD, J., HAENSZEL, W., HAMMOND, E. C., LILIENFELD, A. M., SHIMKIN, M. B. and WYNDER, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* **22** 173–203.

CORREA, J. D., TIAN, J. and BAREINBOIM, E. (2018). Generalized adjustment under confounding and selection biases. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

NATIONAL RESEARCH COUNCIL (2012). The prevention and treatment of missing data in clinical trials. *N. Engl. J. Med.* **367** 1355–1360.

CRASH-3 (2019). Effects of tranexamic acid on death, disability, vascular occlusive events and other morbidities in patients with acute traumatic brain injury (CRASH-3): A randomised, placebo-controlled trial. *Lancet* **394** 1713–1723.

CRO, S., MORRIS, T. P., KAHAN, B. C., CORNELIUS, V. R. and CARPENTER, J. R. (2020). A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. *BMC Med. Res. Methodol.* **20** 208. https://doi.org/10.1186/s12874-020-01089-6

DAGAN, N., BARDA, N., KEPTEN, E., MIRON, O., PERCHIK, S., KATZ, M. A., HERNÁN, M. A., LIPSITCH, M., REIS, B. et al. (2021). Bnt162b2 mrna Covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384** 1412–1423.

DAHABREH, I. J. and HERNÁN, M. A. (2019). Extending inferences from a randomized trial to a target population. *Eur. J. Epidemiol.* **34** 719–722.

DAHABREH, I. J., ROBERTSON, S. E., STEINGRIMSSON, J. A., STUART, E. A. and HERNÁN, M. A. (2020). Extending inferences from a randomized trial to a new target population. *Stat. Med.* **39** 1999–2014. MR4105273 https://doi.org/10.1002/sim.8426

DAHABREH, I. J., ROBERTSON, S. E., TCHETGEN, E. J., STUART, E. A. and HERNÁN, M. A. (2019). Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics* **75** 685–694. MR3999190 https://doi.org/10.1111/biom.13009

DAHABREH, I. J., ROBINS, J. M., HANEUSE, S. J.-P. A., SAEED, I., ROBERTSON, S. E., STUART, E. A. and HERNÁN, M. A. (2023a). Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. *Stat. Med.* **42** 2029–2043. MR4594699 https://doi.org/10.1002/sim.9550

DAHABREH, I. J., ROBINS, J. M., HANEUSE, S. J.-P. A., SAEED, I., ROBERTSON, S. E., STUART, E. A. and HERNÁN, M. A. (2023b). Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. *Stat. Med.* **42** 2029–2043. MR4594699 https://doi.org/10.1002/sim.9550

Dahabreh, I. J., Robins, J. M. and Hernán, M. A. (2020). Benchmarking observational methods by comparing randomized trials and their emulations. *Epidemiology* **31** 614–619.

Dawid, P., Humphreys, M. and Musio, M. (2019). Bounding causes of effects with mediators Technical Report. arXiv:1907.00399.

Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* **210** 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Deaton, A., Case, S. C., Côté, N., Drèze, J., Easterly, W., Khera, R., Pritchett, L. and Reddy, C. R. (2019). *Randomization in the Tropics Revisited: A Theme and Eleven Variations. Randomized Controlled Trials in the Field of Development: A Critical Perspective*. Oxford Univ. Press, London. Forthcoming.

Degtiar, I., Layton, T., Wallace, J. and Rose, S. (2021). Conditional cross-design synthesis estimators for generalizability in medicaid.

Degtiar, I. and Rose, S. (2023). A review of generalizability and transportability. *Annu. Rev. Stat. Appl.* **10** 501–524. MR4567803 https://doi.org/10.1146/annurev-statistics-042522-103837

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

Dewan, Y., Komolafe, E., Mejìa-Mantilla, J., Perel, P., Roberts, I. and Shakur-Still, H. (2012). CRASH-3: Tranexamic acid for the treatment of significant traumatic brain injury: Study protocol for an international randomized, double-blind, placebo-controlled trial. *Trials* **13** 87.

Didelez, V., Kreiner, S. and Keiding, N. (2010). Graphical models for inference under outcome-dependent sampling. *Statist. Sci.* **25** 368–387. MR2791673 https://doi.org/10.1214/10-STS340

Egami, N. and Hartman, E. (2021). Covariate selection for generalizing experimental results: Application to a large-scale development program in Uganda. *J. Roy. Statist. Soc. Ser. A* **184** 1524–1548. MR4344647 https://doi.org/10.1111/rssa.12734

FDA (2018). Framework for fda's real-world evidence program.

Frieden, T. R. (2017). Evidence for health decision making—beyond randomized, controlled trials. *N. Engl. J. Med.* **377** 465–475. https://doi.org/10.1056/NEJMra1614394

Geneletti, S., Richardson, S. and Best, N. (2008). Adjusting for selection bias in retrospective, case–control studies. *Biostatistics* **10** 17–31.

Gordon, B. R., Zettelmeyer, F., Bhargava, N. and Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Mark. Sci.* **38** 193–225.

Green, L. W. and Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Eval. Health Prof.* **29** 126–153. https://doi.org/10.1177/0163278705284445

Guo, F. R. and Perković, E. (2022). Efficient least squares for estimating total effects under linearity and causal sufficiency. *J. Mach. Learn. Res.* **23** Paper No. [104], 41. MR4576689

Guo, R., Cheng, L., Li, J., Hahn, P. R. and Liu, H. (2018). A survey of learning causality with data: Problems and methods. Preprint. Available at arXiv:1809.09337.

Guo, W., Wang, S., Ding, P., Wang, Y. and Jordan, M. I. (2021). Multi-source causal inference using control variates. Preprint. Available at arXiv:2103.16689.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46.

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 https://doi.org/10.1111/rssa.12094

He, Z., Tang, X., Yang, X., Guo, Y., George, T., Charness, N., Hem, K., Hogan, W. and Bian, J. (2020). Clinical trial generalizability assessment in the big data era: A review. *Clin. Transl. Sci.* **13**.

Hernán, M. and Robins, J. (2006). Instruments for causal inference: An epidemiologist's dream? In *Epidemiology* **17** 360–72, Cambridge, Mass.

Hernán, M. A., Cole, S. R., Margolick, J., Cohen, M. and Robins, J. M. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol. Drug Saf.* **14** 477–491. https://doi.org/10.1002/pds.1064

Hobbs, B. P., Sargent, D. J. and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Anal.* **7** 639–673. MR2981631 https://doi.org/10.1214/12-BA722

Hotz, V. J., Imbens, G. W. and Mortimer, J. H. (2005). Predicting the efficacy of future training programs using past experiences at other locations. *J. Econometrics* **125** 241–270. MR2143377 https://doi.org/10.1016/j.jeconom.2004.04.009

Huang, M. (2022). Sensitivity analysis in the generalization of experimental results.

Huang, M., Egami, N., Hartman, E. and Miratrix, L. (2023). Leveraging population outcomes to improve the generalization of experimental results: Application to the JTPA study. *Ann. Appl. Stat.* **17** 2139–2164. MR4637661 https://doi.org/10.1214/22-aoas1712

Huang, Y. and Valtorta, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, *UAI'06* 217–224. AUAI Press, Arlington, VA, USA.

Huang, Y. and Valtorta, M. (2012). Pearl's calculus of intervention is complete.

Huitfeldt, A., Swanson, S. A., Stensrud, M. J. and Suzuki, E. (2019). Effect heterogeneity and variable selection for standardizing causal effects to a target population. *Eur. J. Epidemiol.* **34** 1119–1129.

Hünermund, P. and Bareinboim, E. (2019). Causal inference and data-fusion in econometrics. Preprint. Available at arXiv:1912.09104.

Imbens, G. (2003). Sensitivity to exogeneity assumptions in program evaluation. *Amer. Econ. Rev.*

Imbens, G. W. (2014). Instrumental variables: An econometrician's perspective. *Statist. Sci.* **29** 323–358. MR3264545 https://doi.org/10.1214/14-STS480

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

Jeong, S. and Namkoong, H. (2022). Assessing external validity over worst-case subpopulations. arXiv:2007.02411 [cs, econ, stat].

Jiang, W., Josse, J., Lavielle, M. and Traumabase Group (2020). Logistic regression with missing covariates—parameter estimation, model selection and prediction within a joint-modeling framework. *Comput. Statist. Data Anal.* **145** 106907, 20. MR4053706 https://doi.org/10.1016/j.csda.2019.106907

Josey, K. P., Berkowitz, S. A., Ghosh, D. and Raghavan, S. (2021). Transporting experimental results with entropy balancing. *Stat. Med.* **40** 4310–4326. MR4300088 https://doi.org/10.1002/sim.9031

JOSSE, J., PROST, N., SCORNET, E. and VAROQUAUX, G. (2019). On the consistency of supervised learning with missing values. Preprint. Available at arXiv:1902.06931.

JUNG, Y., TIAN, J. and BAREINBOIM, E. (2020a). Estimating causal effects using weighting-based estimators. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34** 10186–10193.

JUNG, Y., TIAN, J. and BAREINBOIM, E. (2020b). Learning causal effects via weighted empirical risk minimization. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 12697–12709. Curran Associates, Red Hook.

JUNG, Y., TIAN, J. and BAREINBOIM, E. (2021). Estimating identifiable causal effects through double machine learning. *Proc. AAAI Conf. Artif. Intell.* **35** 12113–12122.

KALLUS, N., PULI, A. M. and SHALIT, U. (2018). Removing hidden confounding by experimental grounding. In *Advances in Neural Information Processing Systems* 10888–10897.

KARVANEN, J., TIKKA, S. and HYTTINEN, A. (2020). Do-search—a tool for causal inference and study design with multiple data sources.

KEIDING, N. and LOUIS, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. Roy. Statist. Soc. Ser. A* **179** 319–376. MR3461587 https://doi.org/10.1111/rssa.12136

KENNEDY, L. and GELMAN, A. (2021). Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample. *Psychol. Methods* **26** 547.

KENWARD, M. (2013). The handling of missing data in clinical trials. *Clin. Invest.* **3** 241–250.

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

KRUEGER, A. B. (1999). Experimental estimates of education production functions. *Q. J. Econ.* **114** 497–532.

LAURITZEN, S. L. and RICHARDSON, T. S. (2008). Discussion of mccullagh: Sampling bias and logistic models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 671.

LÊ, S., JOSSE, J. and HUSSON, F. (2008). FactoMineR: A package for multivariate analysis. *J. Stat. Softw.* **25** 1–18.

LEE, D., GHOSH, S. and YANG, S. (2022). Transporting survival of an HIV clinical trial to the external target populations. Preprint. Available at arXiv:2210.02571.

LEE, D., YANG, S., DONG, L., WANG, X., ZENG, D. and CAI, J. (2023). Improving trial generalizability using observational studies. *Biometrics* **79** 1213–1225. MR4606348 https://doi.org/10.1111/biom.13609

LEE, D., YANG, S. and WANG, X. (2022). Doubly robust estimators for generalizing treatment effects on survival outcomes from randomized controlled trials to a target population. *J. Causal Inference* **10** 415–440. MR4519820 https://doi.org/10.1515/jci-2022-0004

LEE, S., CORREA, J. and BAREINBOIM, E. (2020a). General transportability—synthesizing observations and experiments from heterogeneous domains. *Proc. AAAI Conf. Artif. Intell.* **34** 10210–10217.

LEE, S., CORREA, J. D. and BAREINBOIM, E. (2020b). General identifiability with arbitrary surrogate experiments. In *Proceedings of Machine Learning Research* (R. P. Adams and V. Gogate, eds.) **115** 389–398. PMLR.

LESKO, C. R., BUCHANAN, A. L., WESTREICH, D., EDWARDS, J. K., HUDGENS, M. G. and COLE, S. R. (2017). Generalizing study results: A potential outcomes perspective. *Epidemiology* **28** 553–561. https://doi.org/10.1097/EDE.0000000000000664

LESKO, C. R., COLE, S. R., HALL, H. I., WESTREICH, D., MILLER, W. C., ERON, J. J., LI, J., MUGAVERO, M. J. and FOR THE CNICS INVESTIGATORS (2016). The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–2011. *Int. J. Epidemiol.* **45** 140–150.

LI, F., BUCHANAN, A. L. and COLE, S. R. (2021). Generalizing trial evidence to target populations in non-nested designs: Applications to aids clinical trials.

LI, F., HONG, H. and STUART, E. A. (2023). A note on semiparametric efficient generalization of causal effects from randomized trials to target populations. *Comm. Statist. Theory Methods* **52** 5767–5798. MR4608916 https://doi.org/10.1080/03610926.2021.2020291

LI, P. and STUART, E. A. (2019). Best (but oft-forgotten) practices: Missing data methods in randomized controlled nutrition trials. *Am. J. Clin. Nutr.* **109** 504–508.

HERNÁN, M. A. and VAN DER WEELE, T. J. (2011). Compound treatments and transportability of causal inference. *Epidemiology* **22** 368–377.

LING, A. Y., MONTEZ-RATH, M. E., CARITA, P., CHANDROSS, K., LUCATS, L., MENG, Z., SEBASTIEN, B., KAPPHAHN, K. and DESAI, M. (2022). A critical review of methods for real-world applications to generalize or transport clinical trial findings to target populations of interest.

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

LODI, S., PHILLIPS, A., LUNDGREN, J., LOGAN, R., SHARMA, S., COLE, S., BABIKER, A., LAW, M., CHU, H. et al. (2019). Effect estimates in randomized trials and observational studies: Comparing apples with apples. *Amer. J. Epidemiol.* **188**.

MARTEL GARCIA, F. and WANTCHEKON, L. (2010). Theory, external validity, and experimental inference: Some conjectures. *Ann. Am. Acad. Polit. Soc. Sci.* **628** 132–147.

MAYER, I., JOSSE, J. and TRAUMABASE GROUP (2023). Generalizing treatment effects with incomplete covariates: Identifying assumptions and multiple imputation algorithms. *Biom. J.* **65** Paper No. 2100294, 30. MR4603527 https://doi.org/10.1002/bimj.202100294

MAYER, I., SVERDRUP, E., GAUSS, T., MOYER, J.-D., WAGER, S. and JOSSE, J. (2020). Doubly robust treatment effect estimation with missing attributes. *Ann. Appl. Stat.* **14** 1409–1431. MR4152139 https://doi.org/10.1214/20-AOAS1356

MAYER, I., ZHAO, P., GREIFER, N., HUNTINGTON-KLEIN, N. and JOSSE, J. (2022). Cran task view: Causal inference.

NEYMAN, J. (1923). Sur les applications de la thar des probabilities aux experiences Agaricales: Essay de principle. English translation of excerpts by Dabrowska, D. and Speed, T.. *Statist. Sci.* **5** 465–472.

NGUYEN, T., ACKERMAN, B., SCHMID, I., COLE, S. and STUART, E. (2018). Sensitivity analyses for effect modifiers not observed in the target population when generalizing treatment effects from a randomized controlled trial: Assumptions, models, effect scales, data scenarios, and implementation details. *PLoS ONE* **13** e0208795.

NGUYEN, T. Q., EBNESAJJAD, C., COLE, S. R. and STUART, E. A. (2017). Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects. *Ann. Appl. Stat.* **11** 225–247. MR3634322 https://doi.org/10.1214/16-AOAS1001

NIE, X., IMBENS, G. and WAGER, S. (2021). Covariate balancing sensitivity analysis for extrapolating randomized trials across locations.

O'KELLY, M. and RATITCH, B. (2014). *Clinical Trials with Missing Data: A Guide for Practitioners*. Wiley, New York.

O'MUIRCHEARTAIGH, C. and HEDGES, L. V. (2014). Generalizing from unrepresentative experiments: A stratified propensity score approach. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 195–210. MR3234340 https://doi.org/10.1111/rssc.12037

PEARL, J. (1993). [Bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statist. Sci.* **8** 266–269.

PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82** 669–710. MR1380809 https://doi.org/10.1093/biomet/82.4.669

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

PEARL, J. (2009a). Causal inference in statistics: An overview. *Stat. Surv.* **3** 96–146. MR2545291 https://doi.org/10.1214/09-SS057

PEARL, J. (2009b). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

PEARL, J. (2015). Generalizing experimental findings. *J. Causal Inference* **3** 259–266. MR4289440 https://doi.org/10.1515/jci-2015-0025

PEARL, J. and BAREINBOIM, E. (2011). Transportability of causal and statistical relations: A formal approach. In 2011 *IEEE* 11*th International Conference on Data Mining Workshops* (*ICDMW*) 540–547. IEEE.

PEARL, J. and BAREINBOIM, E. (2014). External validity: From do-calculus to transportability across populations. *Statist. Sci.* **29** 579–595. MR3300360 https://doi.org/10.1214/14-STS486

POCOCK, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29** 175–188.

PRENTICE, R. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat. Med.* **8** 431–440.

RICHARDSON, T. S. and ROBINS, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128:2013.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512.

ROBINS, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis, MN*, 1997). *IMA Vol. Math. Appl.* **116** 95–133. Springer, New York. MR1731682 https://doi.org/10.1007/978-1-4612-1284-3_2

ROSENBAUM, P. R. (2002). Sensitivity to hidden bias. In *Observational Studies* 105–170. Springer, Berlin.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). Assessing the sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. Roy. Statist. Soc. Ser. B* **45** 212–218.

ROTHWELL, P. M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* **365** 82–93.

ROTNITZKY, A. and SMUCLER, E. (2020). Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J. Mach. Learn. Res.* **21** Paper No. 188, 86. MR4209474

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

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

RUDOLPH, K. E., SCHMIDT, N. M., GLYMOUR, M. M., CROWDER, R., GALIN, J., AHERN, J. and OSYPUK, T. L. (2018). Composition or context: Using transportability to understand drivers of site differences in a large-scale housing experiment. *Epidemiology* **29** 199–206.

RUDOLPH, K. E. and VAN DER LAAN, M. J. (2017). Robust estimation of encouragement design intervention effects transported across sites. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1509–1525. MR3731673 https://doi.org/10.1111/rssb.12213

SCHMIDLI, H., GSTEIGER, S., ROYCHOUDHURY, S., O'HAGAN, A., SPIEGELHALTER, D. and NEUENSCHWANDER, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* **70** 1023–1032. MR3295763 https://doi.org/10.1111/biom.12242

SHAKUR-STILL, H., ROBERTS, I., BAUTISTA, R., CABALLERO, J., COATS, T., DEWAN, Y., EL-SAYED, H., TAMAR, G., GUPTA, S. et al. (2009). Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): A randomised, placebo-controlled trial. *Lancet* **376** 23–32.

SHPITSER, I. and PEARL, J. (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the* 21*st National Conference on Artificial Intelligence—Volume* 2, *AAAI*'06 1219–1226. AAAI Press, Menlo Park.

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812 https://doi.org/10.1214/09-STS313

STUART, E. A., ACKERMAN, B. and WESTREICH, D. (2018). Generalizability of randomized trial results to target populations: Design and analysis possibilities. *Res. Soc. Work Pract.* **28** 532–537.

STUART, E. A., BRADSHAW, C. P. and LEAF, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prev. Sci.* **16** 475–485. https://doi.org/10.1007/s11121-014-0513-z

STUART, E. A., COLE, S. R., BRADSHAW, C. P. and LEAF, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *J. Roy. Statist. Soc. Ser. A* **174** 369–386. MR2898850 https://doi.org/10.1111/j.1467-985X.2010.00673.x

STUART, E. A. and RHODES, A. (2017). Generalizing treatment effect estimates from sample to population: A case study in the difficulties of finding sufficient data. *Eval. Rev.* **41** 357–388. PMID: 27491758.

SUGIYAMA, M. and KAWANABE, M. (2012). *Machine Learning in Non-stationary Environments: Introduction to Covariate Shift Adaptation*. MIT Press, Cambridge.

TAN, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Amer. Statist. Assoc.* **101** 1619–1637. MR2279484 https://doi.org/10.1198/016214506000000023

R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

TIAN, J. and PEARL, J. (2000). Probabilities of causation: Bounds and identification. *Ann. Math. Artif. Intell.* **28** 287–313. MR1797625 https://doi.org/10.1023/A:1018912507879

TIBSHIRANI, J., ATHEY, S. and WAGER, S. (2020). *grf: Generalized random forests*. R package version 1.2.0.

TIKKA, S., HYTTINEN, A. and KARVANEN, J. (2019). Causal effect identification from multiple incomplete data sources: A general search-based approach. Preprint. Available at arXiv:1902.01073.

TIKKA, S. and KARVANEN, J. (2017). Identifying causal effects with the R package causaleffect. *J. Stat. Softw.* **76** 1–30.

TIPTON, E. (2013). Improving generalizations from experiments using propensity score subclassification: Assumptions, properties, and contexts. *J. Educ. Behav. Stat.* **38** 239–266.

TIPTON, E., HALLBERG, K., HEDGES, L. V. and CHAN, W. (2017). Implications of small samples for generalization: Adjustments and rules of thumb. *Eval. Rev.* **41** 472–505. https://doi.org/10.1177/0193841X16655665

TWALA, B., JONES, M. and HAND, D. J. (2008). Good methods for coping with missing data in decision trees. *Pattern Recogn. Lett.* **29** 950–956.

VAN DER WEELE, T. J. and ROBINS, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology* **18** 561–568.

WESTREICH, D., EDWARDS, J. K., LESKO, C. R., COLE, S. R. and STUART, E. A. (2018). Target validity and the hierarchy of study designs. *Amer. J. Epidemiol.* **188** 438–443.

WESTREICH, D., EDWARDS, J. K., LESKO, C. R., STUART, E. and COLE, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *Amer. J. Epidemiol.* **186** 1010–1014.

WITTE, J., HENCKEL, L., MAATHUIS, M. H. and DIDELEZ, V. (2020). On efficient adjustment in causal graphs. *J. Mach. Learn. Res.* **21** Paper No. 246, 45. MR4213427

WU, L. and YANG, S. (2022). Integrative r-learner of heterogeneous treatment effects combining experimental and observational studies. In *Proceedings of the* 1*st Conference on Causal Learning and Reasoning*.

WU, L. and YANG, S. (2023). Transfer learning of individualized treatment rules from experimental to real-world data. *J. Comput. Graph. Statist.* **32** 1036–1045. MR4641479 https://doi.org/10.1080/10618600.2022.2141752

YANG, S. and DING, P. (2020). Combining multiple observational data sources to estimate causal effects. *J. Amer. Statist. Assoc.* **115** 1540–1554. MR4143484 https://doi.org/10.1080/01621459.2019.1609973

YANG, S. and KIM, J. K. (2020). Statistical data integration in survey sampling: A review. *Jpn. J. Stat. Data Sci.* **3** 625–650. MR4181993 https://doi.org/10.1007/s42081-020-00093-w

YANG, S., KIM, J. K. and SONG, R. (2020). Doubly robust inference when combining probability and non-probability samples with high dimensional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 445–465. MR4084171

YANG, S. and WANG, X. (2022). RWD-integrated randomized clinical trial analysis. In *ASA Biopharmaceutical Report Real World Evidence* (H. Pang, L. Wang and K. L. Griffiths, eds.) **29** 15–21.

YANG, S., WANG, X. and ZENG, D. (2023). Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. *J. Roy. Statist. Soc. Ser. B* https://doi.org/10.1093/jrsssb/qkad017

YANG, S., ZENG, D. and WANG, X. (2020). Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. Preprint. Available at arXiv:2007.12922.

YAO, L., CHU, Z., LI, S., LI, Y., GAO, J. and ZHANG, A. (2020). A survey on causal inference. Preprint. Available at arXiv:2002.02770.

ZHONG, Y., KENNEDY, E. H., BODNAR, L. M. and NAIMI, A. I. (2021). Aipw: An R package for augmented inverse probability weighted estimation of average causal effects. *Amer. J. Epidemiol.*

ZIVICH, P., KLOSE, M., COLE, S., EDWARDS, J. and SHOOK-SA, B. (2022). Delicatessen: M-estimation in Python.

Statistical Science
2024, Vol. 39, No. 1, 192–208
https://doi.org/10.1214/22-STS878

A Conversation with Stephen M. Stigler

Sam Behseta and Robert E. Kass

Abstract. Stephen M. Stigler received his Ph.D. in Statistics from the University of California, Berkeley, with a dissertation on the asymptotic distribution of linear functions of order statistics. Starting in 1967, he taught at the University of Wisconsin, Madison, then in 1979 moved to the University of Chicago where he taught from 1979 to 2021. Stigler has worked on a variety of topics in mathematical statistics, ranging from asymptotic theory to the theory of experimental design, and on applications of statistics including in anthropology, forensic science, paleontology, psychology, information transfer and sports. In recent years, he has concentrated on the history of statistics, with inquiries ranging from the development of statistical methods in astronomy and geodesy and their spread to biological and social sciences, to lotteries, to the modern development of statistical theory. He has published four books, *The History of Statistics* (1986), *Statistics on the Table* (1999), *The Seven Pillars of Statistical Wisdom* (2016) and *Casanova's Lottery* (2022). A recent research focus has been upon the way the work of Francis Galton on the statistics of inheritance led to the creation of modern multivariate analysis and made a true Bayesian inference possible, and on how R. A. Fisher's transformation of Karl Pearson's path breaking research led to a modern period of statistical enlightenment.

Stigler is an elected member of the American Academy of Arts and Sciences and of the American Philosophical Society; he has served as President of the Institute of Mathematical Statistics and of the International Statistical Institute. In 2005, he received the Humboldt Foundation Research Award; in 2010, he was elected Membre Associé of the Académie royale de Belgique, Classe des Sciences. Stigler served as Theory and Methods Editor for the Journal of the American Statistical Association 1979–1982. He was a Guggenheim Fellow in 1977, and received awards for undergraduate teaching at the University of Wisconsin (1971) and University of Chicago (1998).

This interview with Stigler was conducted remotely in July 2021.

Key words and phrases: Statistical training, history of statistics, University of Chicago, academic life, statistical narrative.

REFERENCES

ANSCOMBE, F. R. (2003). Quiet contributor: The civic career and times of John W. Tukey. *Statist. Sci.* **18** 287–310. MR2056571 https://doi.org/10.1214/ss/1076102417

BANKS, D. L. (1996). A conversation with I. J. Good. *Statist. Sci.* **11** 1–19. MR1437124 https://doi.org/10.1214/ss/1032209661

EATON, M. L. (2007). William H. Kruskal and the development of coordinate-free methods. *Statist. Sci.* **22** 264–265. MR2408962 https://doi.org/10.1214/088342306000000367

FIENBERG, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Anal.* **1** 1–40. MR2227361 https://doi.org/10.1214/06-BA101

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **20** 225–241. MR0030730 https://doi.org/10.1214/aoms/1177730032

KASS, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4** 188–219.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. 2nd ed. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

MOSTELLER, F. and TUKEY, J. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, MA.

Sam Behseta is a Professor of Mathematics, California State University, Fullerton, 800 North College Blvd., Fullerton, California, 92831, USA (e-mail: sbehseta@fullerton.edu). Robert E. Kass is the Maurice Falk Professor of Statistics and Computational Neuroscience, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, USA (e-mail: kass@cmu.edu).

SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **80** 360–361. MR0792734

STIGLER, S. M. (1969). Linear functions of order statistics. *Ann. Math. Stat.* **40** 770–788. MR0264822 https://doi.org/10.1214/aoms/1177697587

STIGLER, S. M. (1973a). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *J. Amer. Statist. Assoc.* **68** 872–879. MR0362575

STIGLER, S. M. (1973b). Studies in the history of probability and statistics. XXXII. Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* **60** 439–445. MR0326872 https://doi.org/10.2307/2334992

STIGLER, S. M. (1976). Contribution to discussion of "On Re-reading R. A. Fisher," by L. J. Savage. *Ann. Statist.* **4** 498–500.

STIGLER, S. M. (1980). Stigler's law of eponymy. In *Science and Social Structure*: *A Festschrift for Robert K. Merton*. *Transactions N.Y. Acad. Sci. Series* **2** 147–157.

STIGLER, S. M. (1983). Who discovered Bayes's theorem? *Amer. Statist.* **37** 290–296.

STIGLER, S. M. (1986a). *The History of Statistics*: *The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR0852410

STIGLER, S. M. (1986b). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378. MR0858515

STIGLER, S. M. (1992). Apollo mathematicus: A story of resistance to quantification in the seventeenth century. *Proc. Amer. Philos. Soc.* **136** 93–126.

STIGLER, S. M. (1999). *Statistics on the Table*: *The History of Statistical Concepts and Methods*. Harvard Univ. Press, Cambridge, MA. MR1712969

STIGLER, S. M. (2003). Casanova, 'Bonaparte', and the loterie de France. *J. Soc. Fr. Stat.* **144** 5–34.

STIGLER, S. M. (2007). The epic story of maximum likelihood. *Statist. Sci.* **22** 598–620. MR2410255 https://doi.org/10.1214/07-STS249

STIGLER, S. M. (2008). Karl Pearson's theoretical errors and the advances they inspired. *Statist. Sci.* **23** 261–271. MR2446501 https://doi.org/10.1214/08-STS256

STIGLER, S. M. (2016). *The Seven Pillars of Statistical Wisdom*. Harvard Univ. Press, Cambridge, MA. MR3585675 https://doi.org/10.4159/9780674970199

STIGLER, S. M. (2018). Richard Price, the first Bayesian. *Statist. Sci.* **33** 117–125. MR3757508 https://doi.org/10.1214/17-STS635

STIGLER, S. M. (2022). *Casanova's Lottery*: *The History of a Revolutionary Game of Chance*. Univ. Chicago Press, Chicago, IL.

VON NEUMANN, J. (1955). *The Mathematical Foundations of Quantum Mechanics*. Princeton Univ. Press, Princeton, NJ. Translated from the 1932 German edition.

WHITEHEAD, A. N. (1917). *The Organization of Thought*. Williams and Norgate, London.

WICHURA, M. J. (2006). *The Coordinate-Free Approach to Linear Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **19**. Cambridge Univ. Press, Cambridge. MR2283455 https://doi.org/10.1017/CBO9780511546822

YOUNG, G. A. (1994). Bootstrap: More than a stab in the dark? *Statist. Sci.* **9** 382–415. MR1325434

Statistical Science
2024, Vol. 39, No. 1, 209–220
https://doi.org/10.1214/23-STS897

In Conversation with Sir David Spiegelhalter and Professor Sylvia Richardson

Bhramar Mukherjee

Abstract. Sir David Spiegelhalter and Professor Sylvia Richardson are two eminent statisticians of our time who have made pioneering contributions to Statistics and Data Science with careers spanning over more than four decades. They have a long and celebrated legacy built through foundational research in Bayesian statistics, impactful collaborations, steadfast professional service and superb scientific communications. They have won many prestigious awards and recognitions throughout their distinguished careers. During my sabbatical in 2022 at the University of Cambridge I had the honor of sitting down for a conversation with these two remarkable individuals. We discussed early career influences and digressions, research philosophy, role of mentors and advice for the future generation. We hope this conversation with David and Sylvia will inspire many future statisticians.

Key words and phrases: Bayesian Statistics, Imperial College, Markov chain Monte Carlo, Medical Research Council, Royal Statistical Society, scientific communication, University of Cambridge.

REFERENCES

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

CLIFFORD, P., RICHARDSON, S. and HÉMON, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* **45** 123–134. MR0999441

COVID-19 TASK FORCE (n.d.). RSS. Retrieved 28 April 2023. Available at https://rss.org.uk/policy-campaigns/policy-groups/covid-19-task-force/.

DE FINETTI, B. (1974). *Theory of Probability: A Critical Introductory Treatment. Vol. 1. Wiley Series in Probability and Mathematical Statistics*. Wiley, London-Sydney. MR0440640

GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice*. *Interdisciplinary Statistics*. CRC Press, London. MR1397966 https://doi.org/10.1007/978-1-4899-4485-6

GREEN, P. J., HJORT, N. L. and RICHARDSON, S. (2003). Highly structured stochastic systems. Oxford Univ. Press, Oxford.

GRIGG, O. A., FAREWELL, V. T. and SPIEGELHALTER, D. J. (2003). Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat. Methods Med. Res.* **12** 147–170. MR1963338 https://doi.org/10.1177/096228020301200205

LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 157–224. MR0964177

LUNN, D. J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* **10** 325–337. https://doi.org/10.1023/A:1008929526011

MCCONWAY, K. and SPIEGELHALTER, D. (2021). Sound human, steer clear of jargon, and be prepared. *Significance* **18** 32–34. https://doi.org/10.1111/1740-9713.01508

MOLITOR, J., PAPATHOMAS, M., JERRETT, M. and RICHARDSON, S. (2010). Bayesian profile regression with an application to the national survey of children's health. *Biostatistics* **11** 484–498. https://doi.org/10.1093/biostatistics/kxq013

NICHOLSON, G., BLANGIARDO, M., BRIERS, M., DIGGLE, P. J., FJELDE, T. E., GE, H., GOUDIE, R. J. B., JERSAKOVA, R., KING, R. E. et al. (2022). Interoperability of statistical models in pandemic preparedness: Principles and reality. *Statist. Sci.* **37** 183–206. https://doi.org/10.1214/22-STS854

POUWELS, K. B., HOUSE, T., PRITCHARD, E., ROBOTHAM, J. V., BIRRELL, P. J., GELMAN, A., VIHTA, K.-D., BOWERS, N., BOREHAM, I. et al. (2021). Community prevalence of Sars-CoV-2 in England from April to November, 2020: Results from the ONS coronavirus infection survey. *Lancet Public Health* **6** e30–e38. https://doi.org/10.1016/S2468-2667(20)30282-6

RICHARDSON, S. and BEST, N. (2003). Bayesian hierarchical models in ecological studies of health–environment effects. *Environmetrics* **14** 129–147. https://doi.org/10.1002/env.571

RICHARDSON, S. and GILKS, W. R. (1993). A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *Amer. J. Epidemiol.* **138** 430–442. https://doi.org/10.1093/oxfordjournals.aje.a116875

Bhramar Mukherjee is a Distinguished University Professor and current Chair, Department of Biostatistics University of Michigan, Ann Arbor, MI 48109-2029, United States (e-mail: bhramar@umich.edu).

RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

ROBERTS, M., DRIGGS, D., THORPE, M., GILBEY, J., YEUNG, M., URSPRUNG, S., AVILES-RIVERO, A. I., ETMANN, C., MCCAGUE, C. et al. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for Covid-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **3** Article 3. https://doi.org/10.1038/s42256-021-00307-0

RUFFIEUX, H., DAVISON, A. C., HAGER, J., INSHAW, J., FAIRFAX, B. P., RICHARDSON, S. and BOTTOLO, L. (2020). A global-local approach for detecting hotspots in multiple-response regression. *Ann. Appl. Stat.* **14** 905–928. MR4117834 https://doi.org/10.1214/20-AOAS1332

SPIEGELHALTER, D. (2019). *The Art of Statistics*: *How to Learn from Data*. Basic Books, New York. MR3967688

SPIEGELHALTER, D., GRIGG, O., KINSMAN, R. and TREASURE, T. (2003). Risk-adjusted sequential probability ratio tests: Applications to Bristol, Shipman and adult cardiac surgery. *Int. J. Qual. Health Care* **15** 7–13. https://doi.org/10.1093/intqhc/15.1.7

. UNIVERSITY OF OXFORD (2023). Covid-19 infection survey, Nuffield Department of Medicine. Available at https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey.

WINTON CENTRE CAMBRIDGE (2021). Latest data from the MHRA on blood clots associated with the Astra Zeneca COVID-19 vaccine. Available at https://wintoncentre.maths.cam.ac.uk/news/latest-data-mhra-blood-clots-associated-astra-zeneca-covid-19-vaccine/.

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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