STATISTICAL SCIENCE

Volume 41, Number 2 May 2026

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 41, Number 2, May 2026. 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 41, Number 2 (227–459) May 2026

Volume 41 Number 2 May 2026

Statistical Science
2026, Vol. 41, No. 2, 227–246
https://doi.org/10.1214/24-STS961

Bayesian Nonparametric Inference for "Species-Sampling" Problems

Cecilia Balocchi, Stefano Favaro and Zacharie Naulet

Abstract. Given an observed sample from a population of individuals belonging to species, "species-sampling" problems (SSPs) call for estimating some features of the unknown species composition of additional unobservable samples from the same population. Within SSPs the problems of estimating coverage probabilities, the number of unseen species and coverages of prevalences have emerged in the past three decades for being the subject of numerous methodological and applied works, mostly in biological sciences but also in statistical machine learning, electrical engineering, theoretical computer science, information theory and forensic statistics. In this paper we focus on these popular SSPs and present an overview of their Bayesian nonparametric (BNP) analysis under the Pitman–Yor process (PYP) prior. While reviewing the literature, we improve on computation and interpretability of existing posterior inferences, typically expressed through complicated combinatorial numbers, by establishing novel posterior representations in terms of simple compound binomial and hypergeometric distributions. We also consider the problem of estimating the discount and scale parameters of the PYP prior, showing a property of Bayesian consistency with respect to estimation through the hierarchical Bayes and empirical Bayes approaches; that is, the discount parameter can be estimated consistently, whereas the scale parameter cannot be estimated consistently, thus advising caution in posterior inference. We conclude our work by discussing some generalizations of SSPs, mostly in the field of biological sciences, which deal with "feature-sampling," multiple populations of individuals sharing species and classes of Markov chains.

Key words and phrases: Bayesian nonparametrics, Bayesian consistency, coverage of prevalences, coverage probabilities, empirical Bayes, hierarchical Bayes, Pitman–Yor process prior, "species-sampling" problems, unseen species.

REFERENCES

ACHARYA, J., BAO, Y., KANG, Y. and SUN, Z. (2018). Improved bounds on minimax risk of estimating missing mass. In *IEEE International Symposium on Information Theory* 326–330.

ANEVSKI, D., GILL, R. D. and ZOHREN, S. (2017). Estimating a probability mass function with unknown labels. *Ann. Statist.* **45** 2708–2735. MR3737907 https://doi.org/10.1214/17-AOS1542

ARBEL, J., FAVARO, S., NIPOTI, B. and TEH, Y. W. (2017). Bayesian nonparametric inference for discovery probabilities: Credible intervals and large sample asymptotics. *Statist. Sinica* **27** 839–858. MR3675033

ASADI, M., TORGHABEH, R. P. and SANTHANAM, N. P. (2014). Stationary and transition probabilities in slow mixing, long memory Markov processes. *IEEE Trans. Inf. Theory* **60** 5682–5701. MR3252414 https://doi.org/10.1109/TIT.2014.2334594

AYED, F., BATTISTON, M., CAMERLENGHI, F. and FAVARO, S. (2021a). On consistent and rate optimal estimation of the missing mass. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1476–1494. MR4291456 https://doi.org/10.1214/20-aihp1126

AYED, F., BATTISTON, M., CAMERLENGHI, F. and FAVARO, S. (2021b). Consistent estimation of small masses in feature sampling. *J. Mach. Learn. Res.* **22** Paper No. 6, 28. MR4253699

BACALLADO, S., BATTISTON, M., FAVARO, S. and TRIPPA, L. (2017). Sufficientness postulates for Gibbs-type priors and hierarchical generalizations. *Statist. Sci.* **32** 487–500. MR3730518 https://doi.

Cecilia Balocchi is Assistant Professor, School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom (e-mail: cecilia.balocchi@ed.ac.uk). Stefano Favaro is Professor, Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto, Torino, Italy (e-mail: stefano.favaro@unito.it). Zacharie Naulet is Assistant Professor, Department of Mathematics, Université Paris-Saclay, Orsay, France (e-mail: zacharie.naulet@universite-paris-saclay.fr).

org/10.1214/17-STS619

BACALLADO, S., FAVARO, S. and TRIPPA, L. (2013). Bayesian nonparametric analysis of reversible Markov chains. *Ann*. *Statist*. **41** 870–896. MR3099124 https://doi.org/10.1214/13-AOS1102

BALOCCHI, C., FAVARO, S. and NAULET, Z. (2026). Supplement to "Bayesian nonparametric inference for "species-sampling" problems." https://doi.org/10.1214/24-STS961SUPP

BEN-HAMOU, A., BOUCHERON, S. and GASSIAT, E. (2018). Pattern coding meets censoring: (almost) adaptive coding on countable alphabets. Preprint. Available at arXiv:1608.08367.

BEN-HAMOU, A., BOUCHERON, S. and OHANNESSIAN, M. I. (2017). Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli* **23** 249–287. MR3556773 https://doi.org/10.3150/15-BEJ743

BERAHA, M. and FAVARO, S. (2023). Transform-scaled process priors for trait allocations in Bayesian nonparametrics. Preprint. Available at arXiv:2303.17844.

BERAHA, M., MASOERO, L., FAVARO, S. and RICHARDSON, T. S. (2023). A nonparametric Bayes approach to online activity prediction. Preprint. Available at arXiv:2401.14722.

BERCU, B. and FAVARO, S. (2024). A martingale approach to Gaussian fluctuations and laws of iterated logarithm for Ewens–Pitman model. *Stochastic Process*. *Appl*. **178** Paper No. 104493, 19. MR4798925 https://doi.org/10.1016/j.spa.2024.104493

BETHLEHEM, J. G., KELLER, W. J. and PANNEKOEK, J. (1990). Disclosure control of microdata. *J*. *Amer*. *Statist*. *Assoc*. **85** 38–45.

BUBECK, S., ERNST, D. and GARIVIER, A. (2013). Optimal discovery with probabilistic expert advice: Finite time analysis and macroscopic optimality. *J*. *Mach*. *Learn*. *Res*. **14** 601–623. MR3033341

BUNGE, J. and FITZPATRICK, M. (1993). Estimating the number of species: A review. *J*. *Amer*. *Statist*. *Assoc*. **88** 364–373.

CAI, D., MITZENMACHER, M. and ADAMS, R. P. (2018). A Bayesian nonparametric view on count–min sketch. In *Advances in Neural Information Processing Systems* **31**.

CAMERLENGHI, F., FAVARO, S., MASOERO, L. and BRODERICK, T. (2024). Scaled process priors for Bayesian nonparametric estimation of the unseen genetic variation. *J*. *Amer*. *Statist*. *Assoc*. **119** 320–331. MR4713895 https://doi.org/10.1080/01621459.2022.2115918

CAMERLENGHI, F., FAVARO, S., NAULET, Z. and PANERO, F. (2021). Optimal disclosure risk assessment. *Ann*. *Statist*. **49** 723–744. MR4255105 https://doi.org/10.1214/20-aos1975

CAMERLENGHI, F., LIJOI, A., ORBANZ, P. and PRÜNSTER, I. (2019). Distribution theory for hierarchical processes. *Ann*. *Statist*. **47** 67–92. MR3909927 https://doi.org/10.1214/17-AOS1678

CEREDA, G. (2017). Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach). *Scand*. *J*. *Stat*. **44** 230–248. MR3619703 https://doi.org/10.1111/sjos.12250

CHAKRABORTY, S., ARORA, A., BEGG, C. B. and SHEN, R. (2019). Using somatic variant richness to mine signals from rare variants in the cancer genome. *Nat*. *Commun*. **10** 5506.

CHANDRA, P., THANGARAJ, A. and RAJARAMAN, N. (2021). Missing mass of rank-2 Markov chains. Preprint. Available at arXiv:2102.01938.

CHAO, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand*. *J*. *Stat*. **11** 265–270. MR0793175

CHAO, A. and LEE, S.-M. (1992). Estimating the number of classes via sample coverage. *J*. *Amer*. *Statist*. *Assoc*. **87** 210–217. MR1158639

CHARALAMBIDES, C. A. (2005). *Combinatorial Methods in Discrete Distributions*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2131068 https://doi.org/10.1002/0471733180

CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev*. **51** 661–703. MR2563829 https://doi.org/10.1137/070710111

DALEY, T. and SMITH, A. D. (2013). Predicting the molecular complexity of sequencing libraries. *Nat*. *Methods* **10** 325–327. https://doi.org/10.1038/nmeth.2375

DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R. H., PRÜNSTER, I. and RUGGIERO, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **37** 212–229. https://doi.org/10.1109/TPAMI.2013.217

DENG, C., DALEY, T., DE SENA BRANDINE, G. and SMITH, A. D. (2019). Molecular heterogeneity in large-scale biological data: Techniques and applications. *Annu*. *Rev*. *Biomed*. *Data Sci*. **2** 39–67.

DEVROYE, L. (2009). Random variate generation for exponentially and polynomially tilted stable distribution. *ACM Trans*. *Model*. *Comput*. *Simul*. **19** 4.

DOLERA, E. and FAVARO, S. (2020a). A Berry–Esseen theorem for Pitman's α-diversity. *Ann*. *Appl*. *Probab*. **30** 847–869. MR4108124 https://doi.org/10.1214/19-AAP1518

DOLERA, E. and FAVARO, S. (2020b). Rates of convergence in de Finetti's representation theorem, and Hausdorff moment problem. *Bernoulli* **26** 1294–1322. MR4058368 https://doi.org/10.3150/19-BEJ1156

EFRON, B. (2003). Robbins, empirical Bayes and microarrays. *Ann*. *Statist*. **31** 366–378. Dedicated to the memory of Herbert E. Robbins. MR1983533 https://doi.org/10.1214/aos/1051027871

EFRON, B. and THISTED, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.

ESTY, W. W. (1982). Confidence intervals for the coverage of low coverage samples. *Ann*. *Statist*. **10** 190–196. MR0642730

ESTY, W. W. (1983). A normal limit law for a nonparametric estimator of the coverage of a random sample. *Ann*. *Statist*. **11** 905–912. MR0707940 https://doi.org/10.1214/aos/1176346256

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor*. *Popul*. *Biol*. **3**. MR0325177 https://doi.org/10.1016/0040-5809(72)90035-4

FALAHATGAR, M., ORLITSKY, A., PICHAPATI, V. and SURESH, A. T. (2016). Learning Markov distributions: Does estimation trump compression? In *Proceedings of the IEEE International Symposium on Information Theory* 2689–2693.

FAVARO, S., LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2009). Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 993–1008. MR2750254 https://doi.org/10.1111/j.1467-9868.2009.00717.x

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2012). A new estimator of the discovery probability. *Biometrics* **68** 1188–1196. MR3040025 https://doi.org/10.1111/j.1541-0420.2012.01793.x

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *Ann*. *Appl*. *Probab*. **23** 1721–1754. MR3114915 https://doi.org/10.1214/12-AAP843

FAVARO, S. and NAULET, Z. (2023b). Near-optimal estimation of the unseen under regularly varying tail populations. *Bernoulli* **29** 3423–3442. MR4632144 https://doi.org/10.3150/23-bej1589

FAVARO, S., NIPOTI, B. and TEH, Y. W. (2016). Rediscovery of Good–Turing estimators via Bayesian nonparametrics. *Biometrics* **72** 136–145. MR3500582 https://doi.org/10.1111/biom.12366

FAVARO, S., PANERO, F. and RIGON, T. (2021). Bayesian nonparametric disclosure risk assessment. *Electron*. *J*. *Stat*. **15** 5626–5651. MR4355694 https://doi.org/10.1214/21-ejs1933

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann*. *Statist*. **1** 209–230. MR0350949

FISHER, R. A., CORBET, A. S. and WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12** 42–58.

FLORENCIO, D. and HERLEY, C. (2007). A large-scale study of web password habits. In *Proceedings of the International Conference on World Wide Web* 657–666.

FRANSSEN, S. E. M. P. and VAN DER VAART, A. W. (2022). Empirical and Full Bayes estimation of the type of a Pitman–Yor process. arXiv preprint. Available at arXiv:2208.14255.

GALE, W. A. and SAMPSON, G. (1995). Good–Turing frequency estimation without tears. *J. Quant. Linguist.* **2** 217–237.

GAO, F. (2013). Moderate deviations for a nonparametric estimator of sample coverage. *Ann. Statist.* **41** 641–669. MR3099116 https://doi.org/10.1214/13-AOS1091

GAO, Z., TSENG, C. H., PEI, Z. and BLASER, M. J. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proc. Natl. Acad. Sci. USA* **104** 2927–2932.

GNEDIN, A., HANSEN, B. and PITMAN, J. (2007). Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Probab. Surv.* **4** 146–171. MR2318403 https://doi.org/10.1214/07-PS092

GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264. MR0061330 https://doi.org/10.1093/biomet/40.3-4.237

GOOD, I. J. and TOULMIN, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63. MR0077039 https://doi.org/10.1093/biomet/43.1-2.45

GOODMAN, L. A. (1949). On the estimation of the number of classes in a population. *Ann. Math. Stat.* **20** 572–579. MR0032165 https://doi.org/10.1214/aoms/1177729949

GRABCHAK, M. and ZHANG, Z. (2017). Asymptotic properties of Turing's formula in relative error. *Mach. Learn.* **106** 1771–1785. MR3708277 https://doi.org/10.1007/s10994-016-5620-6

GRAVEL, S. (2014). Predicting discovery rates of genomic features. *Genetics* **197** 601–610. https://doi.org/10.1534/genetics.114.162149

GRIFFIN, J. E. and LEISEN, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 525–545. MR3611758 https://doi.org/10.1111/rssb.12176

HAAS, P. J., NAUGHTON, J. F., SESHADRI, S. and STOKES, L. (1995). Sampling-based estimation of the number of distinct values of an attribute. In *Proceedings of the Very Large Data Bases Conference* 311–322.

HANSEN, B. and PITMAN, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statist. Probab. Lett.* **46** 251–256. MR1745692 https://doi.org/10.1016/S0167-7152(99)00109-1

HAO, Y. and LI, P. (2020a). Bessel smoothing and multi-distribution property estimation. In *Proceedings of the Conference on Computational Learning Theory* **125** 1817–1876.

HAO, Y. and LI, P. (2020b). Optimal prediction of the number of unseen species with multiplicity. In *Advances in Neural Information Processing Systems* **33**.

HAO, Y. and ORLITSKY, A. (2020). Profile entropy: A fundamental measure for the learnability and compressibility of distributions. In *Advances in Neural Information Processing Systems* **34**.

HAO, Y., ORLITSKY, A. and PICHAPATI, V. (2018). On learning Markov chains. In *Advances in Neural Information Processing Systems* **32**.

HOFERT, M. (2011). Sampling exponentially tilted stable distributions. *ACM Trans. Model. Comput. Simul.* **22** Art. 3, 11. MR2955859 https://doi.org/10.1145/2043635.2043638

IONITA-LAZA, I., LANGE, C. and LAIRD, N. M. (2009). Estimating the number of unseen variants in the human genome. *Proc. Natl. Acad. Sci. USA* **106** 5008–5013. MR2496537 https://doi.org/10.1073/pnas.0807815106

JEGANATHAN, P. and HOLMES, S. P. (2021). A statistical perspective on the challenges in molecular microbial biology. *J. Agric. Biol. Environ. Stat.* **26** 131–160. MR4257010 https://doi.org/10.1007/s13253-021-00447-1

JOHNSON, N. L., KEMP, A. W. and KOTZ, S. (2005). *Univariate Discrete Distributions*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2163227 https://doi.org/10.1002/0471715816

JOHNSON, W. E. (1932). Probability: The deductive and inductive problems. *Mind* **41** 409–423.

KINGMAN, J. F. C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* (2) **18** 374–380. MR0509954 https://doi.org/10.1112/jlms/s2-18.2.374

KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1** 705–711. MR0350950 https://doi.org/10.1214/aop/1176996898

KROES, I., LEPP, P. W. and RELMAN, D. A. (1999). Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci. USA* **96** 14547–14552.

LAIRD, N. M. and LANGE, C. (2011). *The Fundamentals of Modern Statistical Genetics*. *Statistics for Biology and Health*. Springer, New York. MR2762582 https://doi.org/10.1007/978-1-4419-7338-2

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. MR2416792 https://doi.org/10.1093/biomet/asm061

LO, A. Y. (1991). A characterization of the Dirichlet process. *Statist. Probab. Lett.* **12** 185–187. MR1130354 https://doi.org/10.1016/0167-7152(91)90075-3

MAGURRAN, A. E. (2003). *Measuring Biological Diversity*. Wiley, New York.

MAO, C. X. (2004). Predicting the conditional probability of discovering a new class. *J. Amer. Statist. Assoc.* **99** 1108–1118. MR2109499 https://doi.org/10.1198/016214504000001709

MAO, C. X. and LINDSAY, B. G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika* **89** 669–681. MR1929171 https://doi.org/10.1093/biomet/89.3.669

MASOERO, L., BERAHA, M., RICHARDSON, T. S. and FAVARO, S. (2023). Improved prediction of future user activity in online A/B testing. Preprint. Available at arXiv:2402.03231.

MASOERO, L., CAMERLENGHI, F., FAVARO, S. and BRODERICK, T. (2022). More for less: Predicting and maximizing genomic variant discovery via Bayesian nonparametrics. *Biometrika* **109** 17–32. MR4374638 https://doi.org/10.1093/biomet/asab012

MCALLESTER, D. and ORTIZ, L. (2004). Concentration inequalities for the missing mass and for histogram rule error. *J. Mach. Learn. Res.* **4** 895–911. MR2076001 https://doi.org/10.1162/1532443041424292

MCALLESTER, D. and SCHAPIRE, R. E. (2000). On the convergence rate of Good–Turing estimators. In *Proceedings of the Conference on Computational Learning Theory* 1–6.

MOSSEL, E. and OHANNESSIAN, M. I. (2019). On the impossibility of learning the missing mass. *Entropy* **21**. https://doi.org/10.3390/e21010028

MOTWANI, R. and VASSILVITSKII, S. (2006). Distinct values estimators for power law distributions. In *Proceedings of the Eighth Workshop on Algorithm Engineering and Experiments and the Third Workshop on Analytic Algorithmics and Combinatorics* 230–237. SIAM, Philadelphia, PA. MR2498155

NACU, Ş. (2006). Increments of random partitions. *Combin. Probab. Comput.* **15** 589–595. MR2238047 https://doi.org/10.1017/S0963548305007455

OHANNESSIAN, M. I. and DAHLEH, M. A. (2012). Rare probability estimation under regularly varying heavy tails. In *Proceedings of the Conference on Learning Theory* **23** 2110–2124.

ORLITSKY, A., SANTHANAM, N. P. and ZHANG, J. (2003). Always good Turing: Asymptotically optimal probability estimation. *Science* **302** 427–431. MR2024675 https://doi.org/10.1126/science.1088284

ORLITSKY, A., SANTHANAM, N. P. and ZHANG, J. (2004). Universal compression of memoryless sources over unknown alphabets. *IEEE Trans. Inf. Theory* **50** 1469–1481. MR2095850 https://doi.org/10.1109/TIT.2004.830761

ORLITSKY, A., SURESH, A. T. and WU, Y. (2016). Optimal prediction of the number of unseen species. *Proc. Natl. Acad. Sci. USA* **113** 13283–13288. MR3582444 https://doi.org/10.1073/pnas.1607774113

PANANJADY, A., MUTHUKUMAR, V. and THANGARAJ, A. (2024). Just wing it: Near-optimal estimation of missing mass in a Markovian sequence. *J. Mach. Learn. Res.* **25** Paper No. 312, 43. MR4829156

PANDE, V. S., BEAUCHAMP, K. and BOWMAN, G. R. (2010). Everything you wanted to know about Markov state models but were afraid to ask. *Methods* **52** 99–105. https://doi.org/10.1016/j.ymeth.2010.06.002

PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448 https://doi.org/10.1007/BF01205234

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249 https://doi.org/10.1007/BF01213386

PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T. S. Ferguson, L. S. Shapley and J. B. MacQueen, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784 https://doi.org/10.1214/lnms/1215453576

PITMAN, J. (2003). Poisson–Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed* (D. R. Goldstein, ed.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. MR2004330 https://doi.org/10.1214/lnms/1215091133

PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. With a foreword by Jean Picard. MR2245368

PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129 https://doi.org/10.1214/aop/1024404422

POLYANSKIY, Y. and WU, Y. (2020). Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens. Preprint. Available at arXiv:1902.05616.

RAGHUNATHAN, A., VALIANT, G. and ZOU, J. (2017). Estimating the unseen from multiple populations. In *Proceedings of the International Conference on Machine Learning* **70** 2855–2863.

REGAZZINI, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilià. *G. Ist. Ital. Attuari* **41** 77–89.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 157–163. Univ. California Press, Berkeley, Calif. MR0084919

ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Stat.* **35** 1–20. MR0163407 https://doi.org/10.1214/aoms/1177703729

ROBBINS, H. E. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Stat.* **39** 256–257. MR0221695 https://doi.org/10.1214/aoms/1177698526

ROBINS, H. S., CAMPREGHER, P. V., SRIVASTAVA, S. K., WACHER, A., TURTLE, C. J., KAHSAI, O., RIDDELL, S. R., WARREN, E. H. and CARLSON, C. S. (2009). Comprehensive assessment of T-cell receptor β-chain diversity in $\alpha\beta$ T cells. *Blood* **114** 4099–4107.

SIMS, D., SUDBERY, I., ILOTT, N., HEGER, A. and PONTING, C. (2014). Sequencing depth and coverage: Key considerations in genomic analysis. *Nat. Rev. Genet.* **15** 121–132.

SKINNER, C. J. and ELLIOT, M. J. (2002). A measure of disclosure risk for microdata. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 855–867. MR1979391 https://doi.org/10.1111/1467-9868.00365

SKINNER, C. J., MARSH, C., OPENSHAW, S. and WYMER, C. (1994). Disclosure control for census microdata. *J. Off. Stat.* **10** 31–51.

SKORSKI, M. (2020a). Revisiting concentration of missing mass. Preprint. Available at arXiv:2005.10018.

SKORSKI, M. (2020b). Missing mass in Markov chains. Preprint. Available at arXiv:2001.03603.

SKORSKI, M. (2023). Bernstein-type bounds for beta distribution. *Mod. Stoch. Theory Appl.* **10** 211–228. MR4573679

TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

THISTED, R. and EFRON, B. (1987). Did Shakespeare write a newly-discovered poem? *Biometrika* **74** 445–455. MR0909350 https://doi.org/10.1093/biomet/74.3.445

THOMPSON, W. K. (2003). *Sampling Rare or Elusive Species*. Island Press, United States.

WILLENBORG, L. and DE WAAL, T. (2001). *Elements of Statistical Disclosure Control. Lecture Notes in Statistics* **155**. Springer, New York. MR1866909 https://doi.org/10.1007/978-1-4613-0121-9

WOLFER, G. and KONTOROVICH, A. (2019). Minimax learning of ergodic Markov chains. In *Algorithmic Learning Theory* 2019. *Proc. Mach. Learn. Res.* (*PMLR*) **98** [903–929], 27. MR3932874

WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory* **62** 3702–3720. MR3506758 https://doi.org/10.1109/TIT.2016.2548468

WU, Y. and YANG, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann. Statist.* **47** 857–883. MR3909953 https://doi.org/10.1214/17-AOS1665

WU, Y. and YANG, P. (2021). Polynomial methods in statistical inference: Theory and practice. *Found. Trends Commun. Inf. Theory* **17** 402–586.

ZABELL, S. L. (1992). Predicting the unpredictable. *Synthese* **90** 205–232. MR1148566 https://doi.org/10.1007/BF00485351

ZABELL, S. L. (1997). The continuum of inductive methods revisited. In *The Cosmos of Science: Essays in Exploration* (J. Earman and J. D. Norton, eds.). Univ. Pittsburgh Press, United States.

ZABELL, S. L. (2005). The continuum of inductive methods revisited. In *Symmetry and Its Discontents: Essays on the History of Inductive Probability. Cambridge Studies in Probability, Induction, and Decision Theory*. Cambridge Univ. Press, New York. With a preface by Brian Skyrms. MR2199124 https://doi.org/10.1017/CBO9780511614293

ZHANG, C.-H. (2005). Estimation of sums of random variables: Examples and information bounds. *Ann. Statist.* **33** 2022–2041. MR2211078 https://doi.org/10.1214/009053605000000390

ZHANG, C.-H. and ZHANG, Z. (2009). Asymptotic normality of a nonparametric estimator of sample coverage. *Ann*. *Statist*. **37** 2582–2595. MR2543704 https://doi.org/10.1214/08-AOS658

ZOU, J., VALIANT, G., VALIANT, P., KARCZEWSKI, K., CHAN, S. O., SAMOCHA, K., LEK, M., SUNYAEV, S., DALY, M. et al. (2016). Quantifying unobserved protein-coding variants in human populations provides a roadmap for large-scale sequencing projects. *Nat*. *Commun*. **7** 13293.

Statistical Science
2026, Vol. 41, No. 2, 247–259
https://doi.org/10.1214/24-STS962

On Matérn Covariance and Gaussian Markov Random Fields: A Spectral Analysis

Patrick E. Brown and Jamie E. Stafford

Abstract. That Gaussian Markov random fields can be used to approximate Gaussian random fields with Matérn covariances functions gained widespread attention due to the seminal work of Lindgren, Rue, and Lindström (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 423–498). This was the culmination of a rich history within the statistical literature with pivotal contributions by Whittle (*Biometrika* **41** (1954) 434–449), Moran (*J. Appl. Probab.* **10** (1973) 54–62), and Besag (*J. Roy. Statist. Soc. Ser. B* (1974) 192–236). There developments often relied on analogies to time series analysis where the Markov properties of autoregressive processes were well understood. We revisit this literature while simultaneously offering an alternative demonstration of the result by Lindgren, Rue, and Lindström (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 423–498) that also relies on time series techniques, namely, spectral analysis. New insight is gained into understanding the role of the curse of dimensionality: as dimension increases, neighbourhoods of increasing size are required in order to exploit the GMRF-GRF approximation. Here the rigour of the differential equation argument in Lindgren, Rue, and Lindström (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 423–498) is sacrificed for the sake of clarity and simplicity. The result is accessible to students enrolled in their first course in time series analysis.

Key words and phrases: Autoregression, curse of dimensionality, frequency domain, lattice processes, spatial statistics, spectral analysis.

REFERENCES

BELLMAN, R. (1961). *Adaptive Control Processes*: *A Guided Tour*. Princeton Univ. Press, Princeton, NJ. MR0134403

BESAG, J. (1981). On a system of two-dimensional recurrence equations. *J. Roy. Statist. Soc. Ser. B* **43** 302–309. MR0637942

BESAG, J. and MONDAL, D. (2005). First-order intrinsic autoregressions and the de Wijs process. *Biometrika* **92** 909–920. MR2234194 https://doi.org/10.1093/biomet/92.4.909

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822 https://doi.org/10.1007/BF00116466

BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* 192236.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis*: *Forecasting and Control*, Revised ed. *Holden-Day Series in Time Series Analysis*. Holden-Day, San Francisco, Calif.-Düsseldorf-Johannesburg. MR0436499

DE WIJS, H. J. (1951). Statistics of ore distribution. Part I: Frequency distribution of assay values. *Geol. Mijnb.* **13** 365–75.

NIST Digital Library of Mathematical Functions. (2017). Retrieved September 18, 2017, from http://dlmf.nist.gov.

EUBANK, R. L. and WANG, S. (2002). The equivalence between the Cholesky decomposition and the Kalman filter. *Amer. Statist.* **56** 39–43. MR1939395 https://doi.org/10.1198/000313002753631349

GRADSHTEYN, I. S. and RYZHIK, I. M. (1965). *Table of Integrals*, *Series*, *and Products*. Academic Press, New York. MR0197789

HARARI, O., BINGHAM, D., DEAN, A. and HIGDON, D. (2018). Computer experiments: Prediction accuracy, sample size and model complexity revisited. *Statist. Sinica* **28** 899–919. MR3791093

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

KRIGE, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afr.* **52** 119–139.

LÉVY, P. (1948). Chaînes doubles de Markoff et fonctions aléatoires de deux variables. *C. R. Acad. Sci. Paris* **226** 53–55. MR0023477

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

LITVINENKO, A., SUN, Y., GENTON, M. G. and KEYES, D. E.

Patrick E. Brown is Associate Professor, Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada (e-mail: patrick.brown@utoronto.ca). Jamie E. Stafford is Professor, Department of Statistical Sciences, University of Toronto, (e-mail: jamie.stafford@utoronto.ca).

(2019). Likelihood approximation with hierarchical matrices for large spatial datasets. *Comput*. *Statist*. *Data Anal*. **137** 115–132. MR3921064 https://doi.org/10.1016/j.csda.2019.02.002

MATÉRN, B. (1947). Metoder att uppskatta noggrannheten vid linje- och provytetaxering. (Tech. Rep. No. 1). Meddelanden frân Statens Skogsforskningsinstitut.

MATÉRN, B. (1960). *Spatial Variation*: *Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Statens Skogsforskningsinstitut, Stockholm. Meddelanden Fran Statens Skogsforskningsinstitut, Band 49, Nr. 5. MR0169346

MORAN, P. A. P. (1973). A Gaussian Markovian process on a square lattice. *J*. *Appl*. *Probab*. **10** 54–62. MR0353437 https://doi.org/10.2307/3212495

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. *Vol*. 1. *Probability and Mathematical Statistics*. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York. MR0628735

RENSHAW, E. and FORD, E. D. (1983). The interpretation of process from pattern using two-dimensional spectral analysis: Methods and problems of interpretation. *J*. *Appl*. *Stat*. **51**.

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

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450 https://doi.org/10.1093/biomet/41.3-4.434

WHITTLE, P. (1963). Stochastic processes in several dimensions. *Bull*. *Inst*. *Int*. *Stat*. **40** 974–994. MR0173287

Statistical Science
2026, Vol. 41, No. 2, 260–280
https://doi.org/10.1214/24-STS964

Overlap Measures Against ROC Summary Indices

M. Carmen Pardo and Alba M. Franco-Pereira

Abstract. Several indices of accuracy have been proposed to summarize ROC curves. Based on these indices, statistical tests have been developed to compare the accuracy of diagnostic tests. We briefly review three of them, *AUC*, Youden index, and the length of the ROC curve. Besides, we explore overlap measures as alternative measures for assessing the effectiveness of a diagnostic marker. Both parametric and nonparametric estimators as well as different methods for constructing confidence intervals are proposed for these measures. We show that they are related to distances and divergences and provide some properties. We then identify situations in which the overlap measures outperform the ROC summary indices through a simulation study. Furthermore, we compare the methods for constructing confidence intervals in terms of coverage and width. Our approaches are illustrated using a real data set.

Key words and phrases: ROC curve, *AUC*, Youden index, length of ROC curve, Weitzman measure, Kullback–Leibler measure, Matusita measure, bootstrap, kernel.

REFERENCES

AL-SALEH, M. and SAMAWI, H. (2007). Inference on overlapping coefficients in two exponential populations. *J. Mod. Stat. Sci.* **6** 503–516.

ALONSO, R., NAKAS, C. T. and PARDO, M. C. (2020). A study of indices useful for the assessment of diagnostic markers in non-parametric ROC curve analysis. *Comm. Statist. Simulation Comput.* **49** 2102–2113. MR4143435 https://doi.org/10.1080/03610918.2018.1511806

BANTIS, L. E., NAKAS, C. T. and REISER, B. (2014). Construction of confidence regions in the ROC space after the estimation of the optimal Youden index-based cut-off point. *Biometrics* **70** 212–223. MR3251682 https://doi.org/10.1111/biom.12107

BANTIS, L. E., NAKAS, C. T. and REISER, B. (2019). Construction of confidence intervals for the maximum of the Youden index and the corresponding cutoff point of a continuous biomarker. *Biom. J.* **61** 138–156. MR3899470 https://doi.org/10.1002/bimj.201700107

BANTIS, L. E., TSIMIKAS, J. V., CHAMBERS, G. R., CAPELLO, M., HANASH, S. and FENG, Z. (2021). The length of the receiver operating characteristic curve and the two cutoff Youden index within a robust framework for discovery, evaluation, and cutoff estimation in biomarker studies involving improper receiver operating characteristic curves. *Stat. Med.* **40** 1767–1789. MR4229801 https://doi.org/10.1002/sim.8869

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. (With discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252. MR0192611

CLEMONS, T. E. and BRADLEY, E. L. (2000). A nonparametric measure of the overlapping coefficient. *Comput. Statist. Data Anal.* **34** 51–61.

CARVALHO, M., BARNEY, B. J. and PAGE, G. L. (2020). Affinity-based measures of biomarker performance evaluation. *Stat. Methods Med. Res.* **29** 837–853. MR4078252 https://doi.org/10.1177/0962280219846157

DHAKER, H., NGOM, P., IBRAHIMOU, B. and MBODJI, M. (2019). Overlap coefficients based on Kullback-Leibler of two normal densities: Equal means case. *J. Math. Res.* **11** 114–124.

DHAKER, H., NGOM, P. and MBODJI, M. (2017). Overlap coefficients based on Kullback-Leibler: Exponential populations case. *Int. J. Appl. Math. Res.* **6** 135–140.

FARAGGI, D. and REISER, B. (2002). Estimation of the area under the ROC curve. *Stat. Med.* **21** 3093–3106.

FRANCO-PEREIRA, A. M., NAKAS, C. T. and PARDO, M. C. (2020). Biomarker assessment in ROC curve analysis using the length of the curve as an index of diagnostic accuracy: The binormal model framework. *AStA Adv. Stat. Anal.* **104** 625–647. MR4173397 https://doi.org/10.1007/s10182-020-00371-8

FRANCO-PEREIRA, A. M., NAKAS, C. T., REISER, B. and PARDO, M. C. (2021). Inference on the overlap coefficient: The binormal approach and alternatives. *Stat. Methods Med. Res.* **30** 2672–2684. MR4345219 https://doi.org/10.1177/09622802211046386

M. Carmen Pardo is Professor, Department of Statistics and O.R., Complutense University of Madrid, 28040-Madrid, Spain; Instituto de Matemática Interdisciplinar (IMI), Complutense University of Madrid, 28040-Madrid, Spain (e-mail: mcapardo@ucm.es). Alba M. Franco-Pereira is Associate Professor, Department of Statistics and O.R., Complutense University of Madrid, 28040-Madrid, Spain; Instituto de Matemática Interdisciplinar (IMI), Complutense University of Madrid, 28040-Madrid, Spain (e-mail: albfranc@ucm.es).

Hall, P., Marron, J. S. and Park, B. U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* **92** 1–20. MR1156447 https://doi.org/10.1007/BF01205233

Inman, H. F. (1984). *Behavior and Properties of the Overlapping Coefficient as a Measure of Agreement Between Distributions*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—The University of Alabama at Birmingham. MR2633845

Knight, K. (2000). *Mathematical Statistics*. *Chapman & Hall/CRC Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR1739674

Kullback, S. (1967). A lower bound for discrimination in terms of variation. *IEEE Trans*. *Inf. Theory* **4** 126–127.

Le Cam, L. M. (1969). *Théorie Asymptotique de la Décision Statistique*. Les Presses de l'Université de Montréal, Montreal, QC. MR0260085

Lloyd, C. J. and Yong, Z. (1999). Kernel estimators of the ROC curve are better than empirical. *Statist*. *Probab*. *Lett*. **44** 221–228. MR1711637 https://doi.org/10.1016/S0167-7152(99)00012-7

Martínez-Camblor, P., De Uña-Álvarez, J. and Corral, N. (2008). k-sample test based on the common area of kernel density estimators. *J*. *Statist*. *Plann*. *Inference* **138** 4006–4020. MR2455983 https://doi.org/10.1016/j.jspi.2008.02.008

Matusita, K. (1955). Decision rules, based on the distance for problems of fit, two samples, and estimation. *Ann*. *Math*. *Stat*. **26** 631–640. MR0073899 https://doi.org/10.1214/aoms/1177728422

Mulekar, M. S. and Mishra, S. N. (1994). Overlap coefficients of two normal densities: Equal means case. *J*. *Japan Statist*. *Soc*. **24** 169–180. MR1326528

Mulekar, M. S. and Pereira De Andrade, N. (2017). A necessary condition for computation of the overlap measure. *Amer. J. Math*. *Management Sci*. **36** 112–117.

Nakas, C., Yiannoutsos, C. T., Bosch, R. J. and Moyssiadis, C. (2003). Assessment of diagnostic markers by goodness-of-fit tests. *Stat*. *Med*. **10** 2503–2513.

Nielsen, F. (2013). Jeffreys centroids: A closed-form expression for positive histograms and a guaranteed tight approximation for frequency histograms. *IEEE Signal Process*. *Lett*. **20** 657–660.

Österreicher, F. (1996). On a class of perimeter-type distances of probability distributions. *Kybernetika* (*Prague*) **32** 389–393. MR1420130

Österreicher, F. and Vajda, I. (2003). A new class of metric divergences on probability spaces and its applicability in statistics. *Ann*. *Inst*. *Statist*. *Math*. **55** 639–653. MR2007803 https://doi.org/10.1007/BF02517812

Pardo, M. C. and Franco-Pereira, A. M. (2017). Non parametric ROC summary statistics. *REVSTAT* **15** 583–600. MR3719217

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

Samawi, H. M., Yin, J., Rochani, H. and Panchal, V. (2017). Notes on the overlap measure as an alternative to the Youden index: How are they related? *Stat*. *Med*. **36** 4230–4240. MR3713661 https://doi.org/10.1002/sim.7435

Schmid, F. and Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping—theory and empirical application. *Comput*. *Statist*. *Data Anal*. **50** 1583–1596. MR2222059 https://doi.org/10.1016/j.csda.2005.01.014

Shapiro, D. E. (1999). The interpretation of diagnostic tests. *Stat*. *Methods Med*. *Res*. **8** 113–134. https://doi.org/10.1177/096228029900800203

Silva-Fortes, C., Turkman, M. A. and Sousa, L. (2012). Arrow plot: A new graphical tool for selecting up and down regulated genes and genes differentially expressed on sample subgroups. *BMC Bioinform*. **13** 147–148.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0848134 https://doi.org/10.1007/978-1-4899-3324-9

Su, H., Hu, N., Yang, H. H., Wang, C., Takikita, M., Wang, Q. H., Giffen, C., Clifford, R., Hewitt, S. M. et al. (2011). Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin*. *Cancer Res*. **17** 2955–2966.

Weitzman, M. L. (1970). *Measures of Overlap of Income Distributions of White and Negro Families in the United States* **22**. US Bureau of the Census.

Yin, J., Hao, Y., Samawi, H. and Rochani, H. (2016). Rank-based kernel estimation of the area under the ROC curve. *Stat*. *Methodol*. **32** 91–106. MR3529719 https://doi.org/10.1016/j.stamet.2016.04.001

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* **3** 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

Zou, K. H., Liu, A., Bandos, A. I., Ohno-Machado, L. and Rockette, H. E. (2012). *Statistical Evaluation of Diagnostic Performance*: *Topics in ROC Analysis*. *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. MR3052764

Statistical Science
2026, Vol. 41, No. 2, 281–302
https://doi.org/10.1214/24-STS967

Evidence Bounds in Singular Models: Probabilistic and Variational Perspectives

Anirban Bhattacharya, Debdeep Pati, Sean Plummer and Yun Yang

Abstract. In Bayesian statistics, the marginal likelihood, a.k.a. the evidence, contains an intrinsic penalty accounting for larger model sizes and is of fundamental importance in Bayesian model comparison. Over the past two decades, there has been steadily increasing activity to understand the nature of this penalty in singular statistical models, building on pioneering works by Sumio Watanabe. Unlike regular models where the Bayesian information criterion (BIC) encapsulates a first-order expansion of the evidence, parameter counting gets trickier in singular models where a quantity called the real log-canonical threshold (RLCT) summarizes the effective model dimensionality. In this article, we offer a probabilistic treatment to recover nonasymptotic versions of established evidence bounds as well as prove a new result based on the Gibbs variational inequality. In particular, we show that mean-field variational inference correctly recovers the RLCT for any singular model in its standard form. We additionally exhibit sharpness of our bound empirically in dimension $d = 2$ and provide two conjectures concerning the asymptotics of the mean-field ELBO for singular models in standard form.

Key words and phrases: Bayesian model selection, coordinate ascent, Gibbs variational inequality, Laplace approximation, mean-field approximation, real log-canonical threshold.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions*: *With Formulas*, *Graphs*, *and Mathematical Tables* **55**. Courier Corporation.

[2] ACQUISTAPACE, F., BROGLIA, F. and FERNANDO, J. F. (2022). *Topics in Global Real Analytic Geometry*. *Springer Monographs in Mathematics*. Springer, Cham. MR4461035 https://doi.org/10.1007/978-3-030-96666-9

[3] AOYAGI, M. (2010). Stochastic complexity and generalization error of a restricted Boltzmann machine in Bayesian estimation. *J. Mach. Learn. Res.* **11** 1243–1272. MR2645451

[4] AOYAGI, M. (2019). Learning coefficient of Vandermonde matrix-type singularities in model selection. *Entropy* **21** Paper No. 561. MR3978855 https://doi.org/10.3390/e21060561

[5] AOYAGI, M. and WATANABE, S. (2005). Stochastic complexities of reduced rank regression in Bayesian estimation. *Neural Netw.* **18** 924–933. https://doi.org/10.1016/j.neunet.2005.03.014

[6] ARNOLD, V. I., VARCHENKO, A. N. and GUSEIN-ZADE, S. M. (2012). *Singularities of Differentiable Maps*: *Volume II Monodromy and Asymptotic Integrals* **83**. Springer Science & Business Media.

[7] BIBINGER, M. (2013). Notes on the sum and maximum of independent exponentially distributed random variables with different scale parameters. arXiv preprint. Available at arXiv:1307.3945.

[8] 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

[9] 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

[10] COOPER, J. L. B. (1966). Laplace transformations of distributions. *Canad. J. Math.* **18** 1325–1332. MR0203376 https://doi.org/10.4153/CJM-1966-130-6

[11] COX, D. A., LITTLE, J. and O'SHEA, D. (2015). *Ideals*, *Varieties*, *and Algorithms*: *An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed. *Undergraduate Texts in Mathematics*. Springer, Cham. MR3330490 https://doi.org/10.1007/978-3-319-16721-3

[12] DRTON, M., LIN, S., WEIHS, L. and ZWIERNIK, P. (2017). Marginal likelihood and model selection for Gaussian latent tree and forest models. *Bernoulli* **23** 1202–1232. MR3606764 https://

Anirban Bhattacharya is a Professor, Department of Statistics, Texas A&M University, College Station, Texas 77843, USA (e-mail: anirbanb@stat.tamu.edu). Debdeep Pati is a Professor, Department of Statistics, University of Wisconson-Madison, Madison, Wisconsin 53706, USA (e-mail: dpati2@wisc.edu). Sean Plummer is an Assistant Professor, Department of Mathematical Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA (e-mail: seanp@uark.edu). Yun Yang is an Associate Professor, Department of Mathematics, University of Maryland, Maryland 20742, USA (e-mail: yy84@umd.edu).

doi.org/10.3150/15-BEJ775

[13] DRTON, M. and PLUMMER, M. (2017). A Bayesian information criterion for singular models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 323–380. MR3611750 https://doi.org/10.1111/rssb.12187

[14] FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics (New York)*. Wiley, New York. MR1681462

[15] FRIEDLANDER, F. G., JOSHI, M. S., JOSHI, M. and JOSHI, M. C. (1998). *Introduction to the Theory of Distributions*. Cambridge University Press.

[16] FULTON, W. (1989). *Algebraic Curves*. *Advanced Book Classics: An Introduction to Algebraic Geometry*. Addison-Wesley Company, Advanced Book Program, Redwood City, CA. MR1042981

[17] GHORBANI, B., JAVADI, H. and MONTANARI, A. (2018). An instability in variational inference for topic models. In *International Conference on Machine Learning*.

[18] HAYASHI, N. and WATANABE, S. (2017). Tighter upper bound of real log canonical threshold of non-negative matrix factorization and its application to Bayesian inference. In 2017 *IEEE Symposium Series on Computational Intelligence (SSCI)* 1–8. IEEE Press, New York.

[19] HOSINO, T., WATANABE, K. and WATANABE, S. (2005). Stochastic complexity of variational Bayesian hidden Markov models. In *Proceedings*. 2005 *IEEE International Joint Conference on Neural Networks*, 2005. **2** 1114–1119. IEEE Press, New York.

[20] HOSINO, T., WATANABE, K. and WATANABE, S. (2006). Free energy of stochastic context free grammar on variational Bayes. In *ICONIP*.

[21] KASS, R. E., TIERNEY, L. and KADANE, J. B. (1990). The validity of posterior expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* (S. Geisser, J. Hodges, S. Press and A. Zellner, eds.) Elsevier Science Publishers B. V., North Holland.

[22] LEVY, E. (2022). On the density for sums of independent exponential, Erlang and gamma variates. *Statist. Papers* **63** 693–721. MR4426413 https://doi.org/10.1007/s00362-021-01256-x

[23] LIN, S. (2011). Algebraic methods for evaluating integrals in Bayesian statistics. PhD thesis, UC Berkeley.

[24] MACKAY, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge Univ. Press, New York. MR2012999

[25] MATHAI, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Ann. Inst. Statist. Math.* **34** 591–597. MR0695077 https://doi.org/10.1007/BF02481056

[26] MUKHERJEE, S. S., SARKAR, P., WANG, Y. R. and YAN, B. (2018). Mean field for the stochastic blockmodel: Optimization landscape and convergence issues. In *Advances in Neural Information Processing Systems* 10694–10704.

[27] NAKAJIMA, S., SATO, I., SUGIYAMA, M., WATANABE, K. and KOBAYASHI, H. (2014). Analysis of variational Bayesian latent Dirichlet allocation: Weaker sparsity than MAP. *Adv. Neural Inf. Process. Syst.* **27**.

[28] NAKAJIMA, S., WATANABE, K. and SUGIYAMA, M. (2019). *Variational Bayesian Learning Theory*. Cambridge Univ. Press, Cambridge.

[29] NAKAJIMA, S. and WATANABE, S. (2007). Variational Bayes solution of linear neural networks and its generalization performance. *Neural Comput.* **19** 1112–1153. MR2303391 https://doi.org/10.1162/neco.2007.19.4.1112

[30] PATI, D., BHATTACHARYA, A. and YANG, Y. (2018). On statistical optimality of variational Bayes. In *International Conference on Artificial Intelligence and Statistics* 1579–1588.

[31] ROBERT, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR2723361

[32] RUSAKOV, D. and GEIGER, D. (2005). Asymptotic model selection for naive Bayesian networks. *J. Mach. Learn. Res.* **6** 1–35. MR2249813

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

[34] SMAILI, K., KADRI, T. and KADRY, S. (2013). Hypoexponential distribution with different parameters.

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

[36] VAN DER VAART, A. W. and WELLNER, J. A. (1996). Weak convergence. In *Weak Convergence and Empirical Processes* 16–28. Springer.

[37] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

[38] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers, Hanover.

[39] WANG, B. and TITTERINGTON, D. M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Anal.* **1** 625–649. MR2221291 https://doi.org/10.1214/06-BA121

[40] WATANABE, K., SHIGA, M. and WATANABE, S. (2009). Upper bound for variational free energy of Bayesian networks. *Mach. Learn.* **75** 199–215.

[41] WATANABE, K. and WATANABE, S. (2004). Lower bounds of stochastic complexities in variational Bayes learning of Gaussian mixture models. In *IEEE Conference on Cybernetics and Intelligent Systems* **1** 99–104.

[42] WATANABE, K. and WATANABE, S. (2006). Stochastic complexities of Gaussian mixtures in variational Bayesian approximation. *J. Mach. Learn. Res.* **7** 625–643. MR2274381

[43] WATANABE, K. and WATANABE, S. (2007). Stochastic complexities of general mixture models in variational Bayesian learning. *Neural Netw.* **20** 210–219.

[44] WATANABE, K. and WATANABE, S. (2007). Stochastic complexity for mixture of exponential families in generalized variational Bayes. *Theoret. Comput. Sci.* **387** 4–17. MR2359010 https://doi.org/10.1016/j.tcs.2007.07.039

[45] WATANABE, S. (1999). Algebraic analysis for singular statistical estimation. In *Algorithmic Learning Theory (Tokyo*, 1999). *Lecture Notes in Computer Science* **1720** 39–50. Springer, Berlin. MR1788519 https://doi.org/10.1007/3-540-46769-6_4

[46] WATANABE, S. (2001). Algebraic analysis for nonidentifiable learning machines. *Neural Comput.* **13** 899–933. https://doi.org/10.1162/089976601300014402

[47] WATANABE, S. (2001). Algebraic geometrical methods for hierarchical learning machines. *Neural Netw.* **14** 1049–1060. https://doi.org/10.1016/s0893-6080(01)00069-7

[48] WATANABE, S. (2007). Almost all learning machines are singular. In 2007 *IEEE Symposium on Foundations of Computational Intelligence* 383–388.

[49] WATANABE, S. (2009). *Algebraic Geometry and Statistical Learning Theory*. *Cambridge Monographs on Applied and Computational Mathematics* **25**. Cambridge Univ. Press, Cambridge. MR2554932 https://doi.org/10.1017/CBO9780511800474

[50] WATANABE, S. (2013). A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14** 867–897. MR3049492

[51] WATANABE, S. (2018). *Mathematical Theory of Bayesian Statistics*. CRC Press, Boca Raton, FL. MR3839272 https://doi.org/10.1201/9781315373010

[52] WEI, S. and LAU, E. (2024). Variational Bayesian neural networks via resolution of singularities. *J. Comput. Graph. Statist.* **33** 1464–1474. MR4829651 https://doi.org/10.1080/10618600.2024.2325455

[53] YAMAZAKI, K. and WATANABE, S. (2003). Singularities in mixture models and upper bounds of stochastic complexity. *Neural Netw.* **16** 1029–1038.

[54] 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

[55] ZEMANIAN, A. H. (1987). *Distribution Theory and Transform Analysis*: *An Introduction to Generalized Functions*, *with Applications*, 2nd ed. Dover, New York. MR0918977

[56] ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann. Statist.* **48** 2575–2598. MR4152113 https://doi.org/10.1214/19-AOS1898

Statistical Science
2026, Vol. 41, No. 2, 303–315
https://doi.org/10.1214/24-STS968

Doubly Ranked Tests of Location for Grouped Functional Data

Mark J. Meyer

Abstract. Nonparametric tests for functional data are a challenging class of tests to work with because of the potentially high-dimensional nature of the data. One of the main challenges for considering rank-based tests, like the Mann–Whitney or Wilcoxon Rank Sum tests (MWW), is that the unit of observation is typically a curve. Thus, any rank-based test must consider ways of ranking curves. While several procedures, including depth-based methods, have recently been used to create scores for rank-based tests, these scores are not constructed under the null and often introduce additional, uncontrolled for variability. We therefore reconsider the problem of rank-based tests for functional data and develop an alternative approach that incorporates the null hypothesis throughout. Our approach first ranks realizations from the curves at each measurement occurrence, then calculates a summary statistic for the ranks of each subject, and finally reranks the summary statistic in a procedure we refer to as a doubly ranked test. We propose two summaries for the middle step: a sufficient statistic and the average rank. As we demonstrate, doubly rank tests are more powerful while maintaining ideal type I error in the two sample, MWW setting. We also extend our framework to more than two samples, developing a Kruskal–Wallis test for functional data, which exhibits good test characteristics as well. Finally, we illustrate the use of doubly ranked tests in functional data contexts from material science, climatology and public health policy.

Key words and phrases: Nonparametric functional data analysis, Wilcoxon rank sum, Mann–Whitney test, Kruskal–Wallis test, order statistics, sufficient statistics.

REFERENCES

ABRAMOWICZ, K., HÄGER, C. K., PINI, A., SCHELIN, L., SJÖSTEDT DE LUNA, S. and VANTINI, S. (2018). Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scand. J. Stat.* **45** 1036–1061. MR3884899 https://doi.org/10.1111/sjos.12333

ALEMÁN-GÓMEZ, Y., ARRIBAS-GIL, A., DESCO, M., ELÍAS, A. and ROMO, J. (2022). Depthgram: Visualizing outliers in high-dimensional functional data with application to fMRI data exploration. *Stat. Med.* **41** 2005–2024. MR4411871 https://doi.org/10.1002/sim.9342

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* **49** 2457–2490. MR4338371 https://doi.org/10.1214/20-aos2041

BROCKHAUS, S., RÜGAMER, D. and GREVEN, S. (2020). Boosting functional regression models with FDboost. *J. Stat. Softw.* **94** 1–50. https://doi.org/10.18637/jss.v094.i10

CHAKRABORTY, A. and CHAUDHURI, P. (2015). A Wilcoxon–Mann–Whitney-type test for infinite-dimensional data. *Biometrika* **102** 239–246. MR3335109 https://doi.org/10.1093/biomet/asu072

CUESTA-ALBERTOS, J. A. and FEBRERO-BANDE, M. (2010). A simple multiway ANOVA for functional data. *TEST* **19** 537–557. MR2746001 https://doi.org/10.1007/s11749-010-0185-3

DAI, W. and GENTON, M. G. (2018). Multivariate functional data visualization and outlier detection. *J. Comput. Graph. Statist.* **27** 923–934. MR3890881 https://doi.org/10.1080/10618600.2018.1473781

DAI, W., MRKVIČKA, T., SUN, Y. and GENTON, M. G. (2020). Functional outlier detection and taxonomy by sequential transformations. *Comput. Statist. Data Anal.* **149** 106960, 17 pp. MR4087190 https://doi.org/10.1016/j.csda.2020.106960

DI, C.-Z., CRAINICEANU, C. M., CAFFO, B. S. and PUNJABI, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat.* **3** 458–488. MR2668715 https://doi.org/10.1214/08-AOAS206

GASSEN, J. (2022). Download Apple mobility trend reports data. *tidycovid19*. Available at https://joachim-gassen.github.io/tidycovid19/

Mark J. Meyer is an Associate Professor, Department of Mathematics and Statistics at Georgetown University, Washington, DC, USA (e-mail: mjm556@georgetown.edu).

reference/download_apple_mtr_data.html. Accessed June, 2023.

GENTON, M. G., JOHNSON, C., POTTER, K., STENCHIKOV, G. and SUN, Y. (2014). Surface boxplots. *Stat* **3** 1–11. MR4027321 https://doi.org/10.1002/sta4.39

GOLDSMITH, J., SCHEIPL, F., HUANG, L., WROBEL, J., DI, C., GELLAR, J., HAREZLAK, J., MCLEAN, M. W., SWIHART, B. et al. (2022). refund: Regression with functional data. R package version 0.1-28.

GREVEN, S. and SCHEIPL, F. (2017). A general framework for functional regression modelling. *Stat. Model.* **17** 1–35. MR3619335 https://doi.org/10.1177/1471082X16681317

HALL, P. and VAN KEILEGOM, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statist. Sinica* **17** 1511–1531. MR2413533

HOLLANDER, M. and WOLFE, D. A. (1999). *Nonparametric Statistical Methods*, 2nd ed. *Wiley Series in Probability and Statistics*: *Texts and References Section*. Wiley, New York. MR1666064

HUANG, H. and SUN, Y. (2019). A decomposition of total variation depth for understanding functional outliers. *Technometrics* **61** 445–458. MR4031498 https://doi.org/10.1080/00401706.2019.1574241

KATZ, R., BOYCE, M. and GRAEDON, E. (2023). Visualizing the impact of policies on COVID response. *COVID AMP*. Available at https://covidamp.org/. Accessed June, 2023.

KLOKE, J. and MCKEAN, J. W. (2015). *Nonparametric Statistical Methods Using R*. CRC Press, Taylor & Francis Group, Boca Raton, FL.

KRUSKAL, W. H. and WALLIS, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Assoc.* **47** 583–621. https://doi.org/10.2307/2280779

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590 https://doi.org/10.1198/jasa.2009.0108

LÓPEZ-PINTADO, S. and ROMO, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.* **55** 1679–1695. MR2748671 https://doi.org/10.1016/j.csda.2010.10.024

LÓPEZ-PINTADO, S., ROMO, J. and TORRENTE, A. (2010). Robust depth-based tools for the analysis of gene expression data. *Biostatistics* **11** 254–264. https://doi.org/10.1093/biostatistics/kxp056

LÓPEZ-PINTADO, S. and WROBEL, J. (2017). Robust non-parametric tests for imaging data based on data depth. *Stat* **6** 405–419. MR3722938 https://doi.org/10.1002/sta4.168

MANN, H. B. and WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18** 50–60. MR0022058 https://doi.org/10.1214/aoms/1177730491

MELÉNDEZ, R., GIRALDO, R. and LEIVA, V. (2021). Sign, Wilcoxon and Mann–Whitney tests for functional data: An approach based on random projections. *Mathematics* **9** 44. https://doi.org/10.3390/math9010044

MEYER, M. J. (2024). runDRT: Run doubly ranked tests. R package version 0.1.0.

MEYER, M. J. (2026a). Supplement A to "Doubly ranked tests of location for grouped functional data." https://doi.org/10.1214/24-STS968SUPPA

MEYER, M. J. (2026b). Supplement B to "Doubly ranked tests of location for grouped functional data." https://doi.org/10.1214/24-STS968SUPPB

MORRIS, J. S. (2015). Functional regression. *Annu. Rev. Stat. Appl.* **2** 321–359. https://doi.org/10.1146/annurev-statistics-010814-020413

POMANN, G.-M., STAICU, A.-M. and GHOSH, S. (2016). A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65** 395–414. MR3470583 https://doi.org/10.1111/rssc.12130

RAMSAY, J. O., GRAVES, S. and HOOKER, G. (2022). fda: Functional data analysis. R package version 6.0.5.

RAMSAY, J. O., HOOKER, G. and GRAVES, S. (2009). *Functional Data Analysis in R and Matlab*. Springer, New York.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

SGUERA, C. and LÓPEZ-PINTADO, S. (2021). A notion of depth for sparse functional data. *TEST* **30** 630–649. MR4297271 https://doi.org/10.1007/s11749-020-00734-y

SUN, Y. and GENTON, M. G. (2011). Functional boxplots. *J. Comput. Graph. Statist.* **20** 316–334. MR2847798 https://doi.org/10.1198/jcgs.2011.09224

SUN, Y. and GENTON, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics* **23** 54–64. MR2873783 https://doi.org/10.1002/env.1136

SUN, Y., GENTON, M. G. and NYCHKA, D. W. (2012). Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat* **1** 68–74. MR4027415 https://doi.org/10.1002/sta4.8

R CORE TEAM (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

WANG, J. L., CHIOU, J. M. and MÜLLER, H. G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

WILCOXON, F. (1945). Individual comparisons by ranking methods. *Biom. Bull.* **1** 80–83. https://doi.org/10.2307/3001968

WROBEL, J., PARK, S. Y., STAICU, A. M. and GOLDSMITH, J. (2016). Interactive graphics for functional data analyses. *Stat* **5** 108–118. https://doi.org/10.1002/sta4.109

XIAO, L., LI, Y. and RUPPERT, D. (2013). Fast bivariate P-splines: The sandwich smoother. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 577–599. MR3065480 https://doi.org/10.1111/rssb.12007

XIAO, L., ZIPUNNIKOV, V., RUPPERT, D. and CRAINICEANU, C. (2016). Fast covariance estimation for high-dimensional functional data. *Stat. Comput.* **26** 409–421. MR3439382 https://doi.org/10.1007/s11222-014-9485-x

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

Statistical Science
2026, Vol. 41, No. 2, 316–353
https://doi.org/10.1214/24-STS969

A Primer on Bayesian Neural Networks: Review and Debates

Julyan Arbel, Konstantinos Pitas, Mariia Vladimirova and Vincent Fortuin

Abstract. Neural networks have achieved remarkable performance across various problem domains, but their widespread applicability is hindered by inherent limitations such as overconfidence in predictions, lack of interpretability and vulnerability to adversarial attacks. To address these challenges, Bayesian neural networks (BNNs) have emerged as a compelling extension of conventional neural networks, integrating uncertainty estimation into their predictive capabilities.

This comprehensive primer presents a systematic introduction to the fundamental concepts of neural networks and Bayesian inference, elucidating their synergistic integration for the development of BNNs. The target audience comprises statisticians with a potential background in Bayesian methods but lacking deep learning expertise, as well as machine learners proficient in deep neural networks but with limited exposure to Bayesian statistics. We provide an overview of commonly employed priors, examining their impact on model behavior and performance. Additionally, we delve into the practical considerations associated with training and inference in BNNs.

Furthermore, we explore advanced topics within the realm of BNN research, acknowledging the existence of ongoing debates and controversies. By offering insights into cutting-edge developments, this primer not only equips researchers and practitioners with a solid foundation in BNNs, but also illuminates the potential applications of this dynamic field. As a valuable resource, it fosters an understanding of BNNs and their promising prospects, facilitating further advancements in the pursuit of knowledge and innovation.

Key words and phrases: Approximate Bayesian inference, deep learning, neural networks.

REFERENCES

[1] ABDAR, M., POURPANAH, F., HUSSAIN, S., REZAZADEGAN, D., LIU, L., GHAVAMZADEH, M., FIEGUTH, P., CAO, X., KHOSRAVI, A. et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion*.

[2] ABDELFATTAH, M. S., MEHROTRA, A., DUDZIAK, Ł. and LANE, N. D. (2021). Zero-cost proxies for lightweight NAS. arXiv preprint. Available at arXiv:2101.08134.

[3] ADLAM, B., LEE, J., XIAO, L., PENNINGTON, J. and SNOEK, J. (2020). Exploring the uncertainty properties of neural networks' implicit priors in the infinite-width limit. In *International Conference on Learning Representations*.

[4] ADLAM, B., SNOEK, J. and SMITH, S. L. (2020). Cold posteriors and aleatoric uncertainty. In *ICML Workshop on Uncertainty & Robustness in Deep Learning*.

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

[6] AITCHISON, L. (2020). Why bigger is not always better: On finite and infinite neural networks. In *International Conference on Machine Learning*.

[7] AITCHISON, L. (2021). A statistical theory of cold posteriors in deep neural networks. In *International Conference on Learning Representations*.

[8] AITCHISON, L., YANG, A. and OBER, S. W. (2021). Deep kernel processes. In *International Conference on Machine Learning*.

[9] ANTOGNINI, J. M. (2019). Finite size corrections for neural network Gaussian processes. In *ICML Workshop on Theoretical*

Julyan Arbel is an associate researcher, Centre Inria de l'Université Grenoble Alpes, Grenoble, France (e-mail: julyan.arbel@inria.fr). Konstantinos Pitas was a postdoctoral fellow at the time this research was conducted, Centre Inria de l'Université Grenoble Alpes, Grenoble, France (e-mail: pitas.konstantinos@inria.fr). Mariia Vladimirova is a senior researcher, Criteo AI Lab, Paris, France (e-mail: m.vladimirova@criteo.com). Vincent Fortuin is a principal investigator, Helmholtz AI and TU Munich, Munich, Germany (e-mail: vincent.fortuin@tum.de).

Physics for Deep Learning.

[10] ANTORÁN, J., JANZ, D., ALLINGHAM, J. U., DAXBERGER, E., BARBANO, R. R., NALISNICK, E. and HERNÁNDEZ-LOBATO, J. M. (2022). Adapting the linearised Laplace model evidence for modern deep learning. In *International Conference on Machine Learning*.

[11] ARORA, S., DU, S. S., HU, W., LI, Z., SALAKHUTDINOV, R. and WANG, R. (2019). On exact computation with an infinitely wide neural net. In *Advances in Neural Information Processing Systems*.

[12] ARORA, S., GE, R., NEYSHABUR, B. and ZHANG, Y. (2018). Stronger generalization bounds for deep nets via a compression approach. In *International Conference on Machine Learning*.

[13] ASHUKHA, A., LYZHOV, A., MOLCHANOV, D. and VETROV, D. (2020). Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. In *International Conference on Learning Representations*.

[14] BACHMANN, G., NOCI, L. and HOFMANN, T. (2022). How tempering fixes data augmentation in Bayesian neural networks. *Adv. Neural Inf. Process. Syst.*.

[15] BAKER, B., GUPTA, O., RASKAR, R. and NAIK, N. (2017). Accelerating neural architecture search using performance prediction. arXiv preprint. Available at arXiv:1705.10823.

[16] BAND, N., RUDNER, T. G., FENG, Q., FILOS, A., NADO, Z., DUSENBERRY, M. W., JERFEL, G., TRAN, D. and GAL, Y. (2021). Benchmarking Bayesian deep learning on diabetic retinopathy detection tasks. In *NeurIPS Workshop on Distribution Shifts: Connecting Methods and Applications*.

[17] BARBER, D. and BISHOP, C. M. (1998). Ensemble learning in Bayesian neural networks. In *Generalization in Neural Networks and Machine Learning* 215–237. Springer, Berlin.

[18] BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.

[19] BARTLETT, P. L., FOSTER, D. J. and TELGARSKY, M. J. (2017). Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems*.

[20] BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

[21] BEKASOV, A. and MURRAY, I. (2018). Bayesian adversarial spheres: Bayesian inference and adversarial examples in a noiseless setting. In *NeurIPS Workshop on Bayesian Deep Learning*.

[22] BEKASOVS, A. and MURRAY, I. (2018). Bayesian adversarial spheres: Bayesian inference and adversarial examples in a noiseless setting. In *Third Workshop on Bayesian Deep Learning* 2018.

[23] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

[24] BIBI, A., ALFADLY, M. and GHANEM, B. (2018). Analytic expressions for probabilistic moments of PL-DNN with Gaussian input. In *Computer Vision and Pattern Recognition*.

[25] 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.

[26] 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

[27] BLAAS, A. and ROBERTS, S. J. (2021). The effect of prior Lipschitz continuity on the adversarial robustness of Bayesian neural networks. arXiv preprint. Available at arXiv:2101.02689.

[28] 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

[29] BLOMQVIST, K., KASKI, S. and HEINONEN, M. (2019). Deep convolutional Gaussian processes. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 582–597. Springer, Berlin.

[30] BLUMER, A., EHRENFEUCHT, A., HAUSSLER, D. and WARMUTH, M. K. (1989). Learnability and the Vapnik–Chervonenkis dimension. *J. Assoc. Comput. Mach.* **36** 929–965. MR1072253 https://doi.org/10.1145/76359.76371

[31] BLUNDELL, C., CORNEBISE, J., KAVUKCUOGLU, K. and WIERSTRA, D. (2015). Weight uncertainty in neural networks. In *International Conference on Machine Learning*.

[32] BOS, T. and SCHMIDT-HIEBER, J. (2022). Convergence rates of deep ReLU networks for multiclass classification. *Electron. J. Stat.* **16** 2724–2773. MR4406243 https://doi.org/10.1214/22-ejs2011

[33] BUI, T., HERNÁNDEZ-LOBATO, D., HERNANDEZ-LOBATO, J., LI, Y. and TURNER, R. (2016). Deep Gaussian processes for regression using approximate expectation propagation. In *International Conference on Machine Learning* 1472–1481. PMLR.

[34] CANDES, E. J. (1998). *Ridgelets: Theory and Applications*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. MR2698286

[35] CARBONE, G., WICKER, M., LAURENTI, L., PATANE, A., BORTOLUSSI, L. and SANGUINETTI, G. (2020). Robustness of Bayesian neural networks to gradient-based attacks. *Adv. Neural Inf. Process. Syst.* **33** 15602–15613.

[36] 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.

[37] CHATZIAFRATIS, V., NAGARAJAN, S. G. and PANAGEAS, I. (2020). Better depth-width trade-offs for neural networks through the lens of dynamical systems. In *International Conference on Machine Learning*.

[38] CHATZIAFRATIS, V., NAGARAJAN, S. G., PANAGEAS, I. and WANG, X. (2020). Depth-width trade-offs for ReLU networks via Sharkovsky's theorem. In *International Conference on Learning Representations*.

[39] CHEN, T., FOX, E. and GUESTRIN, C. (2014). Stochastic gradient Hamiltonian Monte Carlo. In *International Conference on Machine Learning*.

[40] CHO, Y. and SAUL, L. K. (2009). Kernel methods for deep learning. In *Advances in Neural Information Processing Systems*.

[41] CINQUIN, T., IMMER, A., HORN, M. and FORTUIN, V. (2022). Pathologies in priors and inference for Bayesian transformers. In *Fourth Symposium on Advances in Approximate Bayesian Inference*.

[42] CIOSEK, K., FORTUIN, V., TOMIOKA, R., HOFMANN, K. and TURNER, R. (2020). Conservative uncertainty estimation by fitting prior networks. In *International Conference on Learning Representations*.

[43] COX, R. T. (1961). *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore, MD. MR0130703

[44] CUI, T., HAVULINNA, A., MARTTINEN, P. and KASKI, S. (2022). Informative Bayesian neural network priors for weak signals. *Bayesian Anal.* **17** 1121–1151. MR4506024 https://doi.org/10.1214/21-ba1291

[45] CUTAJAR, K., BONILLA, E. V., MICHIARDI, P. and FILIPPONE, M. (2017). Random feature expansions for deep Gaussian processes. In *International Conference on Machine Learning* 884–893. PMLR.

[46] CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2** 303–314. MR1015670 https://doi.org/10.1007/BF02551274

[47] D'ANGELO, F. and FORTUIN, V. (2021). Annealed Stein variational gradient descent. In *Third Symposium on Advances in Approximate Bayesian Inference*.

[48] D'ANGELO, F. and FORTUIN, V. (2021). Repulsive deep ensembles are Bayesian. *Adv. Neural Inf. Process. Syst.* **34** 3451–3465.

[49] D'ANGELO, F., FORTUIN, V. and WENZEL, F. (2021). On Stein variational neural network ensembles. arXiv preprint. Available at arXiv:2106.10760.

[50] D'ASCOLI, S., TOUVRON, H., LEAVITT, M., MORCOS, A., BIROLI, G. and SAGUN, L. (2021). ConViT: Improving vision transformers with soft convolutional inductive biases. In *International Conference on Machine Learning*.

[51] DAI, Z., DAMIANOU, A. C., GONZÁLEZ, J. and LAWRENCE, N. D. (2016). Variational auto-encoded deep Gaussian processes. In *ICLR*.

[52] DAMIANOU, A. and LAWRENCE, N. D. (2013). Deep Gaussian processes. In *International Conference on Artificial Intelligence and Statistics*.

[53] DAVENPORT, T. and KALAKOTA, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthc. J.* **6** 94.

[54] DAXBERGER, E., KRISTIADI, A., IMMER, A., ESCHENHAGEN, R., BAUER, M. and HENNIG, P. (2021). Laplace redux-effortless Bayesian deep learning. *Adv. Neural Inf. Process. Syst.*.

[55] DAXBERGER, E., NALISNICK, E., ALLINGHAM, J. U., ANTORÁN, J. and HERNÁNDEZ-LOBATO, J. M. (2021). Bayesian deep learning via subnetwork inference. In *International Conference on Machine Learning*.

[56] 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

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

[58] DER KIUREGHIAN, A. and DITLEVSEN, O. (2009). Aleatory or epistemic? Does it matter? *Struct. Saf.* **31** 105–112.

[59] DHAHRI, R., IMMER, A., CHARPENTIER, B., GÜNNEMANN, S. and FORTUIN, V. (2024). Shaving weights with Occam's razor: Bayesian sparsification for neural networks using the marginal likelihood. arXiv preprint. Available at arXiv:2402.15978.

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

[61] DOOB, J. L. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et Ses Applications. Colloq. Internat.* **13** 23–27. CNRS, Paris. MR0033460

[62] DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., WEISSENBORN, D., ZHAI, X., UNTERTHINER, T., DEHGHANI, M., MINDERER, M., HEIGOLD, G. et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*.

[63] DUA, D. and GRAFF, C. (2017). UCI machine learning repository.

[64] DUNLOP, M. M., GIROLAMI, M. A., STUART, A. M. and TECKENTRUP, A. L. (2018). How deep are deep Gaussian processes? *J. Mach. Learn. Res.* **19** Paper No. 54, 46. MR3874162

[65] DUSENBERRY, M., JERFEL, G., WEN, Y., MA, Y., SNOEK, J., HELLER, K., LAKSHMINARAYANAN, B. and TRAN, D. (2020). Efficient and scalable Bayesian neural nets with rank-1 factors. In *International Conference on Machine Learning*.

[66] DUTORDOIR, V., HENSMAN, J., VAN DER WILK, M., EK, C. H., GHAHRAMANI, Z. and DURRANDE, N. (2021). Deep neural networks as point estimates for deep Gaussian processes. arXiv preprint. Available at arXiv:2105.04504.

[67] DUTORDOIR, V., SALIMBENI, H., HAMBRO, E., MCLEOD, J., LEIBFRIED, F., ARTEMEV, A., VAN DER WILK, M., HENSMAN, J., DEISENROTH, M. P. et al. (2021). GPflux: A library for deep Gaussian processes. arXiv preprint. Available at arXiv:2104.05674.

[68] DUTORDOIR, V., WILK, M., ARTEMEV, A. and HENSMAN, J. (2020). Bayesian image classification with deep convolutional Gaussian processes. In *International Conference on Artificial Intelligence and Statistics* 1529–1539. PMLR.

[69] DUVENAUD, D., RIPPEL, O., ADAMS, R. and GHAHRAMANI, Z. (2014). Avoiding pathologies in very deep networks. In *International Conference on Artificial Intelligence and Statistics*.

[70] DYER, E. and GUR-ARI, G. (2020). Asymptotics of wide networks from Feynman diagrams. In *International Conference on Learning Representations*.

[71] DZIUGAITE, G. K., HSU, K., GHARBIEH, W., ARPINO, G. and ROY, D. (2021). On the role of data in pac-Bayes bounds. In *International Conference on Artificial Intelligence and Statistics*.

[72] DZIUGAITE, G. K. and ROY, D. M. (2017). Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data. arXiv preprint. Available at arXiv:1703.11008.

[73] ERHAN, D., COURVILLE, A., BENGIO, Y. and VINCENT, P. (2010). Why does unsupervised pre-training help deep learning? In *International Conference on Artificial Intelligence and Statistics*.

[74] FAVARO, S., FORTINI, S. and STEFANO, P. (2020). Stable behaviour of infinitely wide deep neural networks. In *International Conference on Artificial Intelligence and Statistics*.

[75] FAVARO, S., HANIN, B., MARINUCCI, D., NOURDIN, I. and PECCATI, G. (2024). Quantitative CLTs in deep neural networks. arXiv preprint. Available at arXiv:2307.06092.

[76] FAWZI, A., MOOSAVI-DEZFOOLI, S.-M. and FROSSARD, P. (2016). Robustness of classifiers: From adversarial to random noise. *Adv. Neural Inf. Process. Syst.*.

[77] FEINMAN, R., CURTIN, R. R., SHINTRE, S. and GARDNER, A. B. (2017). Detecting adversarial samples from artifacts. arXiv preprint. Available at arXiv:1703.00410.

[78] FLAM-SHEPHERD, D., REQUEIMA, J. and DUVENAUD, D. (2017). Mapping Gaussian process priors to Bayesian neural networks. In *NeurIPS Workshop on Bayesian Deep Learning*.

[79] FLAM-SHEPHERD, D., REQUEIMA, J. and DUVENAUD, D. (2018). Characterizing and warping the function space of Bayesian neural networks. In *NeurIPS Workshop on Bayesian Deep Learning*.

[80] FOLGOC, L. L., BALTATZIS, V., DESAI, S., DEVARAJ, A., ELLIS, S., MANZANERA, O. E. M., NAIR, A., QIU, H., SCHNABEL, J. et al. (2021). Is MC Dropout Bayesian? arXiv preprint. Available at arXiv:2110.04286.

[81] Foong, A. Y., Burt, D. R., Li, Y. and Turner, R. E. (2019). Pathologies of factorised Gaussian and MC dropout posteriors in Bayesian neural networks. *Stat* **1050**.

[82] Foret, P., Kleiner, A., Mobahi, H. and Neyshabur, B. (2021). Sharpness-aware minimization for efficiently improving generalization. In *International Conference on Learning Representations*.

[83] Fort, S., Hu, H. and Lakshminarayanan, B. (2019). Deep ensembles: a loss landscape perspective. arXiv preprint. Available at arXiv:1912.02757.

[84] Fortuin, V. (2021). On the choice of priors in Bayesian deep learning. Ph.D. thesis, ETH Zurich.

[85] Fortuin, V. (2022). Priors in Bayesian deep learning: A review. *Int. Stat. Rev.* **90** 563–591. MR4524825 https://doi.org/10.1111/insr.12502

[86] Fortuin, V., Dresdner, G., Strathmann, H. and Rätsch, G. (2021). Sparse Gaussian processes on discrete domains. *IEEE Access* **9**.

[87] Fortuin, V., Garriga-Alonso, A., Wenzel, F., Rätsch, G., Turner, R., van der Wilk, M. and Aitchison, L. (2021). Bayesian neural network priors revisited. In *Symposium on Advances in Approximate Bayesian Inference*.

[88] Frankle, J. and Carbin, M. (2019). The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*.

[89] Franssen, S. and Szabó, B. (2022). Uncertainty Quantification for nonparametric regression using Empirical Bayesian neural networks. arXiv preprint. Available at arXiv:2204.12735.

[90] Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Netw.* **2** 183–192.

[91] Gal, Y. (2016). Uncertainty in deep learning Ph.D. thesis, Cambridge Univ. Press, Cambridge.

[92] Gal, Y. and Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*.

[93] Gal, Y. and Smith, L. (2018). Sufficient conditions for idealised models to have no adversarial examples: a theoretical and empirical study with Bayesian neural networks. arXiv preprint. Available at arXiv:1806.00667.

[94] Garipov, T., Izmailov, P., Podoprikhin, D., Vetrov, D. and Wilson, A. G. (2018). Loss surfaces, mode connectivity, and fast ensembling of DNNs. In *Advances in Neural Information Processing Systems*.

[95] Garriga-Alonso, A., Rasmussen, C. E. and Aitchison, L. (2019). Deep convolutional networks as shallow Gaussian processes. In *International Conference on Learning Representations*.

[96] Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P. et al. (2021). A survey of uncertainty in deep neural networks. arXiv preprint. Available at arXiv:2107.03342.

[97] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A. and Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*.

[98] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

[99] 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. arXiv preprint. Available at arXiv:2011.01808.

[100] Germain, P., Bach, F., Lacoste, A. and Lacoste-Julien, S. (2016). PAC-Bayesian theory meets Bayesian inference. *Adv. Neural Inf. Process. Syst.*.

[101] Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature* **521** 452–459.

[102] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. *Stochastic Modelling and Applied Probability* **53**. Springer, New York. Applications of Mathematics (New York). MR1999614

[103] Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics*.

[104] Goan, E. and Fookes, C. (2020). Bayesian neural networks: An introduction and survey. In *Case Studies in Applied Bayesian Data Science*. *Lecture Notes in Math.* **2259** 45–87. Springer, Cham. MR4175776 https://doi.org/10.1007/978-3-030-42553-1_3

[105] Golowich, N., Rakhlin, A. and Shamir, O. (2020). Size-independent sample complexity of neural networks. *Inf. Inference* **9** 473–504. MR4108976 https://doi.org/10.1093/imaiai/iaz007

[106] Graves, A. (2011). Practical variational inference for neural networks. *Adv. Neural Inf. Process. Syst.*.

[107] Graves, A., Mohamed, A. and Hinton, G. E. (2013). Speech recognition with deep recurrent neural networks. In *International Conference on Acoustics, Speech and Signal Processing*.

[108] Grosse, K., Pfaff, D., Smith, M. T. and Backes, M. (2019). The limitations of model uncertainty in adversarial settings. In *NeurIPS Workshop on Bayesian Deep Learning*.

[109] Gruber, S. and Buettner, F. (2022). Better uncertainty calibration via proper scores for classification and beyond. *Adv. Neural Inf. Process. Syst.* **35** 8618–8632.

[110] Grünwald, P. (2012). The safe Bayesian: Learning the learning rate via the mixability gap. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **7568** 169–183. Springer, Heidelberg. MR3042889 https://doi.org/10.1007/978-3-642-34106-9_16

[111] Gunasekar, S., Lee, J., Soudry, D. and Srebro, N. (2018). Implicit bias of gradient descent on linear convolutional networks. In *Advances in Neural Information Processing Systems*.

[112] Guo, C., Pleiss, G., Sun, Y. and Weinberger, K. Q. (2017). On calibration of modern neural networks. In *International Conference on Machine Learning*.

[113] Gurbuzbalaban, M., Şimşekli, U. and Zhu, L. (2021). The heavy-tail phenomenon in SGD. In *International Conference on Machine Learning*.

[114] Hadji, A. and Szabó, B. (2021). Can we trust Bayesian uncertainty quantification from Gaussian process priors with squared exponential covariance kernel? *SIAM/ASA J. Uncertain. Quantificat.* **9** 185–230. MR4212884 https://doi.org/10.1137/19M1253010

[115] Hanin, B. (2023). Random neural networks in the infinite width limit as Gaussian processes. *Ann. Appl. Probab.* **33** 4798–4819. MR4674064 https://doi.org/10.1214/23-aap1933

[116] Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

[117] Hayou, S. (2022). On the infinite-depth limit of finite-width neural networks. arXiv preprint. Available at arXiv:2210.00688.

[118] He, K., Zhang, X., Ren, S. and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *International Conference on Computer Vision*.

[119] He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*.

[120] He, K., Zhang, X., Ren, S. and Sun, J. (2016). Identity mappings in deep residual networks. In *European Conference on Computer Vision*.

[121] Hein, M., Andriushchenko, M. and Bitterwolf, J. (2019). Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Computer Vision and Pattern Recognition*.

[122] Hensman, J., Fusi, N. and Lawrence, N. D. (2013). Gaussian processes for big data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence* 282–290.

[123] Hernández-Lobato, J. M. and Adams, R. (2022). Probabilistic backpropagation for scalable learning of Bayesian neural networks. In *International Conference on Machine Learning*.

[124] Hinton, G. E. and Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Conference on Computational Learning Theory*.

[125] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*. **9** 1735–1780.

[126] Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** 359–366.

[127] Hron, J., Bahri, Y., Novak, R., Pennington, J. and Sohl-Dickstein, J. (2020). Exact posterior distributions of wide Bayesian neural networks. In *ICML Workshop on Uncertainty and Robustness in Deep Learning*.

[128] Hron, J., Bahri, Y., Sohl-Dickstein, J. and Novak, R. (2020). Infinite attention: NNGP and NTK for deep attention networks. In *International Conference on Machine Learning*.

[129] Hron, J., Matthews, A. and Ghahramani, Z. (2018). Variational Bayesian dropout: Pitfalls and fixes. In *International Conference on Machine Learning*.

[130] Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J. E. and Weinberger, K. Q. (2017). Snapshot ensembles: Train 1, get M for free. In *International Conference on Learning Representations*.

[131] Huang, J. and Yau, H.-T. (2020). Dynamics of deep neural networks and neural tangent hierarchy. In *International Conference on Machine Learning*.

[132] Huang, Y., Chouzenoux, E., Elvira, V. and Pesquet, J.-C. (2023). Efficient Bayes inference in neural networks through adaptive importance sampling. *J. Franklin Inst*.

[133] Immer, A., Bauer, M., Fortuin, V., Rätsch, G. and Emtiyaz, K. M. (2021). Scalable marginal likelihood estimation for model selection in deep learning. In *International Conference on Machine Learning*.

[134] Immer, A., Hennigen, L. T., Fortuin, V. and Cotterell, R. (2022). Probing as quantifying inductive bias. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1839–1851.

[135] Immer, A., Korzepa, M. and Bauer, M. (2021). Improving predictions of Bayesian neural nets via local linearization. In *International Conference on Artificial Intelligence and Statistics*.

[136] Immer, A., van der Ouderaa, T., Rätsch, G., Fortuin, V. and van der Wilk, M. (2022). Invariance learning in deep neural networks with differentiable Laplace approximations. *Adv. Neural Inf. Process. Syst.*.

[137] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*.

[138] Izmailov, P., Nicholson, P., Lotfi, S. and Wilson, A. G. (2021). Dangers of Bayesian model averaging under covariate shift. *Adv. Neural Inf. Process. Syst.*.

[139] Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D. and Wilson, A. G. (2019). Averaging weights leads to wider optima and better generalization. In *Uncertainty in Artificial Intelligence*.

[140] Izmailov, P., Vikram, S., Hoffman, M. D. and Wilson, A. G. (2021). What are Bayesian neural network posteriors really like? In *International Conference on Machine Learning*.

[141] Jacot, A., Gabriel, F. and Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*.

[142] Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

[143] Jospin, L. V., Laga, H., Boussaid, F., Buntine, W. and Bennamoun, M. (2022). Hands-on Bayesian neural networks—a tutorial for deep learning users. *IEEE Comput. Intell. Mag.* **17** 29–48.

[144] Kääriäinen, M. and Langford, J. (2005). A comparison of tight generalization error bounds. In *International Conference on Machine Learning*.

[145] Kapoor, S., Maddox, W. J., Izmailov, P. and Wilson, A. G. (2022). On uncertainty, tempering, and data augmentation in Bayesian classification. *Adv. Neural Inf. Process. Syst.*.

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

[147] Khan, M., Nielsen, D., Tangkaratt, V., Lin, W., Gal, Y. and Srivastava, A. (2018). Fast and scalable Bayesian deep learning by weight-perturbation in Adam. In *International Conference on Machine Learning*.

[148] Khan, M. E., Liu, Z., Tangkaratt, V. and Gal, Y. (2017). Vprop: Variational inference using rmsprop. arXiv preprint. Available at arXiv:1712.01038.

[149] Khan, M. E. and Rue, H. (2023). The Bayesian learning rule. *J. Mach. Learn. Res.* **24** 1–46. MR4664718

[150] Khan, M. E. and Swaroop, S. (2021). Knowledge-adaptation priors. *Adv. Neural Inf. Process. Syst.*.

[151] Kingma, D. P., Salimans, T. and Welling, M. (2015). Variational dropout and the local reparameterization trick. In *Advances in Neural Information Processing Systems*.

[152] Kingma, D. P. and Welling, M. (2014). Auto-encoding variational Bayes. In *International Conference on Learning Representations*.

[153] Koh, P. W. and Liang, P. (2017). Understanding black-box predictions via influence functions. In *International Conference on Machine Learning*.

[154] Kohler, M. and Langer, S. (2021). On the rate of convergence of fully connected deep neural network regression estimates. *Ann. Statist.* **49** 2231–2249. MR4319248 https://doi.org/10.1214/20-aos2034

[155] Korattikara Balan, A., Rathod, V., Murphy, K. P. and Welling, M. (2015). Bayesian dark knowledge. *Adv. Neural Inf. Process. Syst.*.

[156] Kristiadi, A., Hein, M. and Hennig, P. (2020). Being Bayesian, even just a bit, fixes overconfidence in ReLU networks. In *International Conference on Machine Learning* 5436–5446.

[157] Kristiadi, A., Immer, A., Eschenhagen, R. and Fortuin, V. (2023). Promises and pitfalls of the linearized Laplace

in Bayesian optimization. In *Fifth Symposium on Advances in Approximate Bayesian Inference*.

[158] KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.

[159] KUMAR, V., SINGH, V., SRIJITH, P. and DAMIANOU, A. (2018). Deep Gaussian processes with convolutional kernels. arXiv preprint. Available at arXiv:1806.01655.

[160] LAKSHMINARAYANAN, B., PRITZEL, A. and BLUNDELL, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.*.

[161] LANGFORD, J. (2005). Tutorial on practical prediction theory for classification. *J. Mach. Learn. Res.* **6** 273–306. MR2249822

[162] LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. and JACKEL, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1** 541–551.

[163] LEE, J., SCHOENHOLZ, S., PENNINGTON, J., ADLAM, B., XIAO, L., NOVAK, R. and SOHL-DICKSTEIN, J. (2020). Finite versus infinite neural networks: An empirical study. In *International Conference on Neural Information Processing Systems*.

[164] LEE, J., SOHL-DICKSTEIN, J., PENNINGTON, J., NOVAK, R., SCHOENHOLZ, S. and BAHRI, Y. (2018). Deep neural networks as Gaussian processes. In *International Conference on Learning Representations*.

[165] LEE, J., XIAO, L., SCHOENHOLZ, S. S., BAHRI, Y., NOVAK, R., SOHL-DICKSTEIN, J. and PENNINGTON, J. (2020). Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems* **32** 8572–8583.

[166] LEE, K., LEE, H., LEE, K. and SHIN, J. (2018). Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *International Conference on Learning Representations*.

[167] LI, C., CHEN, C., CARLSON, D. and CARIN, L. (2016). Preconditioned stochastic gradient Langevin dynamics for deep neural networks. In *Conference on Artificial Intelligence*.

[168] LI, Y. and GAL, Y. (2017). Dropout inference in Bayesian neural networks with alpha-divergences. In *International Conference on Machine Learning* 2052–2061. PMLR.

[169] LI, Y., YU, Q., TAN, M., MEI, J., TANG, P., SHEN, W., YUILLE, A. and XIE, C. (2021). Shape-texture debiased neural network training. In *International Conference on Learning Representations*.

[170] LIM, B., ARIK, S. Ö., LOEFF, N. and PFISTER, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **37** 1748–1764.

[171] LIPTON, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **16** 31–57.

[172] LIU, Q. and WANG, D. (2016). Stein variational gradient descent: A general purpose Bayesian inference algorithm. *Adv. Neural Inf. Process. Syst.* **29**.

[173] LIU, X., LI, Y., WU, C. and HSIEH, C.-J. (2019). Adv-BNN: Improved adversarial defense through robust Bayesian neural network. In *International Conference on Learning Representations*.

[174] LLOYD, J., DUVENAUD, D., GROSSE, R., TENENBAUM, J. and GHAHRAMANI, Z. (2014). Automatic construction and natural-language description of nonparametric regression models. In *International Conference on Artificial Intelligence and Statistics*.

[175] LOOG, M., VIERING, T., MEY, A., KRIJTHE, J. H. and TAX, D. M. (2020). A brief prehistory of double descent. *Proc. Natl. Acad. Sci. USA* **117** 10625–10626.

[176] LOTFI, S., IZMAILOV, P., BENTON, G., GOLDBLUM, M. and WILSON, A. G. (2022). Bayesian model selection, the marginal likelihood, and generalization. arXiv preprint. Available at arXiv:2202.11678.

[177] LOUIZOS, C. and WELLING, M. (2016). Structured and efficient variational deep learning with matrix Gaussian posteriors. In *International Conference on Machine Learning*.

[178] LOUIZOS, C. and WELLING, M. (2017). Multiplicative normalizing flows for variational Bayesian neural networks. In *International Conference on Machine Learning*.

[179] LU, L., SHIN, Y., SU, Y. and KARNIADAKIS, G. E. (2020). Dying ReLU and initialization: Theory and numerical examples. *Commun. Comput. Phys.* **28** 1671–1706. MR4188518 https://doi.org/10.4208/cicp.oa-2020-0165

[180] LYLE, C., SCHUT, L., RU, R., GAL, Y. and VAN DER WILK, M. (2020). A Bayesian perspective on training speed and model selection. *Adv. Neural Inf. Process. Syst.*.

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

[182] MACKAY, D. J. (1994). A practical Bayesian framework for backpropagation networks. *Neural Comput.* **4** 448–472.

[183] MACKAY, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge Univ. Press, New York. MR2012999

[184] MADDOX, W. J., IZMAILOV, P., GARIPOV, T., VETROV, D. P. and WILSON, A. G. (2019). A simple baseline for Bayesian uncertainty in deep learning. *Adv. Neural Inf. Process. Syst.*.

[185] MAHONEY, M. and MARTIN, C. (2019). Traditional and heavy-tailed self-regularization in neural network models. In *International Conference on Machine Learning*.

[186] MALACH, E., YEHUDAI, G., SHALEV-SCHWARTZ, S. and SHAMIR, O. (2020). Proving the lottery ticket hypothesis: Pruning is all you need. In *International Conference on Machine Learning*.

[187] MARTENS, J. and GROSSE, R. (2015). Optimizing neural networks with Kronecker-factored approximate curvature. In *International Conference on Machine Learning*.

[188] MATSUBARA, T., OATES, C. J. and BRIOL, F.-X. (2021). The ridgelet prior: A covariance function approach to prior specification for Bayesian neural networks. *J. Mach. Learn. Res.* **22** Paper No. 157, 57. MR4318513

[189] MATTHEWS, A. G. D. G., ROWLAND, M., HRON, J., TURNER, R. E. and GHAHRAMANI, Z. (2018). Gaussian process behaviour in wide deep neural networks. In *International Conference on Learning Representations*.

[190] MCALLESTER, D. A. (1999). Some PAC-Bayesian theorems. *Mach. Learn.* **37** 355–363.

[191] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. MR4400901 https://doi.org/10.1002/cpa.22008

[192] MELLOR, J., TURNER, J., STORKEY, A. and CROWLEY, E. J. (2021). Neural architecture search without training. In *International Conference on Machine Learning*.

[193] MIANJY, P., ARORA, R. and VIDAL, R. (2018). On the implicit bias of dropout. In *International Conference on Machine Learning*.

[194] MIKKOLA, P., MARTIN, O. A., CHANDRAMOULI, S., HARTMANN, M., PLA, O. A., THOMAS, O., PESONEN, H., CORANDER, J., VEHTARI, A. et al. (2023). Prior knowledge elicitation: The past, present, and future. *Bayesian Anal.*.

[195] MINDERER, M., DJOLONGA, J., ROMIJNDERS, R., HUBIS, F., ZHAI, X., HOULSBY, N., TRAN, D. and LUCIC, M. (2021). Revisiting the calibration of modern neural networks. *Adv. Neural Inf. Process. Syst.*.

[196] MISHKIN, A., KUNSTNER, F., NIELSEN, D., SCHMIDT, M. and KHAN, M. E. (2018). Slang: Fast structured covariance approximations for Bayesian deep learning with natural gradient. *Adv. Neural Inf. Process. Syst.*.

[197] MITROS, J. and MAC NAMEE, B. (2019). On the validity of Bayesian neural networks for uncertainty estimation. arXiv preprint. Available at arXiv:1912.01530.

[198] MODAS, A., RADE, R., ORTIZ-JIMÉNEZ, G., MOOSAVI-DEZFOOLI, S.-M. and FROSSARD, P. (2021). PRIME: a few primitives can boost robustness to common corruptions. arXiv preprint. Available at arXiv:2112.13547.

[199] MOHRI, M., ROSTAMIZADEH, A. and TALWALKAR, A. (2018). *Foundations of Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3931734

[200] MOINS, T., ARBEL, J., DUTFOY, A. and GIRARD, S. (2025). On the use of a local $\hat{R}$ to improve MCMC convergence diagnostic. *Bayesian Anal.* **20** 1433–1458. MR4832254 https://doi.org/10.1214/23-ba1399

[201] MOLCHANOV, D., ASHUKHA, A. and VETROV, D. (2017). Variational dropout sparsifies deep neural networks. In *International Conference on Machine Learning*.

[202] MÖLLENHOFF, T. and KHAN, M. E. (2022). Sam as an optimal relaxation of Bayes. In *International Conference on Learning Representations*.

[203] MONTAVON, G., SAMEK, W. and MÜLLER, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **73** 1–15. MR3737870 https://doi.org/10.1016/j.dsp.2017.10.011

[204] MOOSAVI-DEZFOOLI, S.-M., FAWZI, A., FAWZI, O. and FROSSARD, P. (2017). Universal adversarial perturbations. In *Conference on Computer Vision and Pattern Recognition*.

[205] MOOSAVI-DEZFOOLI, S.-M., FAWZI, A. and FROSSARD, P. (2016). Deepfool: A simple and accurate method to fool deep neural networks. In *Conference on Computer Vision and Pattern Recognition*.

[206] MORRISON, K., GILBY, B., LIPCHAK, C., MATTIOLI, A. and KOVASHKA, A. (2021). Exploring corruption robustness: Inductive biases in vision transformers and MLP-mixers. arXiv preprint. Available at arXiv:2106.13122.

[207] MURPHY, K. P. (2012). *Machine Learning*: *A Probabilistic Perspective*. MIT press, Cambridge.

[208] NABARRO, S., GANEV, S., GARRIGA-ALONSO, A., FORTUIN, V., VAN DER WILK, M. and AITCHISON, L. (2021). Data augmentation in Bayesian neural networks and the cold posterior effect. arXiv preprint. Available at arXiv:2106.05586.

[209] NAEINI, M. P., COOPER, G. and HAUSKRECHT, M. (2015). Obtaining well calibrated probabilities using Bayesian binning. In *Conference on Artificial Intelligence*.

[210] NAKKIRAN, P., KAPLUN, G., BANSAL, Y., YANG, T., BARAK, B. and SUTSKEVER, I. (2021). Deep double descent: Where bigger models and more data hurt. *J. Stat. Mech. Theory Exp.* Paper No. 124003, 32. MR4412831 https://doi.org/10.1088/1742-5468/ac3a74

[211] NALISNICK, E., GORDON, J. and HERNÁNDEZ-LOBATO, J. M. (2021). Predictive complexity priors. In *International Conference on Artificial Intelligence and Statistics*.

[212] NALISNICK, E., HERNÁNDEZ-LOBATO, J. M. and SMYTH, P. (2019). Dropout as a structured shrinkage prior. In *International Conference on Machine Learning*.

[213] NALISNICK, E. T. (2018). On priors for Bayesian neural networks Ph.D. thesis, Univ. of California Press, Irvine.

[214] NAVEH, G., BEN DAVID, O., SOMPOLINSKY, H. and RINGEL, Z. (2021). Predicting the outputs of finite deep neural networks trained with noisy gradients. *Phys. Rev. E* **104** Paper No. 064301, 19. MR4366178 https://doi.org/10.1103/physreve.104.064301

[215] NEAL, R. M. (1992). Bayesian training of backpropagation networks by the hybrid Monte Carlo method Technical report, Citeseer.

[216] NEAL, R. M. (1993). Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems*.

[217] NEAL, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer, Berlin.

[218] NEELAKANTAN, A., VILNIS, L., LE, Q. V., SUTSKEVER, I., KAISER, L., KURACH, K. and MARTENS, J. (2016). Adding gradient noise improves learning for very deep networks. In *International Conference on Learning Representations Workshop*.

[219] NEISWANGER, W., WANG, C. and XING, E. P. (2014). Asymptotically exact, embarrassingly parallel mcmc. In *Conference on Uncertainty in Artificial Intelligence*.

[220] NEYSHABUR, B., BHOJANAPALLI, S., MCALLESTER, D. and SREBRO, N. (2018). A PAC-Bayesian approach to spectrally-normalized margin bounds for neural networks. In *International Conference on Learning Representations*.

[221] NICKL, P., XU, L., TAILOR, D., MÖLLENHOFF, T. and KHAN, M. E. E. (2024). The memory-perturbation equation: Understanding model's sensitivity to data. *Adv. Neural Inf. Process. Syst.* **36**.

[222] NIXON, J., DUSENBERRY, M. W., ZHANG, L., JERFEL, G. and TRAN, D. (2019). Measuring calibration in deep learning. In *CVPR Workshop*.

[223] NOCI, L., BACHMANN, G., ROTH, K., NOWOZIN, S. and HOFMANN, T. (2021). Precise characterization of the prior predictive distribution of deep ReLU networks. In *International Conference on Learning Representations*.

[224] NOVAK, R., XIAO, L., BAHRI, Y., LEE, J., YANG, G., HRON, J., ABOLAFIA, D. A., PENNINGTON, J. and SOHL-DICKSTEIN, J. (2020). Bayesian deep convolutional networks with many channels are Gaussian processes. In *ICML Workshop on Uncertainty and Robustness in Deep Learning*.

[225] OSAWA, K., SWAROOP, S., KHAN, M. E. E., JAIN, A., ESCHENHAGEN, R., TURNER, R. E. and YOKOTA, R. (2019). Practical deep learning with Bayesian principles. *Adv. Neural Inf. Process. Syst.*.

[226] OSBAND, I., WEN, Z., ASGHARI, S. M., DWARACHERLA, V., LU, X., IBRAHIMI, M., LAWSON, D., HAO, B., O'DONOGHUE, B. et al. (2022). The neural testbed: Evaluating joint predictions. *Adv. Neural Inf. Process. Syst.* **35** 12554–12565.

[227] OVADIA, Y., FERTIG, E., REN, J., NADO, Z., SCULLEY, D., NOWOZIN, S., DILLON, J. V., LAKSHMINARAYANAN, B. and SNOEK, J. (2019). Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *Adv. Neural Inf. Process. Syst.*.

[228] PAPAMARKOU, T., SKOULARIDOU, M., PALLA, K., AITCHISON, L., ARBEL, J., DUNSON, D., FILIPPONE, M., FORTUIN, V., HENNIG, P. et al. (2024). Position paper: Bayesian deep learning in the age of large-scale AI. arXiv preprint. Available at arXiv:2402.00809.

[229] PEARCE, T., LEIBFRIED, F. and BRINTRUP, A. (2020). Uncertainty in neural networks: Approximately Bayesian ensembling. In *International Conference on Artificial Intelligence and Statistics*.

[230] PITAS, K. and ARBEL, J. (2022). Cold Posteriors through PAC-Bayes. arXiv preprint. Available at arXiv:2206.11173.

[231] PLATT, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*.

[232] POLSON, N. and ROCKOVA, V. (2018). Posterior concentration for sparse deep learning. In *Advances in Neural Information Processing Systems*.

[233] POOLE, B., LAHIRI, S., RAGHU, M., SOHL-DICKSTEIN, J. and GANGULI, S. (2016). Exponential expressivity in deep neural networks through transient chaos. In *International Conference on Neural Information Processing Systems*.

[234] PRICE, R. (1958). A useful theorem for nonlinear devices having Gaussian inputs. *IRE Trans*. *Inf. Theory* **4** 69–72. MR0122618 https://doi.org/10.1109/tit.1958.1057444

[235] QUIÑONERO-CANDELA, J. and RASMUSSEN, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *J*. *Mach*. *Learn*. *Res*. **6** 1939–1959. MR2249877

[236] RASMUSSEN, C. E. and GHAHRAMANI, Z. (2001). Occam's razor. *Adv*. *Neural Inf*. *Process*. *Syst*..

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

[238] RAWAT, A., WISTUBA, M. and NICOLAE, M.-I. (2017). Adversarial phenomenon in the eyes of Bayesian deep learning. arXiv preprint. Available at arXiv:1711.08244.

[239] RITTER, H., BOTEV, A. and BARBER, D. (2018). A scalable Laplace approximation for neural networks. In *International Conference on Learning Representations*.

[240] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann*. *Math*. *Stat*. **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

[241] ROBERT, C. P. (2016). The expected demise of the Bayes factor [comment on MR3506022]. *J*. *Math*. *Psych*. **72** 33–37. MR3506023 https://doi.org/10.1016/j.jmp.2015.08.002

[242] 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

[243] ROSENBLATT, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol*. *Rev*. **65** 386.

[244] ROTHFUSS, J., FORTUIN, V., JOSIFOSKI, M. and KRAUSE, A. (2021). PACOH: Bayes-optimal meta-learning with PAC-guarantees. In *International Conference on Machine Learning* 9116–9126. PMLR.

[245] ROTHFUSS, J., JOSIFOSKI, M., FORTUIN, V. and KRAUSE, A. (2023). Scalable PAC-Bayesian meta-learning via the PAC-optimal hyper-posterior: From theory to practice. *J*. *Mach*. *Learn*. *Res*. **24** Paper No. [386], 62. MR4720842

[246] RU, R., LYLE, C., SCHUT, L., FIL, M., VAN DER WILK, M. and GAL, Y. (2021). Speedy performance estimation for neural architecture search. *Adv*. *Neural Inf*. *Process*. *Syst*..

[247] RUDNER, T. G., SEJDINOVIC, D. and GAL, Y. (2020). Inter-domain deep Gaussian processes. In *International Conference on Machine Learning* 8286–8294. PMLR.

[248] RUMELHART, D. E., HINTON, G. E. and WILLIAMS, R. J. (1986). Learning representations by back-propagating errors. *Nature* **323** 533–536.

[249] SALIMBENI, H. and DEISENROTH, M. P. (2017). Doubly stochastic variational inference for deep Gaussian processes. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* 4591–4602.

[250] SALIMBENI, H., DUTORDOIR, V., HENSMAN, J. and DEISENROTH, M. (2019). Deep Gaussian processes with importance-weighted variational inference. In *International Conference on Machine Learning* 5589–5598. PMLR.

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

[252] SARKER, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput*. *Sci*. **2** 1–20.

[253] SAVAGE, L. J. (1972). *The Foundations of Statistics*, 2nd ed. Dover, New York. MR0348870

[254] SCHAFFER, J. (2015). What not to multiply without necessity. *Australas*. *J*. *Philos* **93** 644–664.

[255] SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann*. *Statist*. **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875

[256] SCHOENHOLZ, S. S., GILMER, J., GANGULI, S. and SOHL-DICKSTEIN, J. (2017). Deep information propagation. In *International Conference on Learning Representations*.

[257] SELVARAJU, R. R., COGSWELL, M., DAS, A., VEDANTAM, R., PARIKH, D. and BATRA, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *International Conference on Computer Vision*.

[258] SHAFER, G. and VOVK, V. (2008). A tutorial on conformal prediction. *J*. *Mach*. *Learn*. *Res*. **9** 371–421. MR2417240

[259] SHARMA, M., RAINFORTH, T., TEH, Y. W. and FORTUIN, V. (2023). Incorporating unlabelled data into Bayesian neural networks. arXiv preprint. Available at arXiv:2304.01762.

[260] SHEN, Y., DAHEIM, N., CONG, B., NICKL, P., MARCONI, G. M., BAZAN, C., YOKOTA, R., GUREVYCH, I., CREMERS, D. et al. (2024). Variational learning is effective for large deep networks. In *International Conference on Machine Learning*.

[261] SHWARTZ-ZIV, R., GOLDBLUM, M., SOURI, H., KAPOOR, S., ZHU, C., LECUN, Y. and WILSON, A. G. (2022). Pre-train your loss: Easy Bayesian transfer learning with informative priors. *Adv*. *Neural Inf*. *Process*. *Syst*. **35** 27706–27715.

[262] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V. et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature* **529** 484–489.

[263] SNELSON, E. and GHAHRAMANI, Z. (2007). Local and global sparse Gaussian process approximations. In *Artificial Intelligence and Statistics* 524–531. PMLR.

[264] SOKOLIĆ, J., GIRYES, R., SAPIRO, G. and RODRIGUES, M. R. D. (2017). Robust large margin deep neural networks. *IEEE Trans*. *Signal Process*. **65** 4265–4280. MR3684063 https://doi.org/10.1109/TSP.2017.2708039

[265] SOLOMONOFF, R. J. (2009). Algorithmic probability: Theory and applications. In *Information Theory and Statistical Learning*. 1–23.

[266] SOUDRY, D., HOFFER, E., NACSON, M. S., GUNASEKAR, S. and SREBRO, N. (2018). The implicit bias of gradient descent on separable data. *J*. *Mach*. *Learn*. *Res*. **19** 2822–2878. MR3899772

[267] SPRINGER, M. D. and THOMPSON, W. E. (1970). The distribution of products of beta, gamma and Gaussian random variables. *SIAM J*. *Appl*. *Math*. **18** 721–737. MR0264733 https://doi.org/10.1137/0118065

[268] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. and SALAKHUTDINOV, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J*. *Mach*. *Learn*. *Res*. **15** 1929–1958. MR3231592

[269] STERKENBURG, T. F. (2016). Solomonoff prediction and Occam's razor. *Philos*. *Sci*. **83** 459–479. MR3560504 https://doi.org/10.1086/687257

[270] STERKENBURG, T. F. and GRÜNWALD, P. D. (2021). The no-free-lunch theorems of supervised learning. *Synthese* **199**

9979–10015. MR4351422 https://doi.org/10.1007/s11229-021-03233-1

[271] SUN, S., CAO, Z., ZHU, H. and ZHAO, J. (2019). A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern.* **50** 3668–3681.

[272] SUN, S., CHEN, C. and CARIN, L. (2017). Learning structured weight uncertainty in Bayesian neural networks. In *International Conference on Artificial Intelligence and Statistics*.

[273] SUN, S., ZHANG, G., SHI, J. and GROSSE, R. (2019). Functional variational Bayesian neural networks. In *International Conference on Learning Representations*.

[274] SUN, S., ZHANG, G., WANG, C., ZENG, W., LI, J. and GROSSE, R. (2018). Differentiable compositional kernel learning for Gaussian processes. In *International Conference on Machine Learning*.

[275] SUNDARARAJAN, M., TALY, A. and YAN, Q. (2017). Axiomatic attribution for deep networks. In *International Conference on Machine Learning*.

[276] SUZUKI, T. (2018). Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality. In *International Conference on Learning Representations*.

[277] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

[278] TELGARSKY, M. (2016). Benefits of depth in neural networks. In *Conference on Learning Theory*.

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

[280] TITSIAS, M. (2009). Variational learning of inducing variables in sparse Gaussian processes. In *Artificial Intelligence and Statistics* 567–574. PMLR.

[281] TRAN, B.-H., ROSSI, S., MILIOS, D. and FILIPPONE, M. (2022). All you need is a good functional prior for Bayesian deep learning. *J. Mach. Learn. Res.* **23** Paper No. [74], 56. MR4576659

[282] TRAN, D., DUSENBERRY, M. W., VAN DER WILK, M. and HAFNER, D. (2018). Bayesian layers: a module for neural network uncertainty. arXiv preprint. Available at arXiv:1812.03973.

[283] ULYANOV, D., VEDALDI, A. and LEMPITSKY, V. (2018). Deep image prior. In *Computer Vision and Pattern Recognition*.

[284] VAICENAVICIUS, J., WIDMANN, D., ANDERSSON, C., LINDSTEN, F., ROLL, J. and SCHÖN, T. (2019). Evaluating model calibration in classification. In *International Conference on Artificial Intelligence and Statistics*.

[285] VAPNIK, V. N. (1999). An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10** 988–999.

[286] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. and POLOSUKHIN, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*.

[287] VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. and BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718. MR4298989 https://doi.org/10.1214/20-ba1221

[288] VLADIMIROVA, M., ARBEL, J. and GIRARD, S. (2021). Bayesian neural network unit priors and generalized Weibull-tail property. In *Asian Conference on Machine Learning*.

[289] VLADIMIROVA, M., ARBEL, J. and GIRARD, S. (2021). Dependence between Bayesian neural network units. In *NeurIPS Workshop on Bayesian Deep Learning*.

[290] VLADIMIROVA, M., GIRARD, S., NGUYEN, H. and ARBEL, J. (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-exponential properties to heavier tailed distributions. *Stat* **9** e318, 8. MR4193421 https://doi.org/10.1002/sta4.318

[291] VLADIMIROVA, M., VERBEEK, J., MESEJO, P. and ARBEL, J. (2019). Understanding priors in Bayesian neural networks at the unit level. In *International Conference on Machine Learning*.

[292] WAGER, S., WANG, S. and LIANG, P. S. (2013). Dropout training as adaptive regularization. *Adv. Neural Inf. Process. Syst.*.

[293] WANG, Y.-X., FIENBERG, S. and SMOLA, A. (2015). Privacy for free: Posterior sampling and stochastic gradient Monte Carlo. In *International Conference on Machine Learning*.

[294] WEISS, K., KHOSHGOFTAAR, T. M. and WANG, D. (2016). A survey of transfer learning. *J. Big Data* **3** 1–40.

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

[296] WEN, Y., VICOL, P., BA, J., TRAN, D. and GROSSE, R. (2018). Flipout: Efficient pseudo-independent weight perturbations on mini-batches. arXiv preprint. Available at arXiv:1803.04386.

[297] WENZEL, F., ROTH, K., VEELING, B., SWIATKOWSKI, J., TRAN, L., MANDT, S., SNOEK, J., SALIMANS, T., JENATTON, R. et al. (2020). How good is the Bayes posterior in deep neural networks really? In *International Conference on Machine Learning*.

[298] WILSON, A. G. (2020). The case for Bayesian deep learning. arXiv preprint. Available at arXiv:2001.10995.

[299] WILSON, A. G., HU, Z., SALAKHUTDINOV, R. and XING, E. P. (2016). Deep kernel learning. In *International Conference on Artificial Intelligence and Statistics*.

[300] WILSON, A. G. and IZMAILOV, P. (2020). Bayesian deep learning and a probabilistic perspective of generalization. In *International Conference on Neural Information Processing Systems*.

[301] WOLINSKI, P. and ARBEL, J. (2023). Gaussian pre-activations in neural networks: Myth or reality? arXiv preprint. Available at arXiv:2205.12379.

[302] WOLINSKI, P., CHARPIAT, G. and OLLIVIER, Y. (2020). Interpreting a penalty as the influence of a Bayesian prior. arXiv preprint. Available at arXiv:2002.00178.

[303] WOLPERT, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Comput.* **8** 1341–1390.

[304] YAIDA, S. (2020). Non-Gaussian processes and neural networks at finite widths. In *Mathematical and Scientific Machine Learning*.

[305] YANG, G. (2019). Scaling limits of wide neural networks with weight sharing: Gaussian process behavior, gradient independence, and neural tangent kernel derivation. arXiv preprint. Available at arXiv:1902.04760.

[306] YANG, G. (2019). Tensor programs i: Wide feedforward or recurrent neural networks of any architecture are Gaussian processes. *Adv. Neural Inf. Process. Syst.*.

[307] ZADROZNY, B. and ELKAN, C. (2001). Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In *International Conference on Machine Learning*.

[308] ZADROZNY, B. and ELKAN, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *International Conference on Knowledge Discovery and Data Mining*.

[309] ZAVATONE-VETH, J. A., CANATAR, A., RUBEN, B. S. and PEHLEVAN, C. (2022). Asymptotics of representation learning in finite Bayesian neural networks. *J. Stat. Mech. Theory Exp.* Paper No. 114008, 18. MR4535579 https://doi.org/10.1088/1742-5468/ac98a6

[310] ZAVATONE-VETH, J. A. and PEHLEVAN, C. (2021). Exact priors of finite neural networks. In *International Conference on Learning Representations*.

[311] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2017). Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations*.

[312] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* **64** 107–115.

[313] ZHANG, G., SUN, S., DUVENAUD, D. and GROSSE, R. (2018). Noisy natural gradient as variational inference. In *International Conference on Machine Learning*.

[314] ZHANG, R., LI, C., ZHANG, J., CHEN, C. and WILSON, A. G. (2019). Cyclical stochastic gradient MCMC for Bayesian deep learning. In *International Conference on Learning Representations*.

Statistical Science
2026, Vol. 41, No. 2, 354–385
https://doi.org/10.1214/24-STS970

An Overview of Asymptotic Normality in Stochastic Blockmodels: Cluster Analysis and Inference

Joshua Agterberg and Joshua Cape

Abstract. This paper provides a selective review of the statistical network analysis literature focused on clustering and inference problems for stochastic blockmodels and their variants. We survey asymptotic normality results for stochastic blockmodels as a means of thematically linking classical statistical concepts to contemporary research in network data analysis. Multiple different forms of asymptotically Gaussian behavior arise in stochastic blockmodels and are useful for different purposes, pertaining to estimation and testing, the characterization of cluster structure in community detection and understanding latent space geometry. This paper concludes with a discussion of open problems and ongoing research activities addressing asymptotic normality and its implications for statistical network modeling.

Key words and phrases: Random graphs, statistical network analysis, community detection, clustering, limit theorems.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177. MR3827065

[2] ABBE, E., FAN, J. and WANG, K. (2022). An ℓ_p theory of PCA and spectral clustering. *Ann. Statist.* **50** 2359–2385. MR4474494 https://doi.org/10.1214/22-aos2196

[3] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[4] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inf. Theory* **68** 4618–4650. MR4449064 https://doi.org/10.1109/tit.2022.3159085

[5] AGTERBERG, J., TANG, M. and PRIEBE, C. (2020). Nonparametric two-sample hypothesis testing for random graphs with negative and repeated eigenvalues. Available at arXiv:2012.09828.

[6] AIROLDI, E. M., BLEI, D., FIENBERG, S. and XING, E. (2008). Mixed membership stochastic blockmodels. In *Advances in Neural Information Processing Systems* **21**. Curran Associates, Red Hook.

[7] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859 https://doi.org/10.1214/13-AOS1138

[8] ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142. MR4318498

[9] ARROYO, J. and LEVINA, E. (2022). Overlapping community detection in networks via sparse spectral decomposition. *Sankhya A* **84** 1–35. MR4402745 https://doi.org/10.1007/s13171-021-00245-4

[10] ATHREYA, A., CAPE, J. and TANG, M. (2022). Eigenvalues of stochastic blockmodel graphs and random graphs with low-rank edge probability matrices. *Sankhya A* **84** 36–63. MR4402746 https://doi.org/10.1007/s13171-021-00268-x

[11] ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. et al. (2017). Statistical inference on random dot product graphs: A survey. *J. Mach. Learn. Res.* **18** Paper No. 226. MR3827114

[12] ATHREYA, A., LUBBERTS, Z., PRIEBE, C. E., PARK, Y., TANG, M., LYZINSKI, V., KANE, M. and LEWIS, B. W. (2023). Numerical tolerance for spectral decompositions of random matrices and applications to network inference. *J. Comput. Graph. Statist.* **32** 145–156. MR4552943 https://doi.org/10.1080/10618600.2022.2082972

[13] ATHREYA, A., PRIEBE, C. E., TANG, M., LYZINSKI, V., MARCHETTE, D. J. and SUSSMAN, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. MR3494576 https://doi.org/10.1007/s13171-015-0071-x

Joshua Agterberg is an Assistant Professor, Department of Statistics, University of Illinois Urbana-Champaign, Champaign, Illinois 61801, USA (e-mail: jagt@illinois.edu). Joshua Cape is an Assistant Professor, Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA (e-mail: jrcape@wisc.edu).

[14] BAI, Z. D. and PAN, G. M. (2012). Limiting behavior of eigenvectors of large Wigner matrices. *J. Stat. Phys.* **146** 519–549. MR2880031 https://doi.org/10.1007/s10955-011-0407-4

[15] BANDEIRA, A. S. and VAN HANDEL, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **44** 2479–2506. MR3531673 https://doi.org/10.1214/15-AOP1025

[16] BAROT, A., BHAMIDI, S. and DHARA, S. (2021). Community detection using low-dimensional network embedding algorithms. Available at arXiv:2111.05267 [cs, math, stat].

[17] BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[18] BHATTACHARYYA, S. and BICKEL, P. J. (2015). Subsampling bootstrap of count features of networks. *Ann. Statist.* **43** 2384–2411. MR3405598 https://doi.org/10.1214/15-AOS1338

[19] BHATTACHARYYA, S. and CHATTERJEE, S. (2018). Spectral clustering for multiple sparse networks: I. Available at arXiv:1805.10594 [cs, Math, Stat].

[20] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124

[21] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

[22] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *Ann. Statist.* **49** 944–967. MR4255114 https://doi.org/10.1214/20-aos1986

[23] CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

[24] CAPE, J. (2024). On varimax asymptotics in network models and spectral methods for dimensionality reduction. *Biometrika* **111** 609–623. MR4745584 https://doi.org/10.1093/biomet/asad061

[25] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070

[26] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752

[27] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). On spectral embedding performance and elucidating network structure in stochastic blockmodel graphs. *Netw. Sci.* **7** 269–291.

[28] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Proceedings of the* 25*th Annual Conference on Learning Theory* 35.1–35.23. JMLR Workshop and Conference Proceedings. ISSN: 1938-7228.

[29] CHEN, L., ZHOU, J. and LIN, L. (2023). Hypothesis testing for populations of networks. *Comm. Statist. Theory Methods* **52** 3661–3684. MR4571714 https://doi.org/10.1080/03610926.2021.1977961

[30] CHEN, S., LIU, S. and MA, Z. (2022). Global and individualized community detection in inhomogeneous multilayer networks. *Ann. Statist.* **50** 2664–2693. MR4500621 https://doi.org/10.1214/22-aos2202

[31] CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806. Available at arXiv:2012.08496.

[32] CHUNG, F. R. K. (1997). *Spectral Graph Theory*. *CBMS Regional Conference Series in Mathematics* **92**. Amer. Math. Soc., Providence, RI. MR1421568

[33] DASGUPTA, A., HOPCROFT, J. E. and MCSHERRY, F. (2004). Spectral analysis of random graphs with skewed degree distributions. In 45*th Annual IEEE Symposium on Foundations of Computer Science* 602–610. ISSN: 0272-5428.

[34] DRAVES, B. and SUSSMAN, D. L. (2021). Bias-variance tradeoffs in joint spectral embeddings. Available at arXiv:2005.02511 [math, stat].

[35] DU, X. and TANG, M. (2023). Hypothesis testing for equality of latent positions in random graphs. *Bernoulli* **29** 3221–3254. MR4632137 https://doi.org/10.3150/22-bej1581

[36] ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 38. Proceedings of Machine Learning Research PMLR. MR3857310

[37] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505

[38] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990

[39] FAN, X., PENSKY, M., YU, F. and ZHANG, T. (2022). ALMA: Alternating minimization algorithm for clustering mixture multilayer network. *J. Mach. Learn. Res.* **23** Paper No. [330]. MR4577769

[40] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414 https://doi.org/10.1016/j.physrep.2009.11.002

[41] FORTUNATO, S. and HRIC, D. (2016). Community detection in networks: A user guide. *Phys. Rep.* **659** 1–44. MR3566093 https://doi.org/10.1016/j.physrep.2016.09.002

[42] FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828 https://doi.org/10.1007/BF02579329

[43] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. Available at arXiv:1710.00862 [cs, Math, Stat].

[44] GAO, C. and MA, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statist. Sci.* **36** 16–33. MR4194201 https://doi.org/10.1214/19-STS736

[45] GHOSHDASTIDAR, D., GUTZEIT, M., CARPENTIER, A. and LUXBURG, U. V. (2017). Two-sample tests for large random graphs using network statistics. In *Proceedings of the* 2017 *Conference on Learning Theory* 954–977. PMLR. ISSN: 2640-3498.

[46] GHOSHDASTIDAR, D., GUTZEIT, M., CARPENTIER, A. and VON LUXBURG, U. (2020). Two-sample hypothesis testing for inhomogeneous random graphs. *Ann. Statist.* **48** 2208–2229. MR4134792 https://doi.org/10.1214/19-AOS1884

[47] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2009). A survey of statistical network models. *Found. Trends Mach. Learn.*.

[48] GREEN, A. and SHALIZI, C. R. (2022). Bootstrapping exchangeable random graphs. *Electron. J. Stat.* **16** 1058–1095. MR4377133 https://doi.org/10.1214/21-ejs1896

[49] GROVER, A. and LESKOVEC, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. *KDD'*16 855–864. Association for Computing Machinery, New York, NY, USA.

[50] HAN, X., YANG, Q. and FAN, Y. (2023). Universal rank inference via residual subsampling with application to large networks. *Ann. Statist.* **51** 1109–1133. MR4630942 https://doi.org/10.1214/23-aos2282

[51] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

[52] 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

[53] HUANG, S., WENG, H. and FENG, Y. (2023). Spectral clustering via adaptive layer aggregation for multi-layer networks. *J. Comput. Graph. Statist.* **32** 1170–1184. MR4641490 https://doi.org/10.1080/10618600.2022.2134874

[54] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

[55] JIN, J., KE, Z. T. and LUO, S. (2024). Mixed membership estimation for social networks. *J. Econometrics* **239** Paper No. 105369. MR4708611 https://doi.org/10.1016/j.jeconom.2022.12.003

[56] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079

[57] JONES, A. and RUBIN-DELANCHY, P. (2021). The multilayer random dot product graph. Available at arXiv:2007.10455 [cs, stat].

[58] KADAVANKANDY, A. and COUILLET, R. (2019). Asymptotic Gaussian fluctuations of spectral clustering eigenvectors. In 2019 *IEEE* 8*th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (*CAMSAP*) 694–698.

[59] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[60] KE, Z. T. and WANG, J. (2022). Optimal network membership estimation under severe degree heterogeneity. Available at arXiv:2204.12087 [math, stat].

[61] KIM, B., LEE, K. H., XUE, L. and NIU, X. (2018). A review of dynamic network models with latent variables. *Stat. Surv.* **12** 105–135. MR3850294 https://doi.org/10.1214/18-SS121

[62] KLOEPFER, D., AVILES-RIVERO, A. I. and HEYDECKER, D. (2021). Delving into deep walkers: A convergence analysis of random-walk-based vertex embeddings. Available at arXiv:2107.10014 [cs, math, stat].

[63] KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167. MR1781185 https://doi.org/10.2307/3318636

[64] LANCICHINETTI, A. and FORTUNATO, S. (2009). Community detection algorithms: A comparative analysis. *Phys. Rev. E* **80** 056117.

[65] LEE, C. and WILKINSON, D. J. (2019). A review of stochastic block models and extensions for graph clustering. *Appl. Netw. Sci.* **4** 1–50.

[66] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

[67] LEI, J. (2021). Network representation using graph root distributions. *Ann. Statist.* **49** 745–768. MR4255106 https://doi.org/10.1214/20-aos1976

[68] LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. MR4064140 https://doi.org/10.1093/biomet/asz068

[69] LEI, J. and LIN, K. Z. (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *J. Amer. Statist. Assoc.* **118** 2433–2445. MR4681594 https://doi.org/10.1080/01621459.2022.2054817

[70] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

[71] LEI, L. (2019). Unified $\ell_{2\rightarrow\infty}$eigenspace perturbation theory for symmetric random matrices. Available at arXiv:1909.04798 [math, Stat].

[72] LEI, L., LI, X. and LOU, X. (2020). Consistency of spectral clustering on hierarchical stochastic block models. Available at arXiv:2004.14531 [math, Stat].

[73] LEVIN, K., ATHREYA, A., TANG, M., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2019). A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference. Available at arXiv:1705.09355 [stat].

[74] LEVIN, K. and LEVINA, E. (2025). Bootstrapping networks with latent space structure. *Electron. J. Stat.* **19** 745–791. MR4870301 https://doi.org/10.1214/25-ejs2347

[75] LI, Y. and LI, H. (2018). Two-sample test of community memberships of weighted stochastic block models. Available at arXiv:1811.12593 [math, Stat].

[76] LIN, C., SUSSMAN, D. L. and ISHWAR, P. (2023). Ergodic limits, relaxations, and geometric properties of random walk node embeddings. *IEEE Trans. Netw. Sci. Eng.* **10** 346–359. MR4545388 https://doi.org/10.1109/tnse.2022.3208520

[77] LIN, Q., LUNDE, R. and SARKAR, P. (2020). On the theoretical properties of the network jackknife. In *Proceedings of the* 37*th International Conference on Machine Learning* 6105–6115. PMLR. ISSN: 2640-3498.

[78] LIN, Q., LUNDE, R. and SARKAR, P. (2022). Trading off accuracy for speedup: Multiplier bootstraps for subgraph counts. Available at arXiv:2009.06170 [math, stat].

[79] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. American Mathematical Soc. Google-Books-ID: FsFqHLid8sAC.

[80] LU, L. and PENG, X. (2013). Spectra of edge-independent random graphs. *Electron. J. Combin.* **20** Paper 27. MR3158266 https://doi.org/10.37236/3576

[81] LUNDE, R. and SARKAR, P. (2023). Subsampling sparse graphons under minimal assumptions. *Biometrika* **110** 15–32. MR4565441 https://doi.org/10.1093/biomet/asac032

[82] LYU, Z., XIA, D. and ZHANG, Y. (2023). Latent space model for higher-order networks and generalized tensor decomposition. *J. Comput. Graph. Statist.* **32** 1320–1336. MR4669249 https://doi.org/10.1080/10618600.2022.2164289

[83] MA, Z., MA, Z. and YUAN, H. (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.* **21** Paper No. 4. MR4071187

[84] MA, Z. and NANDY, S. (2023). Community detection with contextual multilayer networks. *IEEE Trans. Inf. Theory* **69** 3203–3239. MR4581286 https://doi.org/10.1109/tit.2023.3238352

[85] MACDONALD, P. W., LEVINA, E. and ZHU, J. (2022). Latent space models for multiplex networks with shared structure. *Biometrika* **109** 683–706. MR4472842 https://doi.org/10.1093/biomet/asab058

[86] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2018). Overlapping clustering models, and one (class) SVM to bind them all. In *Advances in Neural Information Processing Systems* **31**. Curran Associates, Red Hook.

[87] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2021). Estimating mixed memberships with sharp eigenvector deviations. *J. Amer. Statist. Assoc.* **116** 1928–1940. MR4353723 https://doi.org/10.1080/01621459.2020.1751645

[88] MATIAS, C. and ROBIN, S. (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. In *MMCS, Mathematical Modelling of Complex Systems. ESAIM Proc. Surveys* **47** 55–74. EDP Sci., Les Ulis. MR3419385 https://doi.org/10.1051/proc/201447004

[89] MAUGIS, P.-A. G., OLHEDE, S. C., PRIEBE, C. E. and WOLFE, P. J. (2020). Testing for equivalence of network distribution using subgraph counts. *J. Comput. Graph. Statist.* **29** 455–465. MR4153173 https://doi.org/10.1080/10618600.2020.1736085

[90] MAUGIS, P. A. (2024). Central limit theorems for local network statistics. *Biometrika* **111** 743–754. MR4794147 https://doi.org/10.1093/biomet/asad080

[91] MELE, A., HAO, L., CAPE, J. and PRIEBE, C. E. (2023). Spectral estimation of large stochastic blockmodels with discrete nodal covariates. *J. Bus. Econom. Statist.* **41** 1364–1376. MR4650468 https://doi.org/10.1080/07350015.2022.2139709

[92] MITRA, A., PRASAD, K. and CAPE, J. (2024). On inference for modularity statistics in structured networks. *J. Comput. Graph. Statist.* **33** 1370–1382. MR4829643 https://doi.org/10.1080/10618600.2024.2336147

[93] MODELL, A. and RUBIN-DELANCHY, P. (2021). Spectral clustering under degree heterogeneity: A case for the random walk Laplacian. Available at arXiv:2105.00987 [stat].

[94] MU, C., MELE, A., HAO, L., CAPE, J., ATHREYA, A. and PRIEBE, C. E. (2022). On spectral algorithms for community detection in stochastic blockmodel graphs with vertex covariates. *IEEE Trans. Netw. Sci. Eng.* **9** 3373–3384. MR4482415 https://doi.org/10.1109/tnse.2022.3177708

[95] NOROOZI, M. and PENSKY, M. (2022). The hierarchy of block models. *Sankhya A* **84** 64–107. MR4402747 https://doi.org/10.1007/s13171-021-00247-2

[96] NOROOZI, M., PENSKY, M. and RIMAL, R. (2021). Sparse popularity adjusted stochastic block model. *J. Mach. Learn. Res.* **22** Paper No. 193. MR4329772

[97] NOROOZI, M., RIMAL, R. and PENSKY, M. (2021). Estimation and clustering in popularity adjusted block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 293–317. MR4250277 https://doi.org/10.1111/rssb.12410

[98] PANTAZIS, K., ATHREYA, A., ARROYO, J., FROST, W. N., HILL, E. S. and LYZINSKI, V. (2022). The importance of being correlated: Implications of dependence in joint spectral inference across multiple networks. *J. Mach. Learn. Res.* **23** Paper No. [141]. MR4577093

[99] PAUL, S. and CHEN, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electron. J. Stat.* **10** 3807–3870. MR3579677 https://doi.org/10.1214/16-EJS1211

[100] PAUL, S. and CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Statist.* **48** 230–250. MR4065160 https://doi.org/10.1214/18-AOS1800

[101] PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47** 2378–2403. MR3953455 https://doi.org/10.1214/18-AOS1751

[102] PENSKY, M. and WANG, Y. (2024). Clustering of diverse multiplex networks. *IEEE Trans. Netw. Sci. Eng.* **11** 3441–3454. MR4762253

[103] PENSKY, M. and ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. MR3914178 https://doi.org/10.1214/19-ejs1533

[104] PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., CONROY, J. M., LYZINSKI, V., TANG, M., ATHREYA, A., CAPE, J. and BRIDGEFORD, E. (2019). On a two-truths phenomenon in spectral graph clustering. *Proc. Natl. Acad. Sci. USA* **116** 5995–6000. MR3939781 https://doi.org/10.1073/pnas.1814462116

[105] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* **26**. Curran Associates, Red Hook.

[106] RAZAEE, Z. S., AMINI, A. A. and LI, J. J. (2019). Matched bipartite block model with covariates. *J. Mach. Learn. Res.* **20** Paper No. 34. MR3948074

[107] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[108] ROHE, K. and ZENG, M. (2023). Vintage factor analysis with Varimax performs statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1037–1060. MR4716148 https://doi.org/10.1093/jrsssb/qkad029

[109] RUBIN-DELANCHY, P., CAPE, J., TANG, M. and PRIEBE, C. E. (2022). A statistical interpretation of spectral embedding: The generalised random dot product graph. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1446–1473. MR4494166 https://doi.org/10.1111/rssb.12509

[110] SARKAR, P. and BICKEL, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43** 962–990. MR3346694 https://doi.org/10.1214/14-AOS1285

[111] SCHAEFFER, S. E. (2007). Graph clustering. *Comput. Sci. Rev.* **1** 27–64.

[112] SENGUPTA, S. and CHEN, Y. (2018). A block model for node popularity in networks with community structure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 365–386. MR3763696 https://doi.org/10.1111/rssb.12245

[113] SHAO, M., XIA, D., ZHANG, Y., WU, Q. and CHEN, S. (2024). Higher-order accurate two-sample network inference and network hashing. Available at arXiv:2208.07573 [math, stat].

[114] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *J. Comput. Graph. Statist.* **26** 344–354. MR3640191 https://doi.org/10.1080/10618600.2016.1193505

[115] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872 https://doi.org/10.3150/15-BEJ789

[116] TANG, M., CAPE, J. and PRIEBE, C. E. (2022). Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral estimator. *Bernoulli* **28** 1049–1073. MR4388929 https://doi.org/10.3150/21-bej1376

[117] TANG, M. and PRIEBE, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.* **46** 2360–2415. MR3845021 https://doi.org/10.1214/17-AOS1623

[118] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803 https://doi.org/10.1007/s11222-007-9033-z

[119] WANG, Y. X. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528. MR3650391 https://doi.org/10.1214/16-AOS1457

[120] WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, Cambridge. Google-Books-ID: CAm2DpIqRUIC.

[121] XIE, F. (2023). Euclidean representation of low-rank matrices and its geometric properties. *SIAM J. Matrix Anal. Appl.* **44** 822–866. MR4597265 https://doi.org/10.1137/22M1489125

[122] XIE, F. (2024). Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals. *Bernoulli* **30** 388–418. MR4665583 https://doi.org/10.3150/23-bej1602

[123] XIE, F. and XU, Y. (2023). Efficient estimation for random dot product graphs via a one-step procedure. *J. Amer. Statist. Assoc.* **118** 651–664. MR4571148 https://doi.org/10.1080/01621459.2021.1948419

[124] XIE, F. and ZHANG, Y. (2024). Higher-order entrywise eigenvectors analysis of low-rank random matrices: Bias correction, Edgeworth expansion, and bootstrap. Available at arXiv:2401.15033 [math, stat].

[125] XU, S., ZHEN, Y. and WANG, J. (2023). Covariate-assisted community detection in multi-layer networks. *J. Bus. Econom. Statist.* **41** 915–926. MR4600858 https://doi.org/10.1080/07350015.2022.2085726

[126] YAN, Y., CHEN, Y. and FAN, J. (2024). Inference for heteroskedastic PCA with missing data. *Ann. Statist.* **52** 729–756. MR4744194 https://doi.org/10.1214/24-aos2366

[127] YOUNG, S. J. and SCHEINERMAN, E. (2008). Directed random dot product graphs. *Internet Math.* **5** 91–111. MR2560264

[128] ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

[129] ZHANG, H., GUO, X. and CHANG, X. (2022). Randomized spectral clustering in large-scale stochastic block models. *J. Comput. Graph. Statist.* **31** 887–906. MR4495720 https://doi.org/10.1080/10618600.2022.2034636

[130] ZHANG, L. and AMINI, A. A. (2023). Adjusted chi-square test for degree-corrected block models. *Ann. Statist.* **51** 2366–2385. MR4682701 https://doi.org/10.1214/23-aos2329

[131] ZHANG, Y., CHEN, K., SAMPSON, A., HWANG, K. and LUNA, B. (2019). Node features adjusted stochastic block model. *J. Comput. Graph. Statist.* **28** 362–373. MR3974886 https://doi.org/10.1080/10618600.2018.1530117

[132] ZHANG, Y., LEVINA, E. and ZHU, J. (2020). Detecting overlapping communities in networks using spectral methods. *SIAM J. Math. Data Sci.* **2** 265–283. MR4081803 https://doi.org/10.1137/19M1272238

[133] ZHANG, Y. and TANG, M. (2022). Perturbation analysis of randomized SVD and its applications to high-dimensional statistics. Available at arXiv:2203.10262 [cs, math, stat].

[134] ZHANG, Y. and TANG, M. (2024). A theoretical analysis of DeepWalk and Node2vec for exact recovery of community structures in stochastic blockmodels. *IEEE Trans. Pattern Anal. Mach. Intell.* **46** 1065–1078.

[135] ZHANG, Y. and XIA, D. (2022). Edgeworth expansions for network moments. *Ann. Statist.* **50** 726–753. MR4404918 https://doi.org/10.1214/21-aos2125

[136] ZHAO, Y. (2017). A survey on theoretical advances of community detection in networks. *Wiley Interdiscip. Rev.: Comput. Stat.* **9** e1403. MR3696558 https://doi.org/10.1002/wics.1403

[137] ZHENG, R. and TANG, M. (2022). Limit results for distributed estimation of invariant subspaces in multiple networks inference and PCA. Available at arXiv:2206.04306 [math, stat].

Statistical Science
2026, Vol. 41, No. 2, 386–399
https://doi.org/10.1214/24-STS971

Approximate Inference with Exponential Tilting Densities: Theory and Applications

Xiaoping Shi, Xiang-Sheng Wang, Augustine Wong and Wei Lin

Abstract. A family of exponential tilting density functions (ETD) is presented and compared with energy functions. This ETD family is shown to be associated with the normal density and the log-gamma density by the minimum cross entropy in information theory. In this paper, we show that ETD minimizes the Kulback–Leiber divergence under some moments constraints. Two examples are provided to illustrate how to approximate a baseline density using ETD. In addition, the normalizing constant of the ETD is approximated by three commonly used approximation methods: Gaussian variational approximation (GVA), Laplace approximation (LA) and saddlepoint approximation (SA). It is shown that the normalizing constant obtained by GVA and SA are asymptotically equivalent, and theoretically, both are more accurate than the one obtained by LA. With the availability of the normalizing constant, likelihood-based asymptotic inference can be obtained. To demonstrate the applicability of the proposed method, it is applied to parameter estimation of the Poisson mixed model and Bayesian inference.

Key words and phrases: Kullback–Leiber divergence, Laplace approximation, saddelpoint approximation, Gaussian variational approximation, Poisson mixed model, Bayesian inference, Poisson log-normal distribution.

REFERENCES

[1] ACKLEY, D. H., HINTON, G. E. and SEJNOWSKI, T. J. (1985). A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9** 147–169.

[2] ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

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

[4] BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and Asymptotics*. *Monographs on Statistics and Applied Probability* **52**. CRC Press, London. MR1317097 https://doi.org/10.1007/978-1-4899-3210-5

[5] BELANGER, D. and MCCALLUM, A. (2016). Structured prediction energy networks. In *Proceedings of the* 33rd *International Conference on Machine Learning*, New York, USA.

[6] 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

[7] BRAZZALE, A. R., DAVISON, A. C. and REID, N. (2007). *Applied Asymptotics*: *Case Studies in Small-Sample Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **23**. Cambridge Univ. Press, Cambridge. MR2342742 https://doi.org/10.1017/CBO9780511611131

[8] BRESLOW, N. E. and CLAYTON, D. G. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **41** 179–312.

[9] CHERNOFF, H. (1972). *Sequential Analysis and Optimal Design*. *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics* **8**. SIAM, Philadelphia, PA. MR0309249

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

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

[12] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. With 1 IBM-PC floppy disk (3.5 inch; HD). MR1478673 https://doi.org/10.1017/CBO9780511802843

Xiaoping Shi is an Assistant Professor, Department of Computer Science, Mathematics, Physics and Statistics, University of British Columbia, Kelowna, British Columbia V1V 1V7, Canada (e-mail: xiaoping.shi@ubc.ca). Xiang-Sheng Wang is an Associate Professor, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana 70503, USA (e-mail: xswang@louisiana.edu). Augustine Wong is a Professor, Department of Mathematics and Statistics,York University, Toronto, Ontario M3J 1P3, Canada (e-mail: august@yorku.ca). Wei Lin is a Lecturer, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada (e-mail: becky_lin@sfu.ca).

[13] DESJARDINS, G., COURVILLE, A., BENGIO, Y., VINCENT, P. and DELALLEAU, O. (2010). Parallel tempering for training of restricted Boltzmann machines. In 13*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*), Chia Laguna Resort, Sardinia, Italy.

[14] DOUCET, A., DE FREITAS, N. and GORDON, N., eds. (2001). *Sequential Monte Carlo Methods in Practice*. *Statistics for Engineering and Information Science*. Springer, New York. MR1847783 https://doi.org/10.1007/978-1-4757-3437-9

[15] EFRON, B. (1981). Nonparametric standard errors and confidence intervals. *Canad*. *J*. *Statist*. **9** 139–172. With discussion and a reply by the author. MR0640014 https://doi.org/10.2307/3314608

[16] EFRON, B. and TIBSHIRANI, R. (1996). Using specially designed exponential families for density estimation. *Ann*. *Statist*. **24** 2431–2461. MR1425960 https://doi.org/10.1214/aos/1032181161

[17] ESSCHER, F. (1932). On the probability function in the collective theory of risk. *Scand*. *Actuar*. *J*. **15** 175–195.

[18] GERBER, H. U. and SHIU, E. S. W. (1994). Option pricing by Esscher transforms. *Trans*. *Soc*. *Actuar*. **46** 99–140.

[19] GRATHWOHL, W., WANG, K.-C., JACOBSEN, J.-H., DUVENAUD, D., NOROUZI, M. and SWERSKY, K. (2020). Your classifier is secretly an energy based model and you should treat it like one. In *International Conference on Learning Representations*.

[20] HALL, P., ORMEROD, J. T. and WAND, M. P. (2011). Theory of Gaussian variational approximation for a Poisson mixed model. *Statist*. *Sinica* **21** 369–389. MR2796867

[21] 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

[22] HINKLEY, D. V., REID, N. and SNELL, E. J., eds. (1991). *Statistical Theory and Modelling*: *In Honour of Sir David Cox*, *FRS*. CRC Press, London. MR1120585

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

[24] HINTON, G. E. (2012). A practical guide to training restricted Boltzmann machines. In *Neural Networks*: *Tricks of the Trade* 599–619.

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

[26] JIANG, B., WU, T.-Y., JIN, Y. and WONG, W. H. (2018). Convergence of contrastive divergence algorithm in exponential family. *Ann*. *Statist*. **46** 3067–3098. MR3851765 https://doi.org/10.1214/17-AOS1649

[27] JORDAN, M. I. (2004). Graphical models. *Statist*. *Sci*. **19** 140–155. MR2082153 https://doi.org/10.1214/088342304000000026

[28] JUNEJA, S. and SHAHABUDDIN, P. (2006). Rare event simulation techniques: An introduction and recent advances. In *Research and Management Science* 291–350. Elsevier, Amsterdam, The Netherlands.

[29] KENT, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69** 19–27. MR0655667 https://doi.org/10.1093/biomet/69.1.19

[30] KULLBACK, S. (1997). *Information Theory and Statistics*. Dover, Mineola, NY. Reprint of the second (1968) edition. MR1461541

[31] KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann*. *Math*. *Stat*. **22** 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

[32] LAPLACE, P. S. (1774). Memoir on the probability of the causes of events. *Savants Estranges* **6** 621–656.

[33] LECUN, Y., CHOPRA, S., HADSELL, R., RANZATO, M. and HUANG, F. (2006). *A Tutorial on Energy-Based Learning*. *Predicting Structured Data*. MIT Press.

[34] LIU, Q., PENG, J. and IHLER, A. T. (2016). Variational bounding for distributed data. In *Proceedings of the* 33*rd International Conference on Machine Learning* (*ICML*) 2531–2540.

[35] MARTENS, J. and GROSSE, R. (2015). Optimizing neural networks with Kronecker-factored approximate curvature. In *Proceedings of the* 32*nd International Conference on Machine Learning* (*ICML*) 2408–2417.

[36] MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann*. *Statist*. **10** 65–80. MR0642719

[37] 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

[38] ORMEROD, J. T. and WAND, M. P. (2012). Gaussian variational approximate inference for generalized linear mixed models. *J*. *Comput*. *Graph*. *Statist*. **21** 2–17. MR2913353 https://doi.org/10.1198/jcgs.2011.09118

[39] OWEN, A. and ZHOU, Y. (2000). Safe and effective importance sampling. *J*. *Amer*. *Statist*. *Assoc*. **95** 135–143. MR1803146 https://doi.org/10.2307/2669533

[40] PACE, L., SALVAN, A. and VENTURA, L. (2006). Likelihood-based discrimination between separate scale and regression models. *J*. *Statist*. *Plann*. *Inference* **136** 3539–3553. MR2256275 https://doi.org/10.1016/j.jspi.2005.01.014

[41] POURHASSAN, B., SAKALLI, İ., SHI, X., FAIZAL, M. and WANI, S. S. (2023). Quantum thermodynamics of an α'-corrected Reissner–Nordström black hole. *Europhys*. *Lett*. **144** 29001.

[42] REID, N. (1988). Saddlepoint methods and statistical inference. *Statist*. *Sci*. **3** 213–238. With comments and a rejoinder by the author. MR0968390

[43] 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

[44] SALAKHUTDINOV, R. and HINTON, G. E. (2009). Deep Boltzmann machines. In *Proceedings of the* 12*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*), Clearwater Beach, FL, USA.

[45] SELVARAJ, D. and DHANASEKARAN, R. (2013). A review on current MRI brain tissue segmentation, feature extraction and classification techniques. *International Journal of Electronics*, *Communication & Instrumentation Engineering Research and Development* **3** 11–30.

[46] SHI, X., WANG, X.-S. and WONG, A. (2022). Explicit Gaussian variational approximation for the Poisson lognormal mixed model. *Mathematics* **10** 4542.

[47] SKOVGAARD, I. M. (2017). *Inference*, *Asymptotics*, *and Applications*. World Scientific Co. Pte. Ltd., Hackensack, NJ. Selected papers of Ib Michael Skovgaard, with introductions by his colleagues, Edited by Nancy Reid and Torben Martinussen. MR3701417

[48] SONG, Y. and ERMON, S. (2019). Generative modeling by estimating gradients of the data distribution. arXiv:1907.05600.

[49] THRUN, S., BURGARD, W. and FOX, D. (2005). *Probabilistic Robotics*. MIT Press.

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

[51] TITTERINGTON, D. M. (2004). Bayesian methods for neural networks and related models. *Statist*. *Sci*. **19** 128–139. MR2082152 https://doi.org/10.1214/088342304000000099

[52] TOMCZAK, J. M. (2022). *Deep Generative Modeling*. Springer, Cham.

[53] WAINWRIGHT, M. J., JAAKKOLA, T. S. and WILLSKY, A. S. (2005). A new class of upper bounds on the log partition function. *IEEE Trans*. *Inf*. *Theory* **51** 2313–2335. MR2246364 https://doi.org/10.1109/TIT.2005.850091

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

[55] WELLING, M., ROSEN-ZVL, M. and HINTON, G. (2004). Exponential family harmoniums with an application to information retrieval. *Adv. Neural Inf. Process*. *Syst*. **17**.

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

[57] XIE, J., LU, Y., GAO, R. and ZHU, S. C. (2016). A theory of generative convNet. In *Proceedings of the* 33*rd International Conference on Machine Learning* **48** 2635–2644.

[58] XIE, J., ZHU, S.-C. and WU, Y. N. (2021). Learning energy-based spatial-temporal generative convNets for dynamic patterns. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **43** 516–531.

Statistical Science
2026, Vol. 41, No. 2, 400–410
https://doi.org/10.1214/25-STS975

A Conversation with Byron J.T. Morgan

Rachel McCrea

Abstract. Byron Morgan is an Emeritus Professor of Statistics, a Fellow of the Learned Society of Wales and represents a true pillar of the applied statistics community, contributing appreciable understanding to a number of fields, most recently statistical ecology.

Byron has offered sustained service to the community, chairing the last Research Excellence Framework statistics panel in 2001 (after which statistics was merged with mathematics) and has reviewed and advised several external agencies at home and abroad. He has been editor of the Journal of Agricultural, Biological and Environmental Statistics, co-editor of Biometrics and of Applied Statistics and Associate Editor and Guest Editor for several other journals. He also served in multiple learned societies, most notably the International Biometric Society, (of which he was the president in 1996/1997 and Elected Honorary Life Member in 2014) and in the Royal Statistical Society, where he was a member of council and a vice-president during 1997–2001. His contributions to the organisation of scientific conferences are many and include his serving as Chair of Scientific Program Committee of the XVIIth International Biometric Conference in Hamilton, Ontario, 1994, and as Chair of the local organising committee for the International Statistical Ecology Conference, held in Canterbury in 2010.

As well as leading many initiatives and grants, he has mentored and inspired generations of researchers, many of whom have gone on to research careers in both academia and NGOs. Anyone who has met Byron will not be surprised that it is a great pleasure to have an excuse to talk to him about his career and for him to relay some of his many entertaining stories.

Key words and phrases: Capture-recapture, integrated population modelling, National Centre for Statistical Ecology, International Statistical Ecology Conference, parameter redundancy, population dynamics.

REFERENCES

[1] BINGHAM, N. H. (1996). A conversation with David Kendall. *Statist. Sci.* **11** 159–188. MR1436646 https://doi.org/10.1214/ss/1032280213

[2] SMITH, A. (1995). A conversation with Dennis Lindley. *Statist. Sci.* **10** 305–319. MR1390520

Rachel McCrea is Professor of Statistics, School of Mathematical Sciences, Lancaster University, Lancaster, United Kingdom (e-mail: r.mccrea@lancaster.ac.uk).

Statistical Science
2026, Vol. 41, No. 2, 411–432
https://doi.org/10.1214/25-STS976

Re-Thinking Spatial Confounding in Spatial Linear Mixed Models

K. Khan and C. Berrett

Abstract. In the last two decades, considerable research has been devoted to a phenomenon known as spatial confounding. Spatial confounding is thought to occur when there is multicollinearity between a covariate and the random effect in a spatial regression model. This multicollinearity is considered highly problematic when the inferential goal is estimating regression coefficients and various methodologies have been proposed to attempt to alleviate it. Recently, it has become apparent that many of these methodologies are flawed, yet the field continues to expand. In this paper, we offer a novel perspective of synthesizing the work in the field of spatial confounding. We propose that at least two distinct phenomena are currently conflated with the term spatial confounding. We refer to these as the "analysis model" and the "data generation" types of spatial confounding. We show that these two issues can lead to contradicting conclusions about whether spatial confounding exists and whether methods to alleviate it will improve inference. Our results also illustrate that in most cases, traditional spatial linear mixed models *do* help to improve inference on regression coefficients. Drawing on the insights gained, we offer a path forward for research in spatial confounding.

Key words and phrases: Spatial regression, geostatistics, areal data, omitted variable bias.

REFERENCES

ADIN, A., GOICOA, T., HODGES, J. S., SCHNELL, P. M. and UGARTE, M. D. (2023). Alleviating confounding in spatio-temporal areal models with an application on crimes against women in India. *Stat. Model.* **23** 9–30. MR4549292 https://doi.org/10.1177/1471082X211015452

AZEVEDO, D. R. M., PRATES, M. O. and BANDYOPADHYAY, D. (2021). MSPOCK: Alleviating spatial confounding in multivariate disease mapping models. *J. Agric. Biol. Environ. Stat.* **26** 464–491. MR4292798 https://doi.org/10.1007/s13253-021-00451-5

AZEVEDO, D. R. M., PRATES, M. O. and BANDYOPADHYAY, D. (2023). Alleviating spatial confounding in frailty models. *Biostatistics* **24** 945–961. MR4656769 https://doi.org/10.1093/biostatistics/kxac028

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* CRC Press, Boca Raton, FL. MR3362184

BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48. https://doi.org/10.18637/jss.v067.i01

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822 https://doi.org/10.1007/BF00116466

BIVAND, R. S., PEBESMA, E. J., GÓMEZ-RUBIO, V. and PEBESMA, E. J. (2008). *Applied Spatial Data Analysis with R* **747248717**. Springer, Berlin.

CHIOU, Y.-H., YANG, H.-D. and CHEN, C.-S. (2019). An adjusted parameter estimation for spatial regression with spatial confounding. *Stoch. Environ. Res. Risk Assess.* **33** 1535–1551.

CLAYTON, D. G., BERNARDINELLI, L. and MONTOMOLI, C. (1993). Spatial correlation in ecological analysis. *Int. J. Epidemiol.* **22** 1193–1202. https://doi.org/10.1093/ije/22.6.1193

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics. *J. R. Stat. Soc., Ser. C* **47** 299–350. MR1626544 https://doi.org/10.1111/1467-9876.00113

DUPONT, E., WOOD, S. N. and AUGUSTIN, N. H. (2022). Spatial+: A novel approach to spatial confounding. *Biometrics* **78** 1279–1290. MR4534358 https://doi.org/10.1111/biom.13656

GARNIER, S., ROSS, N., RUDIS, B., SCIAINI, M., CAMARGO, A. P. and SCHERER, C. (2024). viridis(Lite)—Colorblind-friendly color maps for R. viridis package version 0.6.5.

HANKS, E. M., SCHLIEP, E. M., HOOTEN, M. B. and HOETING, J. A. (2015). Restricted spatial regression in practice: Geostatistical models, confounding, and robustness under model misspecifica-

K. Khan (e-mail: kleighkhan@gmail.com). C. Berrett is Professor, Department of Statistics, Brigham Young University, Provo, Utah 84604, USA (e-mail: cberrett@stat.byu.edu).

tion. *Environmetrics* **26** 243–254. MR3340961 https://doi.org/10.1002/env.2331

HEFLEY, T. J., HOOTEN, M. B., HANKS, E. M., RUSSELL, R. E. and WALSH, D. P. (2017). The Bayesian group lasso for confounded spatial data. *J. Agric. Biol. Environ. Stat.* **22** 42–59. MR3607653 https://doi.org/10.1007/s13253-016-0274-1

HODGES, J. S. and REICH, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *Amer. Statist.* **64** 325–334. MR2758564 https://doi.org/10.1198/tast.2010.10052

HUGHES, J. and CUI, X. (2020). ngspatial: Fitting the centered autologistic and sparse spatial generalized linear mixed models for areal data, Frederick, MD. R package version 1.2-2.

HUGHES, J. and HARAN, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 139–159. MR3008275 https://doi.org/10.1111/j.1467-9868.2012.01041.x

HUI, F. K. C. and BONDELL, H. D. (2022). Spatial confounding in generalized estimating equations. *Amer. Statist.* **76** 238–247. MR4453526 https://doi.org/10.1080/00031305.2021.2009372

KELLER, J. P. and SZPIRO, A. A. (2020a). Selecting a scale for spatial confounding adjustment. *J. Roy. Statist. Soc. Ser. A* **183** 1121–1143. MR4114479 https://doi.org/10.1111/rssa.12556

KELLING, C., GRAIF, C., KORKMAZ, G. and HARAN, M. (2021). Modeling the social and spatial proximity of crime: Domestic and sexual violence across neighborhoods. *J. Quant. Criminol.* **37** 481–516. https://doi.org/10.1007/s10940-020-09454-w

KHAN, K. and BERRETT, C. (2026). Supplement to "Re-thinking spatial confounding in spatial linear mixed models." https://doi.org/10.1214/25-STS976SUPPA, https://doi.org/10.1214/25-STS976SUPPB, https://doi.org/10.1214/25-STS976SUPPC

KHAN, K. and CALDER, C. A. (2022). Restricted spatial regression methods: Implications for inference. *J. Amer. Statist. Assoc.* **117** 482–494. MR4399100 https://doi.org/10.1080/01621459.2020.1788949

KIMELDORF, G. S. and WAHBA, G. (1970). Spline functions and stochastic processes. *Sankhyā Ser. A* **32** 173–180. MR0303594

MARQUES, I., KNEIB, T. and KLEIN, N. (2022). Mitigating spatial confounding by explicitly correlating Gaussian random fields. *Environmetrics* **33** Paper No. e2727, 20 pp. MR4453440 https://doi.org/10.1002/env.2727

NARCISI, M., GRECO, F. and TRIVISANO, C. (2024). On the effect of confounding in linear regression models: An approach based on the theory of quadratic forms. *Environ. Ecol. Stat.*.

NOBRE, W. S., SCHMIDT, A. M. and PEREIRA, J. B. M. (2021). On the effects of spatial confounding in hierarchical models. *Int. Stat. Rev.* **89** 302–322. MR4411907 https://doi.org/10.1111/insr.12407

NYCHKA, D., FURRER, R., PAIGE, J. and SAIN, S. (2021). fields: Tools for spatial data. Univ. Corporation for Atmospheric Research, Boulder, CO, USA. R package version 14.0.

PACIOREK, C. (2009). Technical Vignette 5: Understanding intrinsic Gaussian Markov random field spatial models, including intrinsic conditional autoregressive models. Technical report.

PACIOREK, C. J. (2007). Computational techniques for spatial logistic regression with large data sets. *Comput. Statist. Data Anal.* **51** 3631–3653. MR2364480 https://doi.org/10.1016/j.csda.2006.11.008

PACIOREK, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statist. Sci.* **25** 107–125. MR2741817 https://doi.org/10.1214/10-STS326

PAGE, G. L., LIU, Y., HE, Z. and SUN, D. (2017). Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scand. J. Stat.* **44** 780–797. MR3687973 https://doi.org/10.1111/sjos.12275

PAPADOGEORGOU, G., CHOIRAT, C. and ZIGLER, C. M. (2019). Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics* **20** 256–272. MR3922132 https://doi.org/10.1093/biostatistics/kxx074

PEBESMA, E. (2018). Simple features for R: Standardized support for spatial vector data. *R J.* **10** 439–446. https://doi.org/10.32614/RJ-2018-009

PRATES, M. O., ASSUNÇÃO, R. M. and RODRIGUES, E. C. (2019). Alleviating spatial confounding for areal data problems by displacing the geographical centroids. *Bayesian Anal.* **14** 623–647. MR3959875 https://doi.org/10.1214/18-BA1123

REICH, B. J. and HODGES, J. S. (2008). Identification of the variance components in the general two-variance linear model. *J. Statist. Plann. Inference* **138** 1592–1604. MR2427291 https://doi.org/10.1016/j.jspi.2007.05.046

REICH, B. J., HODGES, J. S. and ZADNIK, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62** 1197–1206. MR2307445 https://doi.org/10.1111/j.1541-0420.2006.00617.x

REICH, B. J., YANG, S., GUAN, Y., GIFFIN, A. B., MILLER, M. J. and RAPPOLD, A. (2021). A review of spatial causal inference methods for environmental and epidemiological applications. *Int. Stat. Rev.* **89** 605–634. MR4411920 https://doi.org/10.1111/insr.12452

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. and TJELMELAND, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Stat.* **29** 31–49. MR1894379 https://doi.org/10.1111/1467-9469.00058

SARKAR, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.

SCHMIDT, A. M. (2021). Discussion on "Spatial+: A novel approach to spatial confounding" by Emiko Dupont, Simon N. Wood, and Nicole H. Augustin. *Biometrics*.

TEAM, R. C. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

THADEN, H. and KNEIB, T. (2018). Structural equation models for dealing with spatial confounding. *Amer. Statist.* **72** 239–252. MR3836447 https://doi.org/10.1080/00031305.2017.1305290

TOBLER, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46** 234–240.

VENABLES, W. N. and RIPLEY, B. D. (2002). *Modern Applied Statistics with S-Plus*, 4th ed. *Statistics and Computing*. Springer, New York. MR1337030 https://doi.org/10.1007/978-1-4899-2819-1

WALLER, L. A. and GOTWAY, C. A. (2004). *Applied Spatial Statistics for Public Health Data. Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2075123 https://doi.org/10.1002/0471662682

WICKHAM, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

WICKHAM, H., FRANCOIS, R., HENRY, L., MÜLLER, K. and VAUGHAN, D. (2023). dplyr: A grammar of data manipulation. R package version 1.1.4.

WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 https://doi.org/10.1201/9781420072884-c8

WOOD, S. and WOOD, M. S. (2015). Package 'mgcv'. R Package Version 1, 729.

WOOD, S. N. (2003). Thin plate regression splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 95–114. MR1959095 https://doi.org/10.1111/1467-9868.00374

WOOD, S. N. (2017). *Generalized Additive Models*: *An Introduction with R*, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3726911

YANG, H.-D., CHIOU, Y.-H. and CHEN, C.-S. (2023). Estimation and selection for spatial confounding regression models. *Comm*. *Statist*. *Theory Methods* **52** 939–955. MR4538929 https://doi.org/10.1080/03610926.2021.1934025

ZIMMERMAN, D. L. and VER HOEF, J. M. (2022). On deconfounding spatial confounding in linear models. *Amer*. *Statist*. **76** 159–167. MR4413148 https://doi.org/10.1080/00031305.2021.1946149

Statistical Science
2026, Vol. 41, No. 2, 433–459
https://doi.org/10.1214/25-STS977

The Intersection of Location-Allocation, Partitional Clustering and Model-Based Clustering Techniques—a Review

Tero Lähderanta, Lauri Lovén, Leena Ruha, Teemu Leppänen, Markku Kuismin, Ilkka Launonen, Susanna Pirttikangas, Jukka Riekki and Mikko J. Sillanpää

Abstract. Location-allocation, partitional spatial clustering and model-based clustering all deal with spatial data, seemingly from different viewpoints. Partitional clustering analyzes data points by partitioning them into separate groups, location-allocation places facilities in locations that best meet the needs of demand points, while model-based clustering provides a probabilistic interpretation of the underlying clusters. However, both partitional clustering and location-allocation can be formulated as optimization problems minimizing the distances of (demand) points from their associated centers (facilities). This common framework further connects to model-based clustering by means of likelihood maximization. Moreover, all of these approaches consider extensions such as weighted data points, different distance metrics, capacity constraints, different membership types, outliers and selecting the number of clusters or facilities.

In this article, we highlight and review the similarities and differences of these techniques. We look at a number of extensions common for all approaches, and consider their treatment under the framework. Finally, we provide a number of detailed examples highlighting the effects of extensions.

Key words and phrases: Clustering, spatial clustering, location-allocation, location analytics, finite mixture model.

REFERENCES

ACKERMAN, M., BEN-DAVID, S., BRÂNZEI, S. and LOKER, D. (2012). Weighted clustering. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. *AAAI'*12 858–863. AAAI Press, Menlo Park, CA.

ACKERMANN, M. R., BLÖMER, J. and SOHLER, C. (2010). Clustering for metric and nonmetric distance measures. *ACM Trans. Algorithms* **6** 1–26. MR2760422 https://doi.org/10.1145/1824777.1824779

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. Akad. Kiadó, Budapest. MR0483125

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **AC-19** 716–723. MR0423716 https://doi.org/10.1109/tac.1974.1100705

ARABANI, A. B. and FARAHANI, R. Z. (2012). Facility location dynamics: An overview of classifications and applications. *Comput. Ind. Eng.* **62** 408–420. https://doi.org/10.1016/j.cie.2011.09.018

ARAS, N., ALTINEL, İ. K. and ORBAY, M. (2007). New heuristic methods for the capacitated multi-facility Weber problem. *Naval Res.*

Tero Lähderanta is a Postdoctoral Researcher at the Research Unit for Mathematical Sciences, University of Oulu, Oulu, Finland (e-mail: tero.lahderanta@oulu.fi). Lauri Lovén is a Postdoctoral Researcher at the Center for Ubiquitous Computing, University of Oulu, Oulu, Finland (e-mail: lauri.loven@oulu.fi). Leena Ruha is a Senior Researcher at the Natural Resources Institute Finland, and University of Oulu, Finland (e-mail: leena.ruha@luke.fi). Teemu Leppänen is a Principal Lecturer at the Oulu University of Applied Sciences, Oulu, Finland (e-mail: teemu.leppanen@oamk.fi). Markku Kuismin is a Postdoctoral Researcher at the Research Unit for Mathematical Sciences, University of Oulu, Oulu, Finland (e-mail: markku.kuismin@oulu.fi). Ilkka Launonen is a Postdoctoral Researcher at the Research Unit for Mathematical Sciences, University of Oulu, Oulu, Finland (e-mail: ilkka.launonen@oulu.fi). Susanna Pirttikangas is a Docent at the Center for Ubiquitous Computing, University of Oulu, Finland (e-mail: susanna.pirttikangas@oulu.fi). Jukka Riekki is the Dean of the Faculty of Information Technology and Electrical Engineering and a full Professor at the Center for Ubiquitous Computing, University of Oulu, Finland (e-mail: jukka.riekki@oulu.fi). Mikko J. Sillanpää is a full Professor at the Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland (e-mail: mikko.sillanpaa@oulu.fi).

Logist. **54** 21–32. MR2285482 https://doi.org/10.1002/nav.20176

ARSLAN, O. (2010). An alternative multivariate skew Laplace distribution: Properties and estimation. *Statist*. *Papers* **51** 865–887. MR2749005 https://doi.org/10.1007/s00362-008-0183-7

AZARMAND, Z. and NEISHABOURI, E. (2009). Location allocation problem. In *Facility Location* 93–109. *Contrib*. *Manage*. *Sci*. https://doi.org/10.1007/978-3-7908-2151-2_5

BAADEL, S., THABTAH, F. and LU, J. (2016). Overlapping clustering: A review. In 2016 *SAI Computing Conference* (*SAI*) 233–237. IEEE. https://doi.org/10.1109/SAI.2016.7555988

BANERJEE, A. and GHOSH, J. (2006). Scalable clustering algorithms with balancing constraints. *Data Min*. *Knowl*. *Discov*. **13** 365–395. MR2252382 https://doi.org/10.1007/s10618-006-0040-z

BANERJEE, A., KRUMPELMAN, C., GHOSH, J., BASU, S. and MOONEY, R. J. (2005a). Model-based overlapping clustering. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining* 532–537. ACM, New York. https://doi.org/10.1145/1081870.1081932

BANERJEE, A., MERUGU, S., DHILLON, I. S. and GHOSH, J. (2005b). Clustering with Bregman divergences. *J*. *Mach*. *Learn*. *Res*. **6** 1705–1749. MR2249870

BANFIELD, J. D. and RAFTERY, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49** 803–821. MR1243494 https://doi.org/10.2307/2532201

BARANWAL, M. and SALAPAKA, S. M. (2017). Clustering with capacity and size constraints: A deterministic approach. In 2017 *Indian Control Conference* (*ICC*) 251–256. https://doi.org/10.1109/INDIANCC.2017.7846483

BEDRICK, E. J., CHRISTENSEN, R. and JOHNSON, W. (1996). A new perspective on priors for generalized linear models. *J*. *Amer*. *Statist*. *Assoc*. **91** 1450–1460. MR1439085 https://doi.org/10.2307/2291571

BEN-ISRAEL, A. and IYIGUN, C. (2008). Probabilistic D-clustering. *J*. *Classification* **25** 5–26. MR2429670 https://doi.org/10.1007/s00357-008-9002-z

BENAGLIA, T., CHAUVEAU, D., HUNTER, D. R. and YOUNG, D. (2009). Mixtools: An R package for analyzing finite mixture models. *J*. *Stat*. *Softw*. **32** 1–29. https://doi.org/10.18637/jss.v032.i06

BERKELAAR, M. et al. (2015). lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs. R package version 5.6.13.

BEZDEK, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York. With a foreword by L. A. Zadeh, Advanced Applications in Pattern Recognition. MR0631231

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

BORGWARDT, S., BRIEDEN, A. and GRITZMANN, P. (2017). An LP-based k-means algorithm for balancing weighted point sets. *European J*. *Oper*. *Res*. **263** 349–355. https://doi.org/10.1016/j.ejor.2017.04.054

BOUGUESSA, M. (2014). A mixture model-based combination approach for outlier detection. *Int*. *J*. *Artif*. *Intell*. *Tools* **23** 1460021. https://doi.org/10.1142/S0218213014600215

BRÈGMAN, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput*. *Math*. *Math*. *Phys*. **7** 200–217. https://doi.org/10.1016/0041-5553(67)90040-7

BRIMBERG, J., HANSEN, P., MLADENOVIĆ, N. and SALHI, S. (2008). A survey of solution methods for the continuous location-allocation problem. *Int*. *J*. *Oper*. *Res*. (*Taichung*) **5** 1–12. MR2406830

BRODERICK, T., KULIS, B. and JORDAN, M. (2013). MAD-Bayes: MAP-based asymptotic derivations from Bayes. In *Proceedings of the* 30*th International Conference on Machine Learning* (S. Dasgupta and D. McAllester, eds.). *Proceedings of Machine Learning Research* **28** 226–234. PMLR, Atlanta, GA.

CANÓS, M. J., IVORRA, C. and LIERN, V. (2001). The fuzzy p-median problem: A global analysis of the solutions. *European J*. *Oper*. *Res*. **130** 430–436.

CEBECAUER, M. and BUZNA, Ľ. (2018). Large-scale test data set for location problems. *Data Brief* **17** 267–274. https://doi.org/10.1016/j.dib.2018.01.008

CELEUX, G. and GOVAERT, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Comput*. *Statist*. *Data Anal*. **14** 315–332. MR1192205 https://doi.org/10.1016/0167-9473(92)90042-E

CELEUX, G. and GOVAERT, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognit*. **28** 781–793.

CERIOLI, A. (2005). K-means cluster analysis and Mahalanobis metrics: A problematic match or an overlooked opportunity. *Stat*. *Appl*. **17** 61–73.

CHARIKAR, M., KHULLER, S., MOUNT, D. M. and NARASIMHAN, G. (2001). Algorithms for facility location problems with outliers (extended abstract). In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* (*Washington*, *DC*, 2001) 642–651. SIAM, Philadelphia, PA. MR1958538

CHOKNIWAL, A. and SINGH, M. (2016). Faster Mahalanobis K-means clustering for Gaussian distributions. In 2016 *International Conference on Advances in Computing*, *Communications and Informatics* (*ICACCI*) 947–952. IEEE.

CHURCH, R. and REVELLE, C. (1974). The maximal covering location problem. *Pap*. *Reg*. *Sci*. **32** 101–118.

CHURCH, R. and WEAVER, J. (1986). Theoretical links between median and coverage location problems. *Ann*. *Oper*. *Res*. **6** 1–19.

CHURCH, R. L. and REVELLE, C. S. (1976). Theoretical and computational links between the p-median, location set-covering, and the maximal covering location problem. *Geogr*. *Anal*. **8** 406–415.

CLAESKENS, G. and HJORT, N. L. (2008). *Model Selection and Model Averaging*. *Cambridge Series in Statistical and Probabilistic Mathematics* **27**. Cambridge Univ. Press, Cambridge. MR2431297 https://doi.org/10.1017/CBO9780511790485

CLEUZIOU, G. (2008). An extended version of the k-means method for overlapping clustering. In 2008 19*th International Conference on Pattern Recognition* 1–4. https://doi.org/10.1109/ICPR.2008.4761079

OPEN GEOSPATIAL CONSORTIUM (2012). OGC white paper geospatial business intelligence (GeoBI).

COOPER, L. (1963). Location-allocation problems. *Oper*. *Res*. **11** 331–343. MR0152365 https://doi.org/10.1287/opre.11.3.331

COOPER, L. (1972). The transportation-location problem. *Oper*. *Res*. **20** 94–108. MR0321500 https://doi.org/10.1287/opre.20.1.94

CORD, A., AMBROISE, C. and COCQUEREZ, J.-P. (2006). Feature selection in robust clustering based on Laplace mixture. *Pattern Recogn*. *Lett*. **27** 627–635. https://doi.org/10.1016/j.patrec.2005.09.028

CUESTA-ALBERTOS, J. A., GORDALIZA, A. and MATRÁN, C. (1997). Trimmed k-means: An attempt to robustify quantizers. *Ann*. *Statist*. **25** 553–576. MR1439314 https://doi.org/10.1214/aos/1031833664

DAS, S., ABRAHAM, A. and KONAR, A. (2007). Automatic clustering using an improved differential evolution algorithm. *IEEE Trans*. *Syst*. *Man Cybern*., *Part A*, *Syst*. *Humans* **38** 218–237.

DE AMORIM, R. C. (2016). A survey on feature weighting based K-means algorithms. *J*. *Classification* **33** 210–242. MR3545533 https://doi.org/10.1007/s00357-016-9208-4

DUNN, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J*. *Cybern*. **3** 32–57. MR0375857 https://doi.org/10.1080/01969727308546046

ELLIOTT, M. R. (2011). A simple method to generate equal-sized homogenous strata or clusters for population-based sampling. *Ann. Epidemiol.* **21** 290–296. https://doi.org/10.1016/j.annepidem.2010.11.016

EVANS, K. N. (2013). *Extensions to Model-Based Clustering for Mixed-Type Data: A New Model Framework, Variable Selection, and Outlier Detection*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of Rochester. MR3218120

DEB, A. B. and DEY, L. (2017). Outlier detection and removal algorithm in K-means and hierarchical clustering. *World Journal of Computer Application and Technology* **5** 24–29.

FARAHANI, R. Z., DREZNER, Z. and ASGARI, N. (2009). Single facility location and relocation problem with time dependent weights and discrete planning horizon. *Ann. Oper. Res.* **167** 353–368. MR2506495 https://doi.org/10.1007/s10479-008-0338-x

FARAHANI, R. Z. and HEKMATFAR, M., eds. (2009). *Facility Location: Concepts, Models, Algorithms and Case Studies. Contributions to Management Science*. Physica-Verlag, Heidelberg. https://doi.org/10.1007/978-3-7908-2151-2

FARAHANI, R. Z., STEADIESEIFI, M. and ASGARI, N. (2010). Multiple criteria facility location problems: A survey. *Appl. Math. Model.* **34** 1689–1709. MR2594560 https://doi.org/10.1016/j.apm.2009.10.005

FORSTER, J. and WARMUTH, M. K. (2002). Relative expected instantaneous loss bounds. *J. Comput. System Sci.* **64** 76–102. MR1896143 https://doi.org/10.1006/jcss.2001.1798

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635 https://doi.org/10.1198/016214502760047131

FRIEDMAN, H. P. and RUBIN, J. (1967). On some invariant criteria for grouping data. *J. Amer. Statist. Assoc.* **62** 1159–1178. MR0223012

GANGANATH, N., CHENG, C. and TSE, C. K. (2014). Data clustering with cluster size constraints using a modified k-means algorithm. In 2014 *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* 158–161. https://doi.org/10.1109/CyberC.2014.36

GARBER, L. (2013). Analytics goes on location with new approaches. *Computer* **46** 14–17. https://doi.org/10.1109/MC.2013.123

GARCÍA-ESCUDERO, L. A., GORDALIZA, A., MATRÁN, C. and MAYO-ISCAR, A. (2010). A review of robust clustering methods. *Adv. Data Anal. Classif.* **4** 89–109. MR2720402 https://doi.org/10.1007/s11634-010-0064-5

GEBRU, I. D., ALAMEDA-PINEDA, X., FORBES, F. and HORAUD, R. (2016). EM algorithms for weighted-data clustering with application to audio-visual scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 2402–2415. https://doi.org/10.1109/TPAMI.2016.2522425

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2020). 22 finite mixture models. In *Bayesian Data Analysis* 3*rd Edition*. Chapman and Hall/CRC, London.

GORMLEY, I. C., MURPHY, T. B. and RAFTERY, A. E. (2023). Model-based clustering. *Annu. Rev. Stat. Appl.* **10** 573–595. MR4567806 https://doi.org/10.1146/annurev-statistics-033121-115326

GRUBESIC, T. H., WEI, R. and MURRAY, A. T. (2014). Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Ann. Assoc. Amer. Geogr.* **104** 1134–1156. https://doi.org/10.1080/00045608.2014.958389

GRÜN, B. (2019). Model-based clustering. In *Handbook of Mixture Analysis. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 157–192. CRC Press, Boca Raton, FL. MR3889693

GUO, Y., WANG, S., ZHOU, A., XU, J., YUAN, J. and HSU, C.-H. (2019). User allocation-aware edge cloud placement in mobile edge computing. *Softw. Pract. Exp.* **50** 489–502. https://doi.org/10.1002/spe.2685

GUPTA, M. R., CHEN, Y. et al. (2011). Theory and use of the EM algorithm. *Found. Trends Signal Process.* **4** 223–296.

GUPTA, S. (2017). A survey on balanced data clustering algorithms. *Int. J. Women Reseachers Eng. Sci. Manag.* **2** 2611–2614.

GUROBI OPTIMIZATION, L. (2018). Gurobi optimizer reference manual.

HAJIAGHAYI, M. T., MAHDIAN, M. and MIRROKNI, V. S. (2003). The facility location problem with general cost functions. *Networks* **42** 42–47. MR1991628 https://doi.org/10.1002/net.10080

HALE, T. S. and MOBERG, C. R. (2003). Location science research: A review. *Ann. Oper. Res.* **123** 21–35. MR2036877 https://doi.org/10.1023/A:1026110926707

HÄMÄLÄINEN, J., JAUHIAINEN, S. and KÄRKKÄINEN, T. (2017). Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms* (*Basel*) **10** 105. MR3708466 https://doi.org/10.3390/a10030105

HAR-PELED, S. and MAZUMDAR, S. (2004). On coresets for k-means and k-median clustering. In *Proceedings of the* 36*th Annual ACM Symposium on Theory of Computing* 291–300. ACM, New York. MR2121611 https://doi.org/10.1145/1007352.1007400

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

HELLER, K. A., WILLIAMSON, S. and GHAHRAMANI, Z. (2008). Statistical models for partial membership. In *Proceedings of the* 25*th International Conference on Machine Learning* 392–399. ACM, New York.

HENNIG, C. and LIAO, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification (with discussion). *J. R. Stat. Soc. Ser. C. Appl. Stat.* **62** 309–369. MR3060621 https://doi.org/10.1111/j.1467-9876.2012.01066.x

HOLMSTRÖM, L. and PASANEN, L. (2017). Statistical scale space methods. *Int. Stat. Rev.* **85** 1–30. MR3637728 https://doi.org/10.1111/insr.12155

HU, C. W., LI, H. and QUTUB, A. A. (2017). Shrinkage clustering: A fast and size-constrained algorithm for biomedical applications. *BMC Bioinformatics* **19** 19. https://doi.org/10.1186/s12859-018-2022-8

HUH, M.-H. and LIM, Y. B. (2009). Weighting variables in k-means clustering. *J. Appl. Stat.* **36** 67–78. MR2662245 https://doi.org/10.1080/02664760802382533

HUOTARI, T., RUSANEN, J., KEISTINEN, T., LÄHDERANTA, R., SILLANPÄÄ, M. J. and ANTIKAINEN, H. (2020). Effect of centralization on geographic accessibility of maternity hospitals in Finland. *BMC Health Serv. Res.* **20** 337. https://doi.org/10.1186/s12913-020-05222-5

IYIGUN, C. and BEN-ISRAEL, A. (2008). Probabilistic distance clustering adjusted for cluster size. *Probab. Engrg. Inform. Sci.* **22** 603–621. MR2452339 https://doi.org/10.1017/S0269964808000351

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

JAUHIAINEN, S. and KÄRKKÄINEN, T. (2017). A simple cluster validation index with maximal coverage. In *ESANN* 2017 *Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* ESANN. Accepted: 2018-05-21T09:20:36Z.

JIANG, K., KULIS, B. and JORDAN, M. I. (2012). Small-variance asymptotics for exponential family Dirichlet process mixture models. In *Advances in Neural Information Processing Systems* 25 (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 3158–3166. Curran Associates, Red Hook.

JIANG, K., SRA, S. and KULIS, B. (2017). Combinatorial topic models using small-variance asymptotics. In *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.). *Proceedings of Machine Learning Research* **54** 421–429. PMLR.

JIN, X. and HAN, J. (2010). Partitional clustering. In *Encyclopedia of Machine Learning* 766–766. Springer US, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_631

JITTA, A. and KLAMI, A. (2018). On controlling the size of clusters in probabilistic clustering. In *Thirty-Second AAAI Conference on Artificial Intelligence* 3350–3357.

KARATAS, M., RAZI, N. and TOZAN, H. (2016). A comparison of p-median and maximal coverage location models with Q-coverage requirement. *Proc. Eng.* **149** 169–176.

KARATAS, M., RAZI, N. and TOZAN, H. (2017). A multi-criteria assessment of the p-median, maximal coverage and p-center location models. *Teh. Vjesn. — Tech. Gaz.* **24** 399–407.

KAUFMAN, L. and ROUSSEEUW, P. J. (1987). Clustering by Means of Medoids. Delft University of Technology: Reports of the Faculty of Technical Mathematics and Informatics. Faculty of Mathematics and Informatics.

KLAMI, A. and JITTA, A. (2016). Probabilistic size-constrained microclustering. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*. *UAI*'16 329–338. AUAI Press, Arlington, VA, United States.

KLAWONN, F. and HÖPPNER, F. (2003). What is fuzzy about fuzzy clustering? Understanding and improving the concept of the fuzzifier. In *International Symposium on Intelligent Data Analysis* 254–264. Springer, Berlin.

KODINARIYA, T. M. and MAKWANA, P. R. (2013). Review on determining number of cluster in K-means clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **1** 90–95.

KOLLIOPOULOS, S. G. and RAO, S. (1999). A nearly linear-time approximation scheme for the Euclidean k-median problem. In *Algorithms—ESA '99* (*Prague*). *Lecture Notes in Computer Science* **1643** 378–389. Springer, Berlin. MR1729137 https://doi.org/10.1007/3-540-48481-7_33

KULIS, B. and JORDAN, M. I. (2012). Revisiting K-means: New algorithms via Bayesian nonparametrics. In *Proceedings of the* 29*th International Coference on International Conference on Machine Learning*. *ICML*'12 1131–1138. Omnipress, Madison, WI.

LAHDELMA, R., SALMINEN, P. and HOKKANEN, J. (2002). Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European J. Oper. Res.* **142** 345–356. https://doi.org/10.1016/S0377-2217(01)00303-4

LÄHDERANTA, T., LEPPÄNEN, T., RUHA, L., LOVÉN, L., HARJULA, E., YLIANTTILA, M., RIEKKI, J. and SILLANPÄÄ, M. J. (2021). Edge computing server placement with capacitated location allocation. *J. Parallel Distrib. Comput.* **153** 130–149. https://doi.org/10.1016/j.jpdc.2021.03.007

LÄHDERANTA, T., LOVÉN, L., RUHA, L. and KUISMIN, M. (2019). rpack. https://github.com/terolahderanta/rpack.

LÄHDERANTA, T., LOVÉN, L., RUHA, L., LEPPÄNEN, T., LAUNONEN, I., RIEKKI, J. and SILLANPÄÄ, M. J. (2024). Capacitated spatial clustering with multiple constraints and attributes. *Eng. Appl. Artif. Intell.* **127** 107182. https://doi.org/10.1016/j.engappai.2023.107182

LÄHDERANTA, T., LOVÉN, L., RUHA, L., LEPPÄNEN, T., KUISMIN, M., LAUNONEN, I., PIRTTIKANGAS, S., RIEKKI, J. and SILLANPÄÄ, M. J (2026). Supplement to "The -ntersection of location-allocation, partitional clustering and model-nased clustering techniques—a review." https://doi.org/10.1214/25-STS977SUPP

LAPIDOT, I. (2018). Convergence problems of Mahalanobis distance-based k-means clustering. In 2018 *IEEE International Conference on the Science of Electrical Engineering in Israel* (*ICSEE*) 1–5. IEEE.

DASKIN, M. S. and MAASS, K. L. (2015). The p-median problem. In *Location Science* (G. Laporte and F. S. Nickel, eds.) 21–45. Springer, Berlin.

LEE, A., CARON, F., DOUCET, A. and HOLMES, C. (2012). Bayesian sparsity-path-analysis of genetic association signal using generalized t priors. *Stat. Appl. Genet. Mol. Biol.* **11** 5. MR2935747 https://doi.org/10.2202/1544-6115.1712

LEE, I. and LEE, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Bus. Horiz.* **58** 431–440.

LEVINE, Z. H., GERRITS, T., MIGDALL, A. L., SAMAROV, D. V., CALKINS, B., LITA, A. E. and NAM, S. W. (2012). Algorithm for finding clusters with a known distribution and its application to photon-number resolution using a superconducting transition-edge sensor. *J. Opt. Soc. Amer. B Opt. Phys.* **29** 2066–2073.

LI, Z., NIE, F., CHANG, X., MA, Z. and YANG, Y. (2018). Balanced clustering via exclusive lasso: A pragmatic approach. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

LIAO, K. and GUO, D. (2008). A clustering-based approach to the capacitated facility location problem. *Trans. GIS* **12** 323–339. https://doi.org/10.1111/j.1467-9671.2008.01105.x

LIN, C., LIU, K. and CHEN, M. (2005). Dual clustering: Integrating data clustering over optimization and constraint domains. *IEEE Trans. Knowl. Data Eng.* **17** 628–637. https://doi.org/10.1109/TKDE.2005.75

LIU, H., HAN, J., NIE, F. and LI, X. (2017). Balanced clustering with least square regression. In *Thirty-First AAAI Conference on Artificial Intelligence*.

LIU, H., LI, J., WU, Y. and FU, Y. (2021). Clustering with outlier removal. *IEEE Trans. Knowl. Data Eng.* **33** 2369–2379.

LOVÉN, L., LÄHDERANTA, T., RUHA, L., LEPPÄNEN, T., PELTONEN, E., RIEKKI, J. and SILLANPÄÄ, M. J. (2020). Scaling up an edge server deployment. In 2020 *IEEE International Conference on Pervasive Computing and Communications Workshops* (*PerCom Workshops*) 1–7. IEEE, Austin, TX, US.

LOVÉN, L., LÄHDERANTA, T., RUHA, L., PELTONEN, E., LAUNONEN, I., SILLANPÄÄ, M. J., RIEKKI, J. and PIRTTIKANGAS, S. (2021). EDISON: An edge-native method and architecture for distributed interpolation. *Sensors* **21** 1–20.

LOVÉN, L., PELTONEN, E., PANDYA, A., LEPPÄNEN, T., GILMAN, E., PIRTTIKANGAS, S. and RIEKKI, J. (2019). Towards EDISON: An edge-native approach to distributed interpolation of environmental data. In 28*th International Conference on Computer Communications and Networks* (*ICCCN*2019), 1*st Edge of Things Workshop* 2019 (*EoT*2019) IEEE, Valencia, Spain.

MACGREGOR, R. and HODGKINSON, A. T. (2006). *Small Business Clustering Technologies*: *Applications in Marketing*, *Management*, *IT and Economics*: *Applications in Marketing*, *Management*, *IT and Economics*. IGI Global.

MALINEN, M. I. and FRÄNTI, P. (2014). Balanced k-means for clustering. In *Structural*, *Syntactic*, *and Statistical Pattern Recognition* (P. Fränti, G. Brown, M. Loog, F. Escolano and M. Pelillo, eds.) 32–41. Springer Berlin Heidelberg, Berlin, Heidelberg.

MANNING, C. D., RAGHAVAN, P. and SCHÜTZE, H. (2008). *Introduction to Information Retrieval*. Cambridge Univ. Press, New York.

MAROUFY, V. and MARRIOTT, P. (2017). Mixture models: Building a parameter space. *Stat. Comput.* **27** 591–597. MR3613587 https://doi.org/10.1007/s11222-016-9641-6

MARUTHO, D., HENDRA HANDAKA, S., WIJAYA, E. and MULJONO (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 *International Seminar on Application for Technology of Information and Communication* 533–538. https://doi.org/10.1109/ISEMANTIC.2018.8549751

MASSON, M.-H. and DENOEUX, T. (2008). ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognit*. **41** 1384–1397.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

MCNICHOLAS, P. D. (2016). Model-based clustering. *J. Classification* **33** 331–373. MR3575621 https://doi.org/10.1007/s00357-016-9211-9

MELNYKOV, I. and MELNYKOV, V. (2014). On K-means algorithm with the use of Mahalanobis distances. *Statist*. *Probab. Lett*. **84** 88–95. MR3131260 https://doi.org/10.1016/j.spl.2013.09.026

MILLER, J. W. and HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist*. *Assoc*. **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

MILLIGAN, G. W. and COOPER, M. C. (1988). A study of standardization of variables in cluster analysis. *J. Classification* **5** 181–204. MR0971152 https://doi.org/10.1007/BF01897163

MLADENOVIĆ, N., BRIMBERG, J., HANSEN, P. and MORENO-PÉREZ, J. A. (2007). The p-median problem: A survey of metaheuristic approaches. *European J. Oper. Res*. **179** 927–939. MR2286521 https://doi.org/10.1016/j.ejor.2005.05.034

MULVEY, J. M. and BECK, M. P. (1984). Solving capacitated clustering problems. *European J. Oper. Res*. **18** 339–348. https://doi.org/10.1016/0377-2217(84)90155-3

MURRAY, A. T. (2018). Evolving location analytics for service coverage modeling. *Geogr. Anal*. **50** 207–222. https://doi.org/10.1111/gean.12146

MURRAY, A. T. and ESTIVILL-CASTRO, V. (1998). Cluster discovery techniques for exploratory spatial data analysis. *Int*. *J*. *Geogr. Inf. Sci*. **12** 431–443. https://doi.org/10.1080/136588198241734

N'CIR, C.-E. B., CLEUZIOU, G. and ESSOUSSI, N. (2015). Overview of overlapping partitional clustering methods. In *Partitional Clustering Algorithms* 245–275. Springer, Berlin.

NAIK, D. N. and PLUNGPONGPUN, K. (2006). A Kotz-type distribution for multivariate statistical inference. In *Advances in Distribution Theory*, *Order Statistics*, *and Inference*. *Stat*. *Ind. Technol*. 111–124. Birkhäuser, Boston, MA. MR2226218 https://doi.org/10.1007/0-8176-4487-3_7

NAINGGOLAN, R., PERANGIN-ANGIN, R., SIMARMATA, E. and TARIGAN, A. F. (2019). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the elbow method. *J. Phys*., *Conf. Ser.* **1361** 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

NANDA, S. J. and PANDA, G. (2014). A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm Evol*. *Comput*. **16** 1–18. https://doi.org/10.1016/j.swevo.2013.11.003

NEGREIROS, M. and PALHANO, A. (2006). The capacitated centred clustering problem. *Comput*. *Oper. Res*. **33** 1639–1663.

NG, S. K., KRISHNAN, T. and MCLACHLAN, G. J. (2012). The EM algorithm. In *Handbook of Computational Statistics—Concepts and Methods*. 1, 2. *Springer Handb*. *Comput*. *Stat*. 139–172. Springer, Heidelberg. MR2985398 https://doi.org/10.1007/978-3-642-21551-3_6

OLUKANMI, P. O. and TWALA, B. (2017). K-means-sharp: Modified centroid update for outlier-robust k-means clustering. In 2017 *Pattern Recognition Association of South Africa and Robotics and Mechatronics* (*PRASA-RobMech*) 14–19. IEEE.

RAHMAN, S. and SMITH, D. K. (2000). Use of location-allocation models in health service development planning in developing nations. *European J. Oper. Res*. **123** 437–452. https://doi.org/10.1016/S0377-2217(99)00289-1

PANIAN, Z. (2012). A new dimension of business intelligence: Location-based intelligence. *World Academy of Science*, *Engineering and Technology*, *International Journal of Social*, *Behavioral*, *Educational*, *Economic*, *Business and Industrial Engineering* **6** 338–343.

PEEL, D. and MCLACHLAN, G. J. (2000). Robust mixture modelling using the t distribution. *Stat*. *Comput*. **10** 339–348.

PHAM, D. T., DIMOV, S. S. and NGUYEN, C. D. (2005). Selection of K in K-means clustering. *Proc*. *Inst*. *Mech*. *Eng*., *Part C*, *J*. *Mech*. *Eng*. *Sci*. **219** 103–119.

RAYKOV, Y. P., BOUKOUVALAS, A., BAIG, F. and LITTLE, M. A. (2016). What to do when k-means clustering fails: A simple yet principled alternative algorithm. *PLoS ONE* **11** 1–28. https://doi.org/10.1371/journal.pone.0162259

RAYMAEKERS, J. and ZAMAR, R. H. (2020). Pooled variable scaling for cluster analysis. *Bioinformatics* **36** 3849–3855. https://doi.org/10.1093/bioinformatics/btaa243

REVELLE, C. S., EISELT, H. A. and DASKIN, M. S. (2008). A bibliography for some fundamental problem categories in discrete location science. *European J. Oper. Res*. **184** 817–848. MR2356873 https://doi.org/10.1016/j.ejor.2006.12.044

RHODES, K. M., TURNER, R. M., WHITE, I. R., JACKSON, D., SPIEGELHALTER, D. J. and HIGGINS, J. P. T. (2016). Implementing information priors for heterogeneity in meta-analysis using meta-regression and pseudo data. *Stat*. *Med*. **35** 5495–5511. MR3573072 https://doi.org/10.1002/sim.7090

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

RITTER, G. (2015). *Robust Cluster Analysis and Variable Selection*. *Monographs on Statistics and Applied Probability* **137**. CRC Press, Boca Raton, FL. MR3241693

ROCHE, A. (2011). EM algorithm and variants: An informal tutorial. https://doi.org/10.48550/ARXIV.1105.1476

ROHANI, A. M. B. M. and CHUA, F.-F. (2018). Location analytics for optimal business retail site selection. In *Computational Science and Its Applications – ICCSA* 2018 (O. Gervasi, B. Murgante, S. Misra, E. Stankova, C. M. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, E. Tarantino and Y. Ryu, eds.) 392–405. Springer, Cham.

SARALA, O., PYHÄJÄRVI, T. and SILLANPÄÄ, M. J. (2023). BELMM: Bayesian model selection and random walk smoothing in time-series clustering. *Bioinformatics* **39** btad686. https://doi.org/10.1093/bioinformatics/btad686

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

SETIADI, T., PRANOLO, A., AZIZ, M., MARDIYANTO, S., HENDRAJAYA, B. and MUNIR (2017). A model of geographic information system using graph clustering methods. In 2017 3*rd International Conference on Science in Information Technology* (*ICSITech*) 727–731. https://doi.org/10.1109/ICSITech.2017.8257208

SHI, W., CAO, J., ZHANG, Q., LI, Y. and XU, L. (2016). Edge computing: Vision and challenges. *IEEE Internet Things J*. **3** 637–646.

STEINBACH, M., KUMAR, V. and TAN, P. (2019). Cluster analysis: Basic concepts and algorithms. In *Introduction to Data Mining*, 2nd ed. Addison-Wesley, Reading.

STEINLEY, D. (2004). Standardizing variables in K-means clustering. In *Classification*, *Clustering*, *and Data Mining Applications*

(D. Banks, F. R. McMorris, P. Arabie and W. Gaul, eds.) 53–60. Springer Berlin Heidelberg, Berlin, Heidelberg.

STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

SUNG, K.-K. and POGGIO, T. (1998). Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 39–51.

KRISHNAPURAM, R. and KELLER, J. M. (1993). A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1** 98–110.

TEAM, S. D. (2022). RStan: The R interface to Stan. R package version 2.21.5.

THEODORIDIS, S. and KOUTROUMBAS, K. (2008). *Pattern Recognition, Fourth Edition*, 4th ed. Academic Press, Orlando, FL.

TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293

TOREGAS, C. and REVELLE, C. (1973). Binary logic solutions to a class of location problem. *Geogr. Anal.* **5** 145–155. https://doi.org/10.1111/j.1538-4632.1973.tb01004.x

TSENG, G. C. (2007). Penalized and weighted K-means for clustering with scattered objects and prior information in high-throughput biological data. *Bioinformatics* **23** 2247–2255. https://doi.org/10.1093/bioinformatics/btm320

TSENG, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** 475–494. MR1835069 https://doi.org/10.1023/A:1017501703105

TZENG, G.-H., TENG, M.-H., CHEN, J.-J. and OPRICOVIC, S. (2002). Multicriteria selection for a restaurant location in Taipei. *Int. J. Contemp. Hosp. Manag.* **21** 171–187.

VESANTO, J. (2001). Importance of individual variables in the k-means algorithm. In *Advances in Knowledge Discovery and Data Mining* (D. Cheung, G. J. Williams and Q. Li, eds.). *Lecture Notes in Computer Science* 513–518. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45357-1_54

WADE, S. (2023). Bayesian cluster analysis. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **381** Paper No. 20220149, 20. MR4590507 https://doi.org/10.1098/rsta.2022.0149

WANG, S., GUO, Y., ZHANG, N., YANG, P., ZHOU, A. and SHEN, X. S. (2021). Delay-aware microservice coordination in mobile edge computing: A reinforcement learning approach. *IEEE Trans. Mob. Comput.* **20** 939–951.

WHANG, J. J., DHILLON, I. S. and GLEICH, D. F. (2015). Non-exhaustive, overlapping k-means. In *Proceedings of the* 2015 *SIAM International Conference on Data Mining* 936–944. SIAM, Philadelphia.

WHEELER, A. P. (2019). Creating optimal patrol areas using the P-median model. *Policing* **42** 318–333.

WINARSO, K. and ROHIM, M. I. A. (2019). Comparation of p-median, p-center, and maximal coverage on facility location problem of bokabo tobacco supply chain, sumenep district. In 1*st International Conference on Life, Innovation, Change and Knowledge* (*ICLICK* 2018) Atlantis Press.

WU, D. (2022). em: Generic EM algorithm. R package version 1.0.0.

XIANG, S., NIE, F. and ZHANG, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognit.* **41** 3600–3612.

XIAO, Y. and YU, J. (2012). Partitive clustering (K-means family). *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2** 209–225.

XIONG, H., WU, J. and CHEN, J. (2008). K-means clustering versus validation measures: A data-distribution perspective. *IEEE Trans. Syst. Man Cybern., Part B, Cybern.* **39** 318–331.

XU, J., WANG, S., BHARGAVA, B. K. and YANG, F. (2019). A blockchain-enabled trustless crowd-intelligence ecosystem on mobile edge computing. *IEEE Trans. Ind. Inform.* **15** 3538–3547.

XU, L., COLLIER, R. and O'HARE, G. M. P. (2017). A survey of clustering techniques in WSNs and consideration of the challenges of applying such to 5G IoT scenarios. *IEEE Internet Things J.* **4** 1229–1249.

XU, S., QIAO, X., ZHU, L., ZHANG, Y., XUE, C. and LI, L. (2016). Reviews on determining the number of clusters. *Appl. Math. Inf. Sci.* **10** 1493–1512.

YI, S., LI, C. and LI, Q. (2015). A survey of fog computing: Concepts, applications and issues. In *Proceedings of the* 2015 *Workshop on Mobile Big Data* 37–42.

YUAN, C. and YANG, H. (2019). Research on k-value selection method of k-means clustering algorithm. *J. Multidisciplinary Sci. J.* **2** 226–235.

ZHAO, Q. and FRÄNTI, P. (2014). WB-index: A sum-of-squares based index for cluster validity. *Data Knowl. Eng.* **92** 77–89. https://doi.org/10.1016/j.datak.2014.07.008

ZHAO, W.-L., DENG, C.-H. and NGO, C.-W. (2018). k-means: A revisit. *Neurocomputing* **291** 195–206. https://doi.org/10.1016/j.neucom.2018.02.072

ZHOU, K. and YANG, S. (2020). Effect of cluster size distribution on clustering: A comparative study of k-means and fuzzy c-means clustering. *PAA Pattern Anal. Appl.* **23** 455–466. MR4061621 https://doi.org/10.1007/s10044-019-00783-6

ZHOU, W. and LI, Z. (2013). The multi-covering emergency service facility location problem with considering disaster losses. In 11*th International Symposium on Operations Research and Its Applications in Engineering, Technology and Management* 2013 (*ISORA* 2013) 1–6.

ZHU, J., ZHENG, J., DI, S., WANG, S. and YANG, J. (2020). A dual spatial clustering method in the presence of heterogeneity and noise. *Trans. GIS* **24** 1799–1826. https://doi.org/10.1111/tgis.12687

ZHU, S., WANG, D. and LI, T. (2010). Data clustering with size constraints. *Knowl.-Based Syst.* **23** 883–889. https://doi.org/10.1016/j.knosys.2010.06.003

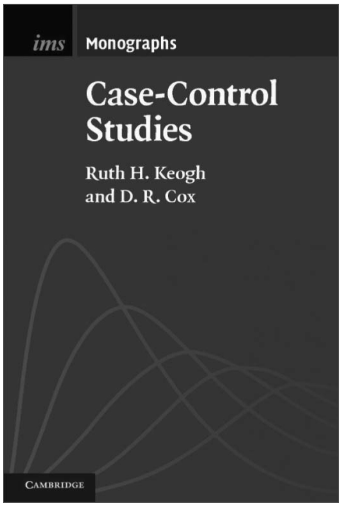
INSTITUTE OF MATHEMATICAL STATISTICS

(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.

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Case-Control Studies

Ruth H. Keogh
and D. R. Cox

The case-control approach is a powerful method for investigating factors that may explain a particular event. It is extensively used in epidemiology to study disease incidence, one of the best-known examples being Bradford Hill and Doll's investigation of the possible connection between cigarette smoking and lung cancer. More recently, case-control studies have been increasingly used in other fields, including sociology and econometrics.

With a particular focus on statistical analysis, this book is ideal for applied and theoretical statisticians wanting an up-to-date introduction to the field. It covers the fundamentals of case-control study design and analysis as well as more recent developments, including two-stage studies, case-only studies and methods for case-control sampling in time. The latter have important applications in large prospective cohorts which require case-control sampling designs to make efficient use of resources. More theoretical background is provided in an appendix for those new to the field.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

Hardback price US$48.00 (non-member price $80.00)

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.