STATISTICAL SCIENCE

Volume 41, Number 1 February 2026

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 41, Number 1, February 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 1 (1–226) February 2026**

Volume 41 **Number 1** **February 2026**

Statistical Science
2026, Vol. 41, No. 1, 1–27
https://doi.org/10.1214/24-STS933

Modern Statistical Models and Methods for Estimating Fatigue-Life and Fatigue-Strength Distributions from Experimental Data

William Q. Meeker, Luis A. Escobar, Francis G. Pascual, Yili Hong, Peng Liu, Wayne M. Falk and Balajee Ananthasayanam

Abstract. Engineers and scientists have been collecting and analyzing fatigue data since the 1800s to ensure the reliability of life-critical structures. Applications include (but are not limited to) bridges, building structures, aircraft and spacecraft components, ships, ground-based vehicles, and medical devices. Engineers need to estimate *S-N* relationships (Stress versus Number of cycles to failure), typically with a focus on estimating small quantiles of the *fatigue-life* distribution. Estimates from this kind of model are used as input to models (e.g., cumulative damage models) that predict failure-time distributions under varying stress patterns. Also, design engineers need to estimate lower-tail quantiles of the closely related *fatigue-strength* distribution. The history of applying incorrect statistical methods is nearly as long and such practices continue to the present. Examples include treating the applied stress (or strain) as the response and the number of cycles to failure as the explanatory variable in regression analyses (because of the need to estimate fatigue-strength distributions) and ignoring or otherwise mishandling censored observations (known as runouts in the fatigue literature). The first part of the paper reviews the traditional modeling approach where a fatigue-life model is specified. Then we show how this specification induces a corresponding fatigue-strength model. The second part of the paper presents a novel alternative modeling approach where a fatigue-strength model is specified and a corresponding fatigue-life model is induced. We explain and illustrate the important advantages of this new modeling approach.

Key words and phrases: Bayesian inference, censored data, failure-time regression, fracture, maximum likelihood, nonlinear regression, reliability, *S-N* curves.

REFERENCES

ASTM (2015). *E739-10 Standard Practice for Statistical Analysis of Linear or Linearized Stress-Life* (*S-N*) *and Strain-Life* (ϵ*-N*) *Fatigue Data*. ASTM International, West Conshohocken, PA.

ASTM (2017). *F3211-17 Standard Guide for Fatigue-to-Fracture* (*FtF*) *Methodology for Cardiovascular Medical Devices*. ASTM International, West Conshohocken, PA.

AWAD, M., DEJACK, M. and KRIVTSOV, V. (2004). Evaluation of fatigue life regression models Technical report, SAE Technical Paper 2004-01-0625.

BABUŠKA, I., SAWLAN, Z., SCAVINO, M., SZABÓ, B. and TEMPONE, R. (2016). Bayesian inference and model comparison for metallic fatigue data. *Comput. Methods Appl. Mech. Engrg.* **304** 171–196.

William Q. Meeker is Distinguished Professor Emeritus of Statistics, Department of Statistics, Iowa State University, Ames, Iowa 50011, USA (e-mail: wqmeeker@iastate.edu). Luis A. Escobar is Professor, Department of Experimental Statistics, Louisiana State University, Baton Rouge, Louisiana 70803, USA (e-mail: luis@lsu.edu). Francis G. Pascual is Emeritus Associate Professor, Department of Mathematics and Statistics, Washington State University, Pullman, WA 99164, USA (e-mail: jave@wsu.edu). Yili Hong is Professor, Department of Statistics, Virginia Tech, Blacksburg, Virginia 24061, USA (e-mail: yilihong@vt.edu). Peng Liu is Principal Research Statistician Developer, JMP Statistical Discovery LLC, Cary, North Carolina 27513, USA (e-mail: Peng.Liu@jmp.com). Wayne M. Falk is Staff Senior Fellow, U.S. Food and Drug Administration, Silver Spring, Maryland 20993, USA (e-mail: wayne.falk@fda.hhs.gov). Balajee Ananthasayanam is Principal Engineer, Honda Aero, Burlington, North Carolina 27215, USA (e-mail: bananthasayanam@na.honda.com).

BASQUIN, O. H. (1910). The exponential law of endurance tests. *Proc. - Am. Soc. Test. Mater.* **10** 625–630.

BASTENAIRE, F., BASTIEN, M. and POMEY, G. (1961). Statistical analysis of fatigue test results and its application to new series of data. *Acta Tech. Acad. Sci. Hung.* **35–36** 7–26.

BASTENAIRE, F. A. (1972). New method for the statistical evaluation of constant stress amplitude fatigue-test results. In *Probabilistic Aspects of Fatigue* American Society for Testing and Materials International, West Conshohocken, PA.

BATHIAS, C. (1999). There is no infinite fatigue life in metallic materials. *Fatigue Fract. Eng. Mater. Struct.* **22** 559–565.

CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression. Monographs on Statistics and Applied Probability*. CRC Press, New York.

CASTILLO, E. and FERNÁNDEZ-CANTELI, A. (2009). *A Unified Statistical Methodology for Modeling Fatigue Damage*. Springer, Berlin.

CASTILLO, E., FERNÁNDEZ-CANTELI, A., ESSLINGER, V. and THÜRLIMANN, B. (1985). Statistical model for fatigue analysis of wires, strands and cables. In *IABSE Proceedings* 1–40.

CASTILLO, E., FERNÁNDEZ-CANTELI, A., PINTO, H. and LÓPEZ-AENLLE, M. (2008). A general regression model for statistical analysis of strain–life fatigue data. *Mater. Lett.* **62** 3639–3642.

CASTILLO, E. and GALAMBOS, J. (1987). Lifetime regression models based on a functional equation of physical nature. *J. Appl. Probab.* **24** 160–169.

CASTILLO, E., MUNIZ-CALVENTE, M., FERNÁNDEZ-CANTELI, A. and BLASÓN, S. (2019). Fatigue assessment strategy using Bayesian techniques. *Materials* **12** 3239.

CROWDER, M. J., KIMBER, A. C., SMITH, R. L. and SWEETING, T. J. (1994). *Statistical Analysis of Reliability Data*, Paperback ed. CRC Press, London.

DOWLING, N. (2013). *Mechanical Behavior of Materials: Engineering Methods for Deformation, Fracture, and Fatigue*, 4th ed. Pearson Education, Upper Saddle River.

EASA (2018). AMC E 510 European Union Aviation Safety Agency Certification Specifications and Acceptable Means of Compliance for Engines CS-E. Available at https://www.easa.europa.eu/en/downloads/67945/en. Accessed: 12 March 2024.

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

FAA (2009a). U.S. Federal Aviation Administration Advisory Circular AC 33.70-1: Guidance Material for Aircraft Engine Life-Limited Parts Requirements. Available at https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_33_70-1_Chg_1.pdf. Accessed: 12 March 2024.

FAA (2009b). U.S. Federal Aviation Administration Advisory Circular AC 33.70-2: Damage Tolerance of Hole Features in High-Energy Turbine Engine Rotors. Available at http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_33_70-2.pdf. Accessed: 12 March 2024.

FALK, W. (2019). A statistically rigorous fatigue strength analysis approach applied to medical devices. In *Fourth Symposium on Fatigue and Fracture of Metallic Medical Materials and Devices*. ASTM International, West Conshohocken, PA.

FDA (2010). U.S. Food and Drug Administration Non-clinical engineering tests and recommended labeling for intravascular stents and associated delivery systems. Available at https://www.fda.gov/media/71639/download. Accessed: 12 March 2024.

FREUDENTHAL, A. M. and GUMBEL, E. J. (1953). On the statistical interpretation of fatigue tests. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **216** 309–332.

FREUDENTHAL, A. M. and GUMBEL, E. J. (1954). Minimum life in fatigue. *J. Amer. Statist. Assoc.* **49** 575–597.

FREUDENTHAL, A. M. and GUMBEL, E. J. (1956). Physical and statistical aspects of fatigue. *Adv. Appl. Mech.* **4** 117–158.

GELMAN, A., SIMPSON, D. and BETANCOURT, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy* **19** 555.

GNEDENKO, B. V., BELYAYEV, Y. K. and SOLOVYEV, A. D. (1969). *Mathematical Methods of Reliability Theory. Probability and Mathematical Statistics, Vol.* 6. Academic Press, New York. Translated by Scripta Technica, Inc, Translation edited by Richard E. Barlow.

GROVE, D. and CAMPEAN, F. (2008). A comparison of two methods of analysing staircase fatigue test data. *Qual. Reliab. Eng. Int.* **24** 485–497.

HADDAD, T., HIMES, A. and CAMPBELL, M. (2014). Fracture prediction of cardiac lead medical devices using Bayesian networks. *Reliab. Eng. Syst. Saf.* **123** 145–157.

HANAKI, S., IWAO, Y., YAMASHITA, M., UCHIDA, H., ZAKO, M. and KURASHIKI, T. (2003). On a decision method of the S–N curve based on fatigue strength distribution. *J. Soc. Material. Sci. Jpn.* **52** 23–27.

HANAKI, S., YAMASHITA, M., UCHIDA, H. and ZAKO, M. (2010). On stochastic evaluation of S–N data based on fatigue strength distribution. *Int. J. Fatigue* **32** 605–609.

HAUTEVILLE, R., HERMITE, X. and LEFEBVRE, F. (2022). A new generic method to analyse fatigue results. *Procedia Struct. Integr.* **38** 507–518.

HONG, Y., MEEKER, W. Q. and ESCOBAR, L. A. (2008). The relationship between confidence intervals for failure probabilities and life time quantiles. *IEEE Trans. Reliab.* **57** 260–266.

ISO (2012). *ISO* 12107: *Metallic Materials-Fatigue Testing-Statistical Planning and Analysis of Data*, 2nd ed. ISO, Geneva, Switzerland.

ISO (2017). *ISO* 5840 *Cardiovascular Implants—Cardiac Valve Prostheses*. International Organization for Standardization, Geneva, Switzerland.

JOHNSON, V. E., FITZGERALD, M. and MARTZ, H. F. (1999). Estimating fatigue curves with the random fatigue-limit model: Discussion. *Technometrics* **41** 294–296.

KING, C. B., HONG, Y., DEHART, S. P., DEFEO, P. A. and PAN, R. (2016). Planning fatigue tests for polymer composites. *J. Qual. Technol.* **48** 227–245.

KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge.

LITTLE, R. E. and JEBE, E. H. (1975). *Statistical Design of Fatigue Experiments*. Halsted, New York.

LIU, P., HONG, Y., ESCOBAR, L. A. and MEEKER, W. Q. (2024). On equivalence of likelihood-based confidence bands for fatigue-life and fatigue-strength distributions. Available at https://arxiv.org/abs/2403.12757. Accessed: 20 March 2024.

LIU, P. and MEEKER, W. Q. (2024). Robust numerical methods for nonlinear regression. Available at https://arxiv.org/abs/2403.12759. Accessed: 20 March 2024.

MANN, N. R., SCHAFER, R. E. and SINGPURWALLA, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York.

MATTINGLY, J. D., HEISER, W. H. and PRATT, D. T. (2002). *Aircraft Engine Design*. AIAA, Washington.

MEEKER, W. Q. and ESCOBAR, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *Amer. Statist.* **49** 48–53.

MEEKER, W. Q., ESCOBAR, L. A. and PASCUAL, F. G. (2022). *Statistical Methods for Reliability Data*, 2nd ed. Wiley, New York.

MEEKER, W. Q, ESCOBAR, L. A, PASCUAL, F. G, HONG, Y., LIU, P., FALK, W. M and ANANTHASAYANAM, B. (2026). Supplement to "Modern statistical models and methods for estimating fatigue-life and fatigue-strength distributions from experimental data." https://doi.org/10.1214/24-STS933SUPP

MEEKER, W. Q., ESCOBAR, L. A. and ZAYAC, S. A. (2003). Use of sensitivity analysis to assess the effect of model uncertainty in analyzing accelerated life test data. In *Case Studies in Reliability and Maintenance*, *Chapter* 12 (W. R. Blischke and D. N. P. Murthy, eds.). *Wiley Ser. Probab. Stat*. 269–292. Wiley-Interscience, Hoboken, NJ.

MEEKER, W. Q., HAHN, G. J. and ESCOBAR, L. A. (2017). *Statistical Intervals*: *A Guide for Practitioners and Researchers*. Wiley, New York.

MMPDS (2021). *Metallic Materials Properties Development and Standardization* (*MMPDS*-16): *Chapters* 1-9 **2**. Battelle Memorial Institute, Columbus, OH.

MÜLLER, C., WÄCHTER, M., MASENDORF, R. and ESDERTS, A. (2017). Accuracy of fatigue limits estimated by the staircase method using different evaluation techniques. *Int*. *J*. *Fatigue* **100** 296–307.

NELSON, W. B. (1973). Analysis of residuals from censored data. *Technometrics* **15** 697–715.

NELSON, W. B. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J*. *Test*. *Eval*. **12** 69–77.

NELSON, W. B. (2004). *Accelerated Testing*: *Statistical Models*, *Test Plans*, *and Data Analyses*, Paperback ed. Wiley, New York.

NISHIJIMA, S. (1980). Statistical analysis of small sample fatigue data. *Trans*. *Jpn*. *Soc*. *Mech*. *Eng*. *Ser*. *A* **46** 1303–1313.

NISHIJIMA, S. (1985). Statistical analysis of small sample fatigue data. *Trans*. *Natl*. *Res*. *Inst*. *Met*. **27** 234–245.

PASCUAL, F. G. (2003). The random fatigue-limit model in multifactor experiments. *J*. *Stat*. *Comput*. *Simul*. **73** 733–752.

PASCUAL, F. G. and MEEKER, W. Q. (1997). Regression analysis of fatigue data with runouts based on a model with nonconstant standard deviation and a fatigue limit parameter. *J*. *Test*. *Eval*. **25** 292–301.

PASCUAL, F. G. and MEEKER, W. Q. (1999). Estimating fatigue curves with the random fatigue-limit model (with discussion). *Technometrics* **41** 277–302.

PAWITAN, Y. (2013). *All Likelihood*: *Statistical Modelling and Inference Using Likelihood*, Paperback ed. Oxford Univ. Press, London.

POLLAK, R., PALAZOTTO, A. and NICHOLAS, T. (2006). A simulation-based investigation of the staircase method for fatigue strength testing. *Mech*. *Mater.* **38** 1170–1181.

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

SEVERINI, T. A. (2000). *Likelihood Methods in Statistics*. *Oxford Statistical Science Series* **22**. Oxford Univ. Press, Oxford.

SHANMUGAM, V., ZHAO, T., KRAMS, W., JOSHI, A., CAO, H. and SCHMIDT, P. (2019). Fatigue reliability analysis framework for medical devices based on a probabilistic finite element approach. In *Fourth Symposium on Fatigue and Fracture of Metallic Medical Materials and Devices* 132–147. ASTM International, West Conshohocken, PA.

SHIMOKAWA, T. and HAMAGUCHI, Y. (1987). Statistical evaluation of fatigue life and fatigue strength in circular-hole notched specimens of a carbon eight-harness-satin/epoxy laminate. In *Statistical Research on Fatigue and Fracture* (T. Tanaka, S. Nishijima and M. Ichikawa, eds.) 159–176. Elsevier, Amsterdam.

SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90.

SPINDEL, J. and HAIBACH, E. (1979). The method of maximum likelihood applied to the statistical analysis of fatigue data *Int*. *J*. *Fatigue* **1** 81–88.

STAN DEVELOPMENT TEAM (2022a). RStan: The R Interface to Stan. R Package Version 2.30.

STAN DEVELOPMENT TEAM (2022b). Stan User's Guide, Version 2.30.

STROMEYER, C. (1914). The determination of fatigue limits under alternating stress conditions. *Proc*. *R*. *Soc*. *Lond*. *Ser*. *A*, *Contain*. *Pap*. *Math*. *Phys*. *Character* **90** 411–425.

TIAN, Q., LEWIS-BECK, C., NIEMI, J. B. and MEEKER, W. Q. (2024). Specifying prior distributions in reliability applications. *Appl*. *Stoch*. *Models Bus*. *Ind*. **40** 5–62.

TOASA CAIZA, P. D. and UMMENHOFER, T. (2018). A probabilistic Stüssi function for modelling the SN curves and its application on specimens made of steel S355J2 $+$ N. *Int*. *J*. *Fatigue* **117** 121–134.

WEAVER, J., SENA, G., AYCOCK, K., ROIKO, A., FALK, W., SIVAN, S. and BERG, B. (2023). Rotary bend fatigue of nitinol to one billion cycles. *Shape Mem*. *Superelast*. **9** 50–73.

WEIBULL, W. (1956). Scatter of fatigue life and fatigue strength in aircraft structural materials and parts. In *Fatigue in Aircraft Structures* (A. Freudenthal, ed.) 126–145. Academic Press, San Diego.

WOO, S. (2020). *Reliability Design of Mechanical Systems*, 2nd ed. Springer, Berlin.

WU, C. F. J. and TIAN, Y. (2014). Three-phase optimal design of sensitivity experiments. *J*. *Statist*. *Plann*. *Inference* **149** 1–15.

Statistical Science
2026, Vol. 41, No. 1, 28–48
https://doi.org/10.1214/24-STS939

Analysis of Linked Files: A Missing Data Perspective

Gauri Kamat and Roee Gutman

Abstract. In many applications, researchers seek to identify overlapping entities across multiple data files. Record linkage algorithms facilitate this task, in the absence of unique identifiers. As these algorithms rely on semi-identifying information, they may miss records that represent the same entity, or incorrectly link records that do not represent the same entity. Analysis of linked files commonly ignores such linkage errors, resulting in biased, or overly precise estimates of the associations of interest. We view record linkage as a missing data problem, and delineate the linkage mechanisms that underpin analysis methods with linked files. Following the missing data literature, we group these methods under three broad categories: likelihood and Bayesian methods, imputation methods, and weighting methods. We summarize the assumptions and limitations of the methods, and evaluate their performance in a wide range of simulation scenarios.

Key words and phrases: Record linkage, missing data, Bayesian, imputation, likelihood.

REFERENCES

ABOWD, J. M., ABRAMOWITZ, J., LEVENSTEIN, M. C., MCCUE, K., PATKI, D., RAGHUNATHAN, T. E., RODGERS, A. M., SHAPIRO, M. D., WASI, N. et al. (2021). Finding needles in haystacks: Multiple-imputation record linkage using machine learning. Working Paper No. 21-35 Center for Economic Studies, U.S. Census Bureau.

ALESHIN-GUENDEL, S. and SADINLE, M. (2023). Multifile partitioning for record linkage and duplicate detection. *J. Amer. Statist. Assoc.* **118** 1786–1795. https://doi.org/10.1080/01621459.2021.2013242

ASHER, J., RESNICK, D., BRITE, J., BRACKBILL, R. and CONE, J. (2020). An introduction to probabilistic record linkage with a focus on linkage processing for WTC registries. *Int. J. Environ. Res. Public Health* **17**. https://doi.org/10.3390/ijerph17186937

BAXTER, R. and CHRISTEN, P. (2003). A comparison of fast blocking methods for record linkage, CMIS technical report 03/139. In *Proceedings of ACM SIGKDD'03 Workshop on Data Cleaning, Record Linkage, and Object Consolidation* 39–48.

BELIN, T. R. and RUBIN, D. B. (1995). A method for calibrating false-match rates in record linkage. *J. Amer. Statist. Assoc.* **90** 694–707.

BILENKO, M., KAMATH, B. and MOONEY, R. (2006). Adaptive blocking: Learning to scale up record linkage. In *Proceedings of the Sixth IEEE International Conference on Data Mining* 87–96.

BILENKO, M. and MOONEY, R. J. (2003). On evaluation and training-set construction for duplicate detection. In *Proceedings of the KDD-2003 Workshop on Data Cleaning, Record Linkage, and Object Consolidation* 7–12.

BINETTE, O. and STEORTS, R. C. (2022). (Almost) all of entity resolution. *Sci. Adv.* **8** eabi8021. https://doi.org/10.1126/sciadv.abi8021

BIRD, S. M. and KING, R. (2018). Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annu. Rev. Stat. Appl.* **5** 95–118. https://doi.org/10.1146/annurev-statistics-031017-100641

BOHENSKY, M. (2015). Bias in data linkage studies. In *Methodological Developments in Data Linkage* 63–82 4. Wiley, New York.

BRENNER, H., SCHMIDTMANN, I. and STEGMAIER, C. (1997). Effects of record linkage errors on registry-based follow-up studies. *Stat. Med.* **16** 2633–2643. https://doi.org/10.1002/(sici)1097-0258(19971215)16:23<2633::aid-sim702>3.0.co;2-1

BRISCOLINI, D., DI CONSIGLIO, L., LISEO, B., TANCREDI, A. and TUOTO, T. (2018). New methods for small area estimation with linkage uncertainty. *Internat. J. Approx. Reason.* **94** 30–42.

CAMPBELL, K. M., DECK, D. and KRUPSKI, A. (2008). Record linkage software in the public domain: A comparison of link plus, the link king, and a 'basic' deterministic algorithm. *Health Inform. J.* **14** 5–15.

CAMPBELL, S. R., RESNICK, D. M., COX, C. S. and MIREL, L. B. (2021). Using supervised machine learning to identify efficient blocking schemes for record linkage. *Stat. J. IAOS* **37** 673–680. https://doi.org/10.3233/sji-200779

CANGUL, M. Z., CHRETIEN, Y. R., GUTMAN, R. and RUBIN, D. B. (2009). Testing treatment effects in unconfounded studies under model misspecification: Logistic regression, discretization, and their combination. *Stat. Med.* **28** 2531–2551. https://doi.org/10.1002/sim.3633

CHAMBERS, R. (2009). Regression analysis of probability-linked data. *Statisphere Off. Stat.* **4**.

Gauri Kamat is Ph.D. candidate, Department of Biostatistics, Brown University, Providence, Rhode Island 02903, USA (e-mail: gaurik1.edu@gmail.com). Roee Gutman is Associate Professor, Department of Biostatistics, Brown University, Providence, Rhode Island 02903, USA (e-mail: rgutman@stat.brown.edu).

CHAMBERS, R. and DINIZ DA SILVA, A. (2020). Improved secondary analysis of linked data: A framework and an illustration. *J. R. Stat. Soc., Ser. A* **183** 37–59.
CHAMBERS, R., SALVATI, N., FABRIZI, E. and DINIZ DA SILVA, A. (2019). Domain estimation under informative linkage. *Stat. Theory Relat. Fields* **3** 90–102.
CHAMBERS, R. L., FABRIZI, E., RANALLI, M. G., SALVATI, N. and WANG, S. (2022). Robust regression using probabilistically linked data. *Wiley Interdiscip. Rev.: Comput. Stat.* e1596.
CHIPPERFIELD, J. (2019). A weighting approach to making inference with probabilistically linked data. *Stat. Neerl.* **73** 333–350.
CHIPPERFIELD, J. O. and CHAMBERS, R. L. (2015). Using the bootstrap to account for linkage errors when analysing probabilistically linked categorical data. *J. Off. Stat.* **31** 397.
CHRISTEN, P. (2007). A two-step classification approach to unsupervised record linkage. In *Proceedings of the Sixth Australasian Conference on Data Mining and Analytics* 111–119.
CHRISTEN, P. (2008a). Automatic record linkage using seeded nearest neighbour and support vector machine classification. In *Proceedings of the* 14*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 151–159.
CHRISTEN, P. (2008b). Automatic training example selection for scalable unsupervised record linkage. In *Advances in Knowledge Discovery and Data Mining*: 12*th Pacific-Asia Conference*, *PAKDD* (T. Washio, E. Suzuki, K. M. Ting and A. Inokuchi, eds.) 511–518.
CHRISTEN, P. and GOISER, K. (2005). Assessing deduplication and data linkage quality: What to measure? In *Proceedings of the Fourth Australasian Data Mining Conference*.
CHRISTEN, P. and GOISER, K. (2007). Quality and complexity measures for data linkage and deduplication. In *Quality Measures in Data Mining* 127–151. Springer, Berlin.
CHRISTEN, P., RANBADUGE, T. and SCHNELL, R. (2020). *Linking Sensitive Data*: *Methods and Techniques for Practical Privacy-Preserving Information Sharing*. Springer, Cham.
COCHINWALA, M., KURIEN, V., LALK, G. and SHASHA, D. (2001). Efficient data reconciliation. *Inform. Sci.* **137** 1–15.
COLLINS, L. M., SCHAFER, J. L. and KAM, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Methods* **6** 330–351.
CONTI, G., FRÜHWIRTH-SCHNATTER, S., HECKMAN, J. J. and PIATEK, R. (2014). Bayesian exploratory factor analysis. *J. Econometrics* **183** 31–57. https://doi.org/10.1016/j.jeconom.2014.06.008
COOK, L. J., OLSON, L. M. and DEAN, J. M. (2001). Probabilistic record linkage: Relationships between file sizes, identifiers and match weights. *Methods Inf. Med.* **40** 196–203.
COPAS, J. B. and HILTON, F. J. (1990). Record linkage: Statistical models for matching computer records. *J. Roy. Statist. Soc. Ser. A* **153** 287–320.
D'ORAZIO, M., DI ZIO, M. and SCANU, M. (2006). *Statistical Matching*: *Theory and Practice*. Wiley, Hoboken, NJ.
DAGGY, J., XU, H., HUI, S. and GRANNIS, S. (2014). Evaluating latent class models with conditional dependence in record linkage. *Stat. Med.* **33** 4250–4265.
DALZELL, N. M. and REITER, J. P. (2018). Regression modeling and file matching using possibly erroneous matching variables. *J. Comput. Graph. Statist.* **27** 728–738.
DASYLVA, A., TITUS, R.-C. and THIBAULT, C. (2014). Overcoverage in the 2011 Canadian census. In *Proceedings of Statistics Canada Symposium*.
DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.
DI CONSIGLIO, L. and TUOTO, T. (2015). Coverage evaluation on probabilistically linked data. *J. Off. Stat.* **31** 415–429.
DI CONSIGLIO, L. and TUOTO, T. (2018). Population size estimation and linkage errors: The multiple lists case. *J. Off. Stat.* **34** 889–908.
DOIDGE, J. C. and HARRON, K. (2018). Demystifying probabilistic linkage: Common myths and misconceptions. *Int. J. Popul. Data Sci.* **3** 410.
DUSETZINA, S. B., TYREE, S., MEYER, A.-M., MEYER, A., GREEN, L. and CARPENTER, W. R. (2014). Linking data for health services research: A framework and instructional guide, rockville, MD: Agency for healthcare research and quality (US).
ELFEKY, M. G., VERYKIOS, V. S., ELMAGARMID, A. K., GHANEM, T. M. and KUWAIT, W. A. R. (2003). Record linkage: A machine learning approach, a toolbox, and a digital government web service. Purdue E-Pubs. Purdue University, West Lafayette.
ENAMORADO, T., FIFIELD, B. and IMAI, K. (2019). Using a probabilistic model to assist merging of large-scale administrative records. *Amer. Polit. Sci. Rev.*, **113** 353–371.
FELLEGI, I. P. and SUNTER, A. B. (1969). A theory for record linkage. *J. Amer. Statist. Assoc.* **64** 1183–1210.
FIENBERG, S. E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika* **59** 591–603.
FIENBERG, S. E. and MANRIQUE-VALLIER, D. (2009). Integrated methodology for multiple systems estimation and record linkage using a missing data formulation. *AStA Adv. Stat. Anal.* **93** 49–60.
FISHER, R. A. and YATES, F. (1963). *Statistical Tables for Biological*, *Agricultural and Medical Research*. Oliver and Boyd, Edinburgh, UK.
FORTINI, M., LISEO, B., NUCCITELLI, A. and SCANU, M. (2001). On Bayesian record linkage. *Res. Off. Stat.* **4** 185–198.
FREEDMAN, D. A. (2006). On the so-called "Huber sandwich estimator" and "Robust standard errors". *Amer. Statist.* **60** 299–302.
GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. CRC Press, Boca Raton.
GILBERT, R., LAFFERTY, R., HAGGER-JOHNSON, G., HARRON, K., ZHANG, L. C., SMITH, P., DIBBEN, C. and GOLDSTEIN, H. (2017). GUILD: GUidance for information about linking data sets. *J. Public Health* **40** 191–198.
GOLDEN, C. and MIREL, L. B. (2021). Enhancement of health surveys with data linkage. In *Administrative Records for Survey Methodology* (A. Y. Chun, M. Larsen, G. Durrant and J. P. Reiter, eds.) 105–138. Wiley, New York.
GOLDSTEIN, H., CARPENTER, J., KENWARD, M. G. and LEVIN, K. A. (2009). Multilevel models with multivariate mixed response types. *Stat. Model.* **9** 173–197.
GOLDSTEIN, H. and HARRON, K. (2015). Record linkage: A missing data problem. In *Methodological Developments in Data Linkage* (K. Harron, H. Goldstein and C. Dibben, eds.) **1** 109–124. Wiley, New York.
GOLDSTEIN, H., HARRON, K. and WADE, A. (2012). The analysis of record-linked data using multiple imputation with data value priors. *Stat. Med.* **31** 3481–3493. https://doi.org/10.1002/sim.5508
GOMATAM, S., CARTER, R., ARIET, M. and MITCHELL, G. (2002). An empirical comparison of record linkage procedures. *Stat. Med.* **21** 1485–1496.
GREEN, P. J. and MARDIA, K. V. (2006). Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika* **93** 235–254.
GU, L. and BAXTER, R. (2006). Decision models for record linkage. In *Data Mining*, *Lecture Notes in Computer Science* (G. J. Williams and S. J. Simoff, eds.) 146–160. Springer, Berlin, Heidelberg.
GUTMAN, R., AFENDULIS, C. C. and ZASLAVSKY, A. M. (2013). A Bayesian procedure for file linking to analyze end-of-life medical costs. *J. Amer. Statist. Assoc.* **108** 34–47. https://doi.org/10.1080/01621459.2012.726889

GUTMAN, R., SAMMARTINO, C. J., GREEN, T. C. and MONTAGUE, B. T. (2016). Error adjustments for file linking methods using encrypted unique client identifier (eUCI) with application to recently released prisoners who are HIV+. *Stat. Med.* **35** 115–129. https://doi.org/10.1002/sim.6586

HAAS, J. S., BRANDENBURG, J. A., UDVARHELYI, I. S. and EPSTEIN, A. M. (1994). Creating a comprehensive database to evaluate health coverage for pregnant women: The completeness and validity of a computerized linkage algorithm. *Med. Care* **32** 1053–1057. https://doi.org/10.1097/00005650-199410000-00006

HALL, R. and FIENBERG, S. (2012). Valid statistical inference on automatically matched files. In *Proceedings of the International Conference on Privacy in Statistical Databases* (J. Domingo-Ferrer and K. Muralidhar, eds.) 131–142.

HAN, Y. (2018). Statistical inference using data from multiple files combined through record linkage. Ph.D. thesis Univ. Maryland.

HAN, Y. and LAHIRI, P. (2019). Statistical analysis with linked data. *Int. Stat. Rev.* **87** 1013–1038.

HARRON, K., DIBBEN, C., BOYD, J., HJERN, A., AZIMAEE, M., BARRETO, M. L. and GOLDSTEIN, H. (2017). Challenges in administrative data linkage for research. *Big Data Soc.* **4**.

HARRON, K., WADE, A., GILBERT, R., MULLER-PEBODY, B. and GOLDSTEIN, H. (2014). Evaluating bias due to data linkage error in electronic healthcare records. *BMC Med. Res. Methodol.* **14** 1–10.

HOF, M. H. P., RAVELLI, A. C. and ZWINDERMAN, A. H. (2017). A probabilistic record linkage model for survival data. *J. Amer. Statist. Assoc.* **112** 1504–1515.

HOF, M. H. P. and ZWINDERMAN, A. H. (2012). Methods for analyzing data from probabilistic linkage strategies based on partially identifying variables. *Stat. Med.* **31** 4231–4242. https://doi.org/10.1002/sim.5498

HOF, M. H. P. and ZWINDERMAN, A. H. (2015). A mixture model for the analysis of data derived from record linkage. *Stat. Med.* **34** 74–92. https://doi.org/10.1002/sim.6315

ISAKI, C. T. and SCHULTZ, L. K. (1987). The effects of correlation and matching error on dual system estimation. *Comm. Statist. Theory Methods* **16** 2405–2427.

JARO, M. A. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *J. Amer. Statist. Assoc.* **84** 414–420.

JIANG, J., LAHIRI, P. and WAN, S.-M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Ann. Statist.* **30** 1782–1810.

KAMAT, G. and GUTMAN, R. (2026). Supplement to "Analysis of Linked Files: a Missing Data Perspective." https://doi.org/10.1214/24-STS939SUPP

KAMAT, G., SHAN, M. and GUTMAN, R. (2023). Bayesian record linkage with variables in one file. *Stat. Med.* **42** 4931–4951. https://doi.org/10.1002/sim.9894

KIM, G. and CHAMBERS, R. (2012a). Regression analysis under incomplete linkage. *Comput. Statist. Data Anal.* **56** 2756–2770.

KIM, G. and CHAMBERS, R. (2012b). Regression analysis under probabilistic multi-linkage. *Stat. Neerl.* **66** 64–79.

KIM, G. and CHAMBERS, R. (2015). Unbiased regression estimation under correlated linkage errors. *Stat* **4** 32–45.

KREWSKI, D., DEWANJI, A., WANG, Y., BARTLETT, S., ZIELINSKI, J. and MALLICK, R. (2005). The effect of record linkage errors on risk estimates in cohort mortality studies. *Surv. Methodol.* **31** 13–21.

LAHIRI, P. and LARSEN, M. D. (2005). Regression analysis with linked data. *J. Amer. Statist. Assoc.* **100** 222–230.

LARISCY, J. T. (2011). Differential record linkage by Hispanic ethnicity and age in linked mortality studies: Implications for the epidemiologic paradox. *J. Aging Health* **23** 1263–1284. https://doi.org/10.1177/0898264311421369

LARSEN, M. D. (2002). Comments on hierarchical Bayesian record linkage. In *Proceedings of the Survey Methods Section* 1995–2000. Amer. Statist. Assoc., Alexandria.

LARSEN, M. D. (2005). Advances in record linkage theory: Hierarchical Bayesian record linkage theory. In *JSM Proceedings of the Survey Research Methods Section* 3277–3284. Amer. Statist. Assoc., Alexandria.

LARSEN, M. D. and RUBIN, D. B. (2001). Iterative automated record linkage using mixture models. *J. Amer. Statist. Assoc.* **96** 32–41.

LEE, D., ZHANG, L.-C. and KIM, J.-K. (2022). Maximum entropy classification for record linkage. *Surv. Methodol.* **48**.

LEE, J. W. and HAREL, O. (2022). Incomplete clustering analysis via multiple imputation. *J. Appl. Stat.* **0** 1–18.

LEVENSHTEIN, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.* **10** 707–710.

LI, X., XU, H. and GRANNIS, S. (2022). The data-adaptive Fellegi–Sunter model for probabilistic record linkage: Algorithm development and validation for incorporating missing data and field selection. *J. Med. Internet Res.* **24** e33775.

LISEO, B. and TANCREDI, A. (2011). Bayesian estimation of population size via linkage of multivariate normal data sets. *J. Off. Stat.* **27** 491–505.

LITTLE, R. J. A. and RUBIN, D. B. (2019). *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ.

LIU, B., DAI, Y., LI, X., LEE, W. S. and YU, P. S. (2003). Building text classifiers using positive and unlabeled examples. In *Proceedings of Third IEEE International Conference on Data Mining* 179–186.

MADIGAN, D., RAFTERY, A. E., VOLINSKY, C. and HOETING, J. (1996). Bayesian model averaging. In *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, Portland, OR* 77–83.

MADIGAN, D., YORK, J. and ALLARD, D. (1995). Bayesian graphical models for discrete data. *Int. Stat. Rev.* **63** 215–232.

MARCHANT, N. G., KAPLAN, A., ELAZAR, D. N., RUBINSTEIN, B. I. P. and STEORTS, R. C. (2021). d-blink: Distributed end-to-end Bayesian entity resolution. *J. Comput. Graph. Statist.* **30** 406–421.

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

MCVEIGH, B., SPAHN, B. T. and MURRAY, J. S. (2020). Scaling Bayesian probabilistic record linkage with post-hoc blocking: An application to the California great registers. ArXiv.

MENG, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statist. Sci.* **9** 538–558.

MEYER, B. D. and MITTAG, N. (2019). Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net. *Am. Econ. J. Appl. Econ.* **11** 176–204.

MEYER, B. D. and MITTAG, N. (2021). Combining administrative and survey data to improve income measurement (A. Y. Chun, M. Larsen, G. Durrant and J. P. Reiter, eds.) 105–138. Wiley, New York.

MICHELSON, M. and KNOBLOCK, C. A. (2006). Learning blocking schemes for record linkage. In *Proceedings of the 21st National Conference on Artificial Intelligence–Volume* 1 440–445.

MOLENBERGHS, G., FITZMAURICE, G., KENWARD, M. G., TSIATIS, A. and VERBEKE, G. (2014). *Handbook of Missing Data Methodology*. CRC Press, Boca Raton.

MURRAY, J. S. (2015). Probabilistic record linkage and deduplication after indexing, blocking, and filtering. *J. Priv. Confid.* **7** https://doi.org/10.29012/jpc.v7i1.643.

MURRAY, J. S. (2018). Multiple imputation: A review of practical and theoretical findings. *Statist. Sci.* **33** 142–159. MR3797707 https://doi.org/10.1214/18-STS644

NASSIRI, V., LOVIK, A., MOLENBERGHS, G. and VERBEKE, G. (2018). On using multiple imputation for exploratory factor analysis of incomplete data. *Behav. Res.* **50** 501–517.

NETER, J., MAYNES, E. S. and RAMANATHAN, R. (1965). The effect of mismatching on the measurement of response errors. *J. Amer. Statist. Assoc.* **60** 1005–1027. MR0193729

NEWCOMBE, H. B., KENNEDY, J. M., AXFORD, S. J. and JAMES, A. P. (1959). Automatic linkage of vital records. *Science* **130** 954–959. https://doi.org/10.1126/science.130.3381.954

NEWMAN, L. M., SAMUEL, M. C., STENGER, M. R., GERBER, T. M., MACOMBER, K., STOVER, J. A. and WISE, W. (2009). Practical considerations for matching STD and HIV surveillance data with data from other sources. *Public Health Rep.* 2–27.

PATKI, D. and SHAPIRO, M. D. (2023). Implicates as instrumental variables: An approach for estimation and inference with probabilistically matched data. *J. Surv. Stat. Methodol.*.

REITER, J. P. (2021). Assessing uncertainty when using linked administrative records. In *Administrative Records for Survey Methodology* (A. Y. Chun, M. Larsen, G. Durrant and J. P. Reiter, eds.) 139–153. Wiley, Hoboken, NJ.

RENCHER, A. C. (2002). *Methods of Multivariate Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, New York. MR1885894 https://doi.org/10.1002/0471271357

ROBERT, C. P. and CASELLA, G. (1999). *Monte Carlo Statistical Methods*. *Springer Texts in Statistics*. Springer, New York. MR1707311 https://doi.org/10.1007/978-1-4757-3071-5

ROOS, L. L., WAJDA, A., NICOL, J. P. and ROBERTS, J. (1992). Record linkage: An overview. Technical Report Medical Effectiveness Research Data Methods, US Department of Health and Human Services.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0899519 https://doi.org/10.1002/9780470316696

RUBIN, D. B. (1996). Multiple imputation after 18+ years. *J. Amer. Statist. Assoc.* **91** 473–489.

SADINLE, M. (2014). Detecting duplicates in a homicide registry using a Bayesian partitioning approach. *Ann. Appl. Stat.* **8** 2404–2434. MR3292503 https://doi.org/10.1214/14-AOAS779

SADINLE, M. (2017). Bayesian estimation of bipartite matchings for record linkage. *J. Amer. Statist. Assoc.* **112** 600–612. MR3671755 https://doi.org/10.1080/01621459.2016.1148612

SADINLE, M. (2018). Bayesian propagation of record linkage uncertainty into population size estimation of human rights violations. *Ann. Appl. Stat.* **12** 1013–1038. MR3834293 https://doi.org/10.1214/18-AOAS1178

SADINLE, M. (2020). BRL: Beta record linkage. R package version 0.1.0.

SADINLE, M. and FIENBERG, S. E. (2013). A generalized Fellegi-Sunter framework for multiple record linkage with application to homicide record systems. *J. Amer. Statist. Assoc.* **108** 385–397. MR3174628 https://doi.org/10.1080/01621459.2012.757231

SCHAFER, J. L. (1997). *Analysis of Incomplete Multivariate Data*. *Monographs on Statistics and Applied Probability* **72**. CRC Press, London. MR1692799 https://doi.org/10.1201/9781439821862

SCHEUREN, F. and WINKLER, W. E. (1993). Regression analysis of data files that are computer matched–part I. *Surv. Methodol.* **19** 106–125.

SCHEUREN, F. and WINKLER, W. E. (1997). Regression analysis of data files that are computer matched–part II. *Surv. Methodol.* **19** 106–125.

SCHÜRLE, J. (2005). A method for consideration of conditional dependencies in the Fellegi and Sunter model of record linkage. *Statist. Papers* **46** 433–449. MR2253218 https://doi.org/10.1007/BF02762843

SCRUCCA, L., FOP, M., MURPHY, T. B. and RAFTERY, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8** 289–317.

SEAMAN, S. R. and WHITE, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22** 278–295. MR3190658 https://doi.org/10.1177/0962280210395740

SHAN, M., THOMAS, K. S. and GUTMAN, R. (2021). A multiple imputation procedure for record linkage and causal inference to estimate the effects of home-delivered meals. *Ann. Appl. Stat.* **15** 412–436. MR4255279 https://doi.org/10.1214/20-aoas1397

SHAN, M., THOMAS, K. S. and GUTMAN, R. (2023). A Bayesian multilayer record linkage procedure to analyze post-acute care recovery of patients with traumatic brain injury. *Biostatistics* **24** 743–759. MR4615251 https://doi.org/10.1093/biostatistics/kxac016

SHLOMO, N. (2019). Overview of data linkage methods for policy design and evaluation. In *Data-Driven Policy Impact Evaluation*: *How Access to Microdata Is Transforming Policy Design* (N. Crato and P. Paruolo, eds.) 47–65. Springer, Berlin.

SLAWSKI, M., DIAO, G. and BEN-DAVID, E. (2021). A pseudo-likelihood approach to linear regression with partially shuffled data. *J. Comput. Graph. Statist.* **30** 991–1003. MR4356600 https://doi.org/10.1080/10618600.2020.1870482

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

STEORTS, R. (2020). blink: Record linkage for empirically motivated priors. R package version 1.1.0.

STEORTS, R. C. (2015). Entity resolution with empirically motivated priors. *Bayesian Anal.* **10** 849–875. MR3432242 https://doi.org/10.1214/15-BA965SI

STEORTS, R. C., HALL, R. and FIENBERG, S. (2014). SMERED: A Bayesian approach to graphical record linkage and de-duplication. In *Artificial Intelligence and Statistics* 922–930.

STEORTS, R. C., HALL, R. and FIENBERG, S. E. (2016). A Bayesian approach to graphical record linkage and deduplication. *J. Amer. Statist. Assoc.* **111** 1660–1672. MR3601725 https://doi.org/10.1080/01621459.2015.1105807

STEORTS, R. C., TANCREDI, A. and LISEO, B. (2018). Generalized Bayesian record linkage and regression with exact error propagation. In *Proceedings of the International Conference on Privacy in Statistical Databases* (J. Domingo-Ferrer and K. Muralidhar, eds.) 297–313.

STEORTS, R. C., VENTURA, S. L., SADINLE, M. and FIENBERG, S. E. (2014). A comparison of blocking methods for record linkage. In *Proceedings of the International Conference on Privacy in Statistical Databases* (J. Domingo-Ferrer and K. Muralidhar, eds.) 253–268.

STERN, R. S. (1986). Record linkage. A powerful tool for epidemiologic analysis. *Arch. Dermatol.* **122** 1383–1384. https://doi.org/10.1001/archderm.122.12.1383

SWARTZ, T., HAITOVSKY, Y., VEXLER, A. and YANG, T. (2004). Bayesian identifiability and misclassification in multinomial data. *Canad. J. Statist.* **32** 285–302. MR2101757 https://doi.org/10.2307/3315930

TANCREDI, A. and LISEO, B. (2011). A hierarchical Bayesian approach to record linkage and population size problems. *Ann.*

Appl. *Stat*. **5** 1553–1585. MR2849786 https://doi.org/10.1214/10-AOAS447

TANCREDI, A. and LISEO, B. (2015). Regression analysis with linked data: Problems and possible solutions. *Statistica* **75** 19–35.

TANCREDI, A., STEORTS, R. and LISEO, B. (2020). A unified framework for de-duplication and population size estimation (with discussion). *Bayesian Anal*. **15** 633–682. MR4122517 https://doi.org/10.1214/19-BA1146

TANG, J., REITER, J. P. and STEORTS, R. C. (2020). Bayesian modeling for simultaneous regression and record linkage. In *Proceedings of the International Conference on Privacy in Statistical Databases* (J. Domingo-Ferrer and K. Muralidhar, eds.) 209–223.

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

TAYLOR, I., KAPLAN, A. and BETANCOURT, B. (2024). Fast Bayesian record linkage for streaming data contexts. *J. Comput. Graph. Statist*. **33** 833–844. MR4785788 https://doi.org/10.1080/10618600.2023.2283571

TEPPING, B. J. (1968). A model for optimum linkage of records. *J. Amer. Statist. Assoc*. **63** 1321–1332.

TROMP, M., RAVELLI, A. C., BONSEL, G., HASMAN, A. and REITSMA, J. B. (2011). Results from simulated data sets: Probabilistic record linkage outperforms deterministic record linkage. *J. Clin. Epidemiol*. **64** 564–572.

TROMP, M., REITSMA, J. B., RAVELLI, A. C., MÉRAY, N. and BONSEL, G. J. (2006). Record linkage: Making the most out of errors in linking variables. In *Proceedings of the AMIA Annual Symposium* 779–783.

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

VAN BUUREN, S. (2018). *Flexible Imputation of Missing Data*. CRC Press, Boca Raton.

VO, T. H., CHAUVET, G., HAPPE, A., OGER, E., PAQUELET, S. and GARÈS, V. (2023). Extending the Fellegi-Sunter record linkage model for mixed-type data with application to the French national health data system. *Comput. Statist. Data Anal*. **179** 107656. MR4510762 https://doi.org/10.1016/j.csda.2022.107656

VO, T. H., GARÉS, V., ZHANG, L.-C., HAPPE, A., OGER, E., PAQUELET, S. and CHAUVET, G. (2024). Cox regression with linked data. *Stat. Med*. **43** 296–314. MR4684817 https://doi.org/10.1002/sim.9960

WANG, Z., BEN-DAVID, E., DIAO, G. and SLAWSKI, M. (2022). Regression with linked datasets subject to linkage error. *Wiley Interdiscip. Rev.: Comput. Stat*. **14** e1570. MR4463249 https://doi.org/10.1002/wics.1570

WINGLEE, M., VALLIANT, R. and SCHEUREN, F. (2005). A case study in record linkage. *Surv. Methodol*. **31** 3–11.

WINKLER, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi–Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, *American Statistical Association* 354–359.

WINKLER, W. E. (1993). Improved decision rules in the Fellegi–Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, *American Statistical Association* 274–279.

WINKLER, W. E. (2000). Frequency-based matching in Fellegi–Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, *American Statistical Association* 778–783.

WINKLER, W. E. (2014). Matching and record linkage. *Wiley Interdiscip. Rev.: Comput. Stat*. **6** 313–325.

WINKLER, W. E. (2021). Cleaning and using administrative lists: Enhanced practices and computational algorithms for record linkage and modeling/editing/imputation. In *Administrative Records for Survey Methodology* (A. Y. Chun, M. Larsen, G. Durrant and J. P. Reiter, eds.) 271–296. Wiley, New York.

WINKLER, W. E. and THIBAUDEAU, Y. (1991). An application of the Fellegi–Sunter model of record linkage to the 1990 US decennial census. Technical Report US Bureau of the Census.

WORTMAN, J. H. and REITER, J. P. (2018). Simultaneous record linkage and causal inference with propensity score subclassification. *Stat. Med*. **37** 3533–3546. MR3862902 https://doi.org/10.1002/sim.7911

XIE, X. and MENG, X.-L. (2017). Dissecting multiple imputation from a multi-phase inference perspective: What happens when God's, imputer's and analyst's models are uncongenial? *Statist. Sinica* **27** 1485–1545. MR3701490

XU, H., LI, X. and GRANNIS, S. (2022). A simple two-step procedure using the Fellegi-Sunter model for frequency-based record linkage. *J. Appl. Stat*. **49** 2789–2804. MR4457289 https://doi.org/10.1080/02664763.2021.1922615

XU, H., LI, X., SHEN, C., HUI, S. L. and GRANNIS, S. (2019). Incorporating conditional dependence in latent class models for probabilistic record linkage: Does it matter? *Ann. Appl. Stat*. **13** 1753–1790. MR4019157 https://doi.org/10.1214/19-AOAS1256

XU, H., LI, X., ZHANG, Z. and GRANNIS, S. (2022). Score test for assessing the conditional dependence in latent class models and its application to record linkage. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **71** 1663–1687. MR4511127 https://doi.org/10.1111/rssc.12590

YANCEY, W. E. (2000). Frequency-dependent probability measures for record linkage. In *Proceedings of the Section on Survey Research Methods*, *American Statistical Association* 752–757.

ZANELLA, G., BETANCOURT, B., WALLACH, H., MILLER, J., ZAIDI, A. and STEORTS, R. C. (2016). Flexible models for microclustering with application to entity resolution. In *Proceedings of the* 30*th International Conference on Neural Information Processing Systems* 1425–1433.

ZHANG, L.-C. and TUOTO, T. (2021). Linkage-data linear regression. *J. Roy. Statist. Soc. Ser. A* **184** 522–547. MR4255909 https://doi.org/10.1111/rssa.12630

ZHOU, X. and REITER, J. P. (2010). A note on Bayesian inference after multiple imputation. *Amer. Statist*. **64** 159–163. MR2757007 https://doi.org/10.1198/tast.2010.09109

ZHU, V. J., OVERHAGE, M. J., EGG, J., DOWNS, S. M. and GRANNIS, S. J. (2009). An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *J. Amer. Med. Inform. Assoc*. **16** 738–745.

ZHU, Y., MATSUYAMA, Y., OHASHI, Y. and SETOGUCHI, S. (2015). When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J. Biomed. Inform*. **56** 80–86. https://doi.org/10.1016/j.jbi.2015.05.012

Statistical Science
2026, Vol. 41, No. 1, 49–68
https://doi.org/10.1214/24-STS944
This research was funded, in whole or in part, by UKRI Medical Research Council, MC_UU_00004/07, MC_UU_00004/09. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions.

On the Mixed-Model Analysis of Covariance in Cluster-Randomized Trials

Bingkai Wang, Michael O. Harhay, Jiaqi Tong, Dylan S. Small, Tim P. Morris and Fan Li

Abstract. In the analyses of cluster-randomized trials, mixed-model analysis of covariance (ANCOVA) is a standard approach for covariate adjustment and handling within-cluster correlations. However, when the normality, linearity, or the random-intercept assumption is violated, the validity and efficiency of the mixed-model ANCOVA estimators for estimating the average treatment effect remain unclear. Under the potential outcomes framework, we prove that the mixed-model ANCOVA estimators for the average treatment effect are consistent and asymptotically normal under arbitrary misspecification of its working model. If the probability of receiving treatment is 0.5 for each cluster, we further show that the model-based variance estimator under mixed-model ANCOVA1 (ANCOVA without treatment-covariate interactions) remains consistent, clarifying that the confidence interval given by standard software is asymptotically valid even under model misspecification. Beyond robustness, we discuss several insights on precision among classical methods for analyzing cluster-randomized trials, including the mixed-model ANCOVA, individual-level ANCOVA, and cluster-level ANCOVA estimators. These insights may inform the choice of methods in practice. Our analytical results and insights are illustrated via simulation studies and analyses of three cluster-randomized trials.

Key words and phrases: Average treatment effect, causal inference, cluster-randomized experiments, covariate adjustment, model robustness, precision.

REFERENCES

BALZER, L. B., PETERSEN, M. L., VAN DER LAAN, M. J. and COLLABORATION SEARCH (2016). Targeted estimation and inference for the sample average treatment effect in trials with and without pair-matching. *Stat. Med.* 35 3717–3732. MR3538043 https://doi.org/10.1002/sim.6965

BALZER, L. B., VAN DER LAAN, M., AYIEKO, J., KAMYA, M., CHAMIE, G., SCHWAB, J., HAVLIR, D. V. and PETERSEN, M. L. (2023). Two-Stage TMLE to reduce bias and improve efficiency in cluster randomized trials. *Biostatistics* **24** 502–517. MR4598783 https://doi.org/10.1093/biostatistics/kxab043

BUGNI, F., CANAY, I., SHAIKH, A. and TABORD-MEEHAN, M. (2023). Inference for Cluster Randomized Experiments with Non-ignorable Cluster Sizes. ArXiv preprint. Available at arXiv:2204.08356.

CAMPBELL, M. K., ELBOURNE, D. R. and ALTMAN, D. G. (2004). CONSORT statement: Extension to cluster randomised trials. *BMJ* **328** 702–708. https://doi.org/10.1136/bmj.328.7441.702

CAMPBELL, M. K., FAYERS, P. M. and GRIMSHAW, J. M. (2005). Determinants of the intracluster correlation coefficient in cluster randomized trials: The case of implementation research. *Clin. Trials* **2** 99–107. https://doi.org/10.1191/1740774505cn071oa

DING, P. (2024). *A first course in causal inference*. CRC Press, Boca Raton, FL, USA.

DONNER, A. and KLAR, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, London.

DRIKVANDI, R., VERBEKE, G. and MOLENBERGHS, G. (2017). Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics* **73** 63–71. MR3632352 https://doi.org/10.1111/biom.12551

Bingkai Wang is an Assistant Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: bingkai@umich.edu). Michael O. Harhay is an Assistant Professor, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: mharhay@pennmedicine.upenn.edu). Jiaqi Tong is a Ph.D. student, Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut 06520 (e-mail: jiaqi.tong@yale.edu). Dylan S. Small is Universal Furniture Professor, Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu). Tim P. Morris is a Principal Research Fellow, MRC Clinical Trials Unit at UCL, London, WC1V 6LJ, United Kingdom (e-mail: tim.morris@ucl.ac.uk). Fan Li is an Associate Professor, Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut 06520, USA (e-mail: fan.f.li@yale.edu).

ELDRIDGE, S. M., ASHBY, D., FEDER, G. S., RUDNICKA, A. R. and UKOUMUNNE, O. C. (2004). Lessons for cluster randomized trials in the twenty-first century: A systematic review of trials in primary care. *Clin. Trials* **1** 80–90. https://doi.org/10.1191/1740774504cn006rr

ELDRIDGE, S. M., ASHBY, D. and KERRY, S. (2006). Sample size for cluster randomized trials: Effect of coefficient of variation of cluster size and analysis method. *Int. J. Epidemiol.* **35** 1292–1300. https://doi.org/10.1093/ije/dyl129

ELDRIDGE, S. M., UKOUMUNNE, O. C. and CARLIN, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *Int. Stat. Rev.* **77** 378–394.

FIERO, M. H., HUANG, S., OREN, E. and BELL, M. L. (2016). Statistical analysis and handling of missing data in cluster randomized trials: A systematic review. *Trials* **17** 1–10.

GREEN, D. P. and VAVRECK, L. (2008). Analysis of cluster-randomized experiments: A comparison of alternative estimation approaches. *Polit. Anal.* **16** 138–152.

HAYES, R. J. and MOULTON, L. H. (2009). *Cluster Randomised Trials*. Taylor & Francis Group, LLC, Boca Raton, FL.

IMAI, K., KING, G. and NALL, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statist. Sci.* **24** 29–53. MR2561126 https://doi.org/10.1214/08-STS274

IVERS, N. M., HALPERIN, I. J., BARNSLEY, J., GRIMSHAW, J. M., SHAH, B. R., TU, K., UPSHUR, R. and ZWARENSTEIN, M. (2012). Allocation techniques for balance at baseline in cluster randomized trials: A methodological review. *Trials* **13** 1–9.

JIANG, J. (2010). *Large Sample Techniques for Statistics*. *Springer Texts in Statistics*. Springer, New York. MR2675055 https://doi.org/10.1007/978-1-4419-6827-2

JIANG, J. (2017). *Asymptotic Analysis of Mixed Effects Models*: *Theory*, *Applications*, *and Open Problems*. *Monographs on Statistics and Applied Probability* **155**. CRC Press, Boca Raton, FL. MR3676375

KAHAN, B. C., LI, F., BLETTE, B., JAIRATH, V., COPAS, A. and HARHAY, M. (2023a). Informative cluster size in cluster-randomised trials: A case study from the TRIGGER trial. *Clin. Trials*.

KAHAN, B. C., LI, F., COPAS, A. J. and HARHAY, M. O. (2023b). Estimands in cluster-randomized trials: Choosing analyses that answer the right question. *Int. J. Epidemiol.* **52** 107–118.

KELLY, E. L., MOEN, P., OAKES, J. M., FAN, W., OKECHUKWU, C., DAVIS, K. D., HAMMER, L. B., KOSSEK, E. E., KING, R. B. et al. (2014). Changing work and work-family conflict: Evidence from the work, family, and health network. *Am. Sociol. Rev.* **79** 485–516.

LI, F., LOKHNYGINA, Y., MURRAY, D. M., HEAGERTY, P. J. and DELONG, E. R. (2016). An evaluation of constrained randomization for the design and analysis of group-randomized trials. *Stat. Med.* **35** 1565–1579. MR3513471 https://doi.org/10.1002/sim.6813

LI, F., TIAN, Z., BOBB, J., PAPADOGEORGOU, G. and LI, F. (2022). Clarifying selection bias in cluster randomized trials. *Clin. Trials* **19** 33–41.

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

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

LITIÈRE, S., ALONSO, A. and MOLENBERGHS, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* **63** 1038–1044. MR2414580 https://doi.org/10.1111/j.1541-0420.2007.00782.x

LITIÈRE, S., ALONSO, A. and MOLENBERGHS, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Stat. Med.* **27** 3125–3144. MR2522153 https://doi.org/10.1002/sim.3157

MCCULLOCH, C. E. and NEUHAUS, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statist. Sci.* **26** 388–402. MR2917962 https://doi.org/10.1214/11-STS361

MCNEISH, D. and STAPLETON, L. M. (2016). Modeling clustered data with very few clusters. *Multivar. Behav. Res.* **51** 495–518.

MIDDLETON, J. A. and ARONOW, P. M. (2015). Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Stat. Polit. Policy* **6** 39–75.

MURRAY, D. M. and BLITSTEIN, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Eval. Rev.* **27** 79–103.

MURRAY, D. M., HANNAN, P. J., PALS, S. P., MCCOWEN, R. G., BAKER, W. L. and BLITSTEIN, J. L. (2006). A comparison of permutation and mixed-model regression methods for the analysis of simulated data in the context of a group-randomized trial. *Stat. Med.* **25** 375–388. MR2225172 https://doi.org/10.1002/sim.2233

MURRAY, D. M. et al. (1998). *Design and Analysis of Group-Randomized Trials* **29**. Oxford Univ. Press, USA.

NEUHAUS, J. M., MCCULLOCH, C. E. and BOYLAN, R. (2013). Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. *Stat. Med.* **32** 2419–2429. MR3067393 https://doi.org/10.1002/sim.5682

OFFORHA, B. C., WALTERS, S. J. and JACQUES, R. M. (2022). Statistical analysis of publicly funded cluster randomised controlled trials: A review of the national institute for health research journals library. *Trials* **23** 115.

OGEDEGBE, G., PLANGE-RHULE, J., GYAMFI, J., CHAPLIN, W., NTIM, M., APUSIGA, K., IWELUNMOR, J., AWUDZI, K. Y., QUAKYI, K. N. et al. (2018). Health insurance coverage with or without a nurse-led task shifting strategy for hypertension control: A pragmatic cluster randomized trial in Ghana. *PLoS Med.* **15** e1002561.

PARK, C. and KANG, H. (2023). Assumption-lean analysis of cluster randomized trials in infectious diseases for intent-to-treat effects and network effects. *J. Amer. Statist. Assoc.* **118** 1195–1206. MR4595487 https://doi.org/10.1080/01621459.2021.1983437

PRAGUE, M., WANG, R., STEPHENS, A., TCHETGEN TCHETGEN, E. and DEGRUTTOLA, V. (2016). Accounting for interactions and complex inter-subject dependency in estimating treatment effect in cluster-randomized trials with missing outcomes. *Biometrics* **72** 1066–1077. MR3591591 https://doi.org/10.1111/biom.12519

RAUDENBUSH, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychol. Methods* **2** 173.

ROCKERS, P. C., ZANOLINI, A., BANDA, B., CHIPILI, M. M., HUGHES, R. C., HAMER, D. H. and FINK, G. (2018). Two-year impact of community-based health screening and parenting groups on child development in Zambia: Follow-up to a cluster-randomized controlled trial. *PLoS Med.* **15** e1002555.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

SCHOCHET, P. Z., PASHLEY, N. E., MIRATRIX, L. W. and KAUTZ, T. (2022). Design-based ratio estimators and central limit theorems for clustered, blocked RCTs. *J. Amer. Statist. Assoc.* **117** 2135–2146. MR4528494 https://doi.org/10.1080/01621459.2021.1906685

SEAMAN, S., PAVLOU, M. and COPAS, A. (2014). Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Stat. Med.* **33** 5371–5387. MR3285402 https://doi.org/10.1002/sim.6277

SU, F. and DING, P. (2021). Model-assisted analyses of cluster-randomized experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 994–1015. MR4349125 https://doi.org/10.1111/rssb.12468

SULLIVAN, T. R., WHITE, I. R., SALTER, A. B., RYAN, P. and LEE, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat. Methods Med. Res.* **27** 2610–2626. MR3843399 https://doi.org/10.1177/0962280216683570

TONG, G., LI, F., CHEN, X., HIRANI, S. P., NEWMAN, S. P., WANG, W. and HARHAY, M. O. (2023). A Bayesian approach for estimating the survivor average causal effect when outcomes are truncated by death in cluster-randomized trials. *Amer. J. Epidemiol.* **192** 1006–1015.

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

TSIATIS, A. A., DAVIDIAN, M., ZHANG, M. and LU, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Stat. Med.* **27** 4658–4677. MR2528575 https://doi.org/10.1002/sim.3113

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

WANG, B., HARHAY, M. O, TONG, J., SMALL, D. S, MORRIS, T. P and LI, F. (2026). Supplement to "On the Mixed-Model Analysis of Covariance in Cluster-Randomized Trials." https://doi.org/10.1214/24-STS944SUPP

WANG, B., OGBURN, E. L. and ROSENBLUM, M. (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. *Biometrics* **75** 1391–1400. MR4041839 https://doi.org/10.1111/biom.13062

WANG, B., PARK, C., SMALL, D. S. and LI, F. (2024). Model-robust and efficient covariate adjustment for cluster-randomized experiments. *J. Amer. Statist. Assoc.* 1–13. https://doi.org/10.1080/01621459.2023.2289693

WANG, B., SUSUKIDA, R., MOJTABAI, R., AMIN-ESMAEILI, M. and ROSENBLUM, M. (2023). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *J. Amer. Statist. Assoc.* **118** 1152–1163. MR4595484 https://doi.org/10.1080/01621459.2021.1981338

WORK, FAMILY AND HEALTH STUDY (WFHS) (2018). Work, Family and Health Network. Inter-university Consortium for Political and Social Research [distributor].

YANG, L. and TSIATIS, A. A. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *Amer. Statist.* **55** 314–321. MR1943328 https://doi.org/10.1198/000313001753272466

ZHANG, D. and DAVIDIAN, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57** 795–802. MR1859815 https://doi.org/10.1111/j.0006-341X.2001.00795.x

TONG, G., ESSERMAN, D. and LI, F., (2022). Accounting for unequal cluster sizes in designing cluster randomized trials to detect treatment effect heterogeneity. *Stat. Med.* **41** 1376–1396. MR4396900 https://doi.org/10.1002/sim.9283

Statistical Science
2026, Vol. 41, No. 1, 69–98
https://doi.org/10.1214/24-STS948

Multivariate Matérn Models—a Spectral Approach

Drew Yarger, Stilian Stoev and Tailen Hsing

Abstract. The classical Matérn model has been a staple in spatial statistics. Novel data-rich applications in environmental and physical sciences, however, call for new, flexible vector-valued spatial and space-time models. Therefore, the extension of the classical Matérn model has been a problem of active theoretical and methodological interest. In this paper, we offer a new perspective to extending the Matérn covariance model to the vector-valued setting. We adopt a spectral, stochastic integral approach, which allows us to address challenging issues on the validity of the covariance structure and at the same time to obtain new, flexible, and interpretable models. In particular, our multivariate extensions of the Matérn model allow for asymmetric covariance structures. Moreover, the spectral approach provides an essentially complete flexibility in modeling the local structure of the process. We establish closed-form representations of the cross-covariances when available, compare them with existing models, simulate Gaussian instances of these new processes, and demonstrate estimation of the model's parameters through maximum likelihood. An application of the new class of multivariate Matérn models to data indicate their success in capturing inherent covariance-asymmetry phenomena.

Key words and phrases: Multivariate spatial statistics, cross-covariance functions, spectral analysis.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*: *With Formulas*, *Graphs and Mathematical Tables*. Dover, New York.

AGREST, M. M. and MAKSIMOV, M. S. (1971). *Theory of Incomplete Cylindrical Functions and Their Applications*. *Die Grundlehren der Mathematischen Wissenschaften*, *Band* 160. Springer, New York. Translated from the Russian by H. E. Fettis, J. W. Goresh and D. A. Lee. MR0346209

ALEGRÍA, A., EMERY, X. and PORCU, E. (2021). Bivariate Matérn covariances with cross-dimple for modeling coregionalized variables. *Spat*. *Stat*. **41** Paper No. 100491, 16. MR4205104 https://doi.org/10.1016/j.spasta.2021.100491

APANASOVICH, T. and GENTON, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97** 15–30. MR2594414 https://doi.org/10.1093/biomet/asp078

APANASOVICH, T. V., GENTON, M. G. and SUN, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J*. *Amer*. *Statist*. *Assoc*. **107** 180–193. MR2949350 https://doi.org/10.1080/01621459.2011.643197

ARGO (2020). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC).

AVERBUCH, A., COIFMAN, R. R., DONOHO, D. L., ELAD, M. and ISRAELI, M. (2006). Fast and accurate polar Fourier transform. *Appl*. *Comput*. *Harmon*. *Anal*. **21** 145–167. MR2259776 https://doi.org/10.1016/j.acha.2005.11.003

BABISTER, A. W. (1967). *Transcendental Functions Satisfying Nonhomogeneous Linear Differential Equations*. Macmillan Co., New York. Macmillan & Co., London. MR0206339

BARRADAS, I. and COHEN, J. E. (1994). Iterated exponentiation, matrix-matrix exponentiation, and entropy. *J*. *Math*. *Anal*. *Appl*. **183** 76–88. MR1273433 https://doi.org/10.1006/jmaa.1994.1132

BOCHNER, S. (1948). *Vorlesungen über Fouriersche Integrale*. Chelsea, New York.

BOLIN, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scand*. *J*. *Stat*. **41** 557–579. MR3249417 https://doi.org/10.1111/sjos.12046

BOLIN, D. and WALLIN, J. (2020). Multivariate type G Matérn stochastic partial differential equation random fields. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **82** 215–239. MR4060983

BYRD, R. H., LU, P., NOCEDAL, J. and ZHU, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J*.

Drew Yarger is Assistant Professor, Department of Statistics, Purdue University, West Lafayette, Indiana 47907, USA (e-mail: anyarger@purdue.edu). Stilian Stoev is Professor, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: sstoev@umich.edu). Tailen Hsing is Professor, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: thsing@umich.edu).

Sci. *Comput*. **16** 1190–1208. MR1346301 https://doi.org/10.1137/0916069

CARDOSO, J. R. and SADEGHI, A. (2018). Conditioning of the matrix-matrix exponentiation. *Numer. Algorithms* **79** 457–477. MR3857921 https://doi.org/10.1007/s11075-017-0446-2

CHILÈS, J.-P. and DELFINER, P. (2012). *Geostatistics*: *Modeling Spatial Uncertainty*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2850475 https://doi.org/10.1002/9781118136188

COOLEY, J. W. and TUKEY, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math*. *Comp*. **19** 297–301. MR0178586 https://doi.org/10.2307/2003354

CRAMÉR, H. (1942). On harmonic analysis in certain functional spaces. In *Breakthroughs in Statistics*: *Foundations and Basic Theory* 179–184 Springer, Berlin. https://doi.org/10.1007/978-1-4612-0919-5_12

CRESSIE, N. and ZAMMIT-MANGION, A. (2016). Multivariate spatial covariance models: A conditional approach. *Biometrika* **103** 915–935. MR3620448 https://doi.org/10.1093/biomet/asw045

DE IACO, S., PALMA, M. and POSA, D. (2003). Covariance functions and models for complex-valued random fields. *Stoch*. *Environ*. *Res*. *Risk Assess*. **17** 145–156. https://doi.org/10.1007/s00477-003-0129-5

DIDIER, G., MEERSCHAERT, M. M. and PIPIRAS, V. (2018). Domain and range symmetries of operator fractional Brownian fields. *Stochastic Process*. *Appl*. **128** 39–78. MR3729530 https://doi.org/10.1016/j.spa.2017.04.003

DIDIER, G. and PIPIRAS, V. (2011). Integral representations and properties of operator fractional Brownian motions. *Bernoulli* **17** 1–33. MR2797980 https://doi.org/10.3150/10-BEJ259

DINGLE, R. B. (1959). Asymptotic expansions and converging factors. V. Lommel, Struve, modified Struve, Anger and Weber functions, and integrals of ordinary and modified Bessel functions. *Proc*. *R*. *Soc*. *Lond*. *Ser*. *A* **249** 284–292. MR0103377 https://doi.org/10.1098/rspa.1959.0023

DU, J. and MA, C. (2011). Spherically invariant vector random fields in space and time. *IEEE Trans*. *Signal Process*. **59** 5921–5929. MR2907890 https://doi.org/10.1109/TSP.2011.2166391

DURAND, A. and ROUEFF, F. (2023). Weakly stationary stochastic processes valued in a separable Hilbert space: Gramian–Cramér representations and applications. *ESAIM Probab. Stat*. **27** 776–809. MR4634740 https://doi.org/10.1051/ps/2023014

EMERY, X. and ARROYO, D. (2018). On a continuous spectral algorithm for simulating non-stationary Gaussian random fields. *Stoch*. *Environ*. *Res*. *Risk Assess*. **32** 905–919. https://doi.org/10.1007/s00477-017-1402-3

EMERY, X., ARROYO, D. and PORCU, E. (2016). An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. *Stoch*. *Environ*. *Res*. *Risk Assess*. **30** 1863–1873. https://doi.org/10.1007/s00477-015-1151-0

EMERY, X., PORCU, E. and WHITE, P. (2022). New validity conditions for the multivariate Matérn coregionalization model, with an application to exploration geochemistry. *Math*. *Geosci*. **54** 1043–1068. MR4455143 https://doi.org/10.1007/s11004-022-10000-6

FAHMY, Y. and GUINNESS, J. (2022). Vecchia approximations and optimization for multivariate Matérn models. *J. Data Sci*. 475–492. https://doi.org/10.6339/22-JDS1074

FISCHER, A., GAUNT, R. E. and SARANTSEV, A. (2023). The variance-gamma distribution: a review. Available at arXiv:2303.05615.

FRIGO, M. (1999). A fast Fourier transform compiler. *ACM SIGPLAN Not*. **34** 169–180. https://doi.org/10.1145/301631.301661

GALASSI, M., DAVIES, J., THEILER, J., GOUGH, B., JUNGMAN, G., ALKEN, P., BOOTH, M., ROSSI, F. and ULERICH, R. (2002). GNU Scientific Library. Network Theory Limited Godalming.

GELFAND, A. E. and BANERJEE, S. (2010). Multivariate spatial process models. In *Handbook of Spatial Statistics* (A. Gelfand, P. Diggle, M. Fuentes and P. Guttorp, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 495–515. CRC Press, Boca Raton, FL. MR2730963 https://doi.org/10.1201/9781420072884-c28

GENTON, M. G. and KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist*. *Sci*. **30** 147–163. MR3353096 https://doi.org/10.1214/14-STS487

GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist*. *Assoc*. **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

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

GOULET, V. (2016). expint: Exponential Integral and Incomplete Gamma Function R package.

GRADSHTEYN, I. S., RYZHIK, I. M., ZWILLINGER, D. and MOLL, V., eds. (2015). *Table of Integrals*, *Series*, *and Products*, 8th ed. Elsevier, Amsterdam. Translated from the Russian, Translation edited and with a preface by Daniel Zwillinger and Victor Moll, Revised from the seventh edition [MR2360010]. MR3307944

GRIGORIU, M. (1998). Simulation of stationary non-Gaussian translation processes. *J*. *Eng*. *Mech*. **124** 121–126. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:2(121)

GRZEBYK, M. and WACKERNAGEL, H. (1994). Multivariate analysis and spatial/temporal scales: Real and complex models. In *Proceedings of XVIIth International Biometric Conference*.

GUINNESS, J. (2022a). Nonparametric spectral methods for multivariate spatial and spatial-temporal data. *J. Multivariate Anal*. **187** Paper No. 104823, 14. MR4319406 https://doi.org/10.1016/j.jmva.2021.104823

GUINNESS, J. (2022b). Inverses of Matérn covariances on grids. *Biometrika* **109** 535–541. MR4430973 https://doi.org/10.1093/biomet/asab017

GUINNESS, J., FUENTES, M., HESTERBERG, D. and POLIZZOTTO, M. (2014). Multivariate spatial modeling of conditional dependence in microscale soil elemental composition data. *Spat*. *Stat*. **9** 93–108. MR3326833 https://doi.org/10.1016/j.spasta.2014.03.009

HANČOVÁ, M., GAJDOŠ, A. and HANČ, J. (2022). A practical, effective calculation of gamma difference distributions with open data science tools. *J*. *Stat*. *Comput*. *Simul*. **92** 2205–2232. MR4445037 https://doi.org/10.1080/00949655.2021.2023873

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. MR0279952

HIGHAM, N. J. (2008). *Functions of Matrices*: *Theory and Computation*. SIAM, Philadelphia, PA. MR2396439 https://doi.org/10.1137/1.9780898717778

HOLMES, R. B. (1979). Mathematical foundations of signal processing. *SIAM Rev.* **21** 361–388. MR0535119 https://doi.org/10.1137/1021053

HU, X., SIMPSON, D., LINDGREN, F. and RUE, H. (2013). Multivariate Gaussian random fields using systems of stochastic partial differential equations. Available at arXiv:1307.1379 [stat].

KALLIANPUR, G. and MANDREKAR, V. (1971). Spectral theory of stationary H-valued processes. *J. Multivariate Anal.* **1** 1–16. MR0301788 https://doi.org/10.1016/0047-259X(71)90026-1

KING, F. W. (2009). *Hilbert Transforms*, *Vol*. 2. *Encyclopedia of Mathematics and Its Applications* **125**. Cambridge Univ. Press, Cambridge. MR2542215

KLAR, B. (2015). A note on gamma difference distributions. *J*. *Stat*. *Comput*. *Simul*. **85** 3708–3715. MR3401104 https://doi.org/10.1080/00949655.2014.996566

KLEIBER, W. (2017). Coherence for multivariate random fields. *Statist. Sinica* **27** 1675–1697. MR3701504

KLEIBER, W. and NYCHKA, D. (2012). Nonstationary modeling for multivariate spatial processes. *J. Multivariate Anal.* **112** 76–91. MR2957287 https://doi.org/10.1016/j.jmva.2012.05.011

KLEIBER, W. and PORCU, E. (2015). Nonstationary matrix covariances: Compact support, long range dependence and quasi-arithmetic constructions. *Stoch. Environ. Res. Risk Assess.* **29** 193–204. https://doi.org/10.1007/s00477-014-0867-6

KOTZ, S., KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston, MA. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1

KUUSELA, M. and STEIN, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. R. Soc. A, Math. Phys. Eng. Sci.* **474**. https://doi.org/10.1098/rspa.2018.0400

LI, B. and ZHANG, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structures. *J. Multivariate Anal.* **102** 1445–1453. MR2819961 https://doi.org/10.1016/j.jmva.2011.05.010

LIM, S. C. and EAB, C. H. (2021). Tempered fractional Brownian motion with variable index and variable tempering parameter. Available at arXiv:2104.04749 [math].

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. With discussion and a reply by the authors. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

MARTÍNEZ-HERNÁNDEZ, I. and GENTON, M. G. (2020). Recent developments in complex and spatially correlated functional data. *Braz. J. Probab. Stat.* **34** 204–229. MR4093256 https://doi.org/10.1214/20-BJPS466

MILLER, M. J. and REICH, B. J. (2022). Bayesian spatial modeling using random Fourier frequencies. *Spat. Stat.* **48** Paper No. 100598, 16. MR4388461 https://doi.org/10.1016/j.spasta.2022.100598

MOREVA, O. and SCHLATHER, M. (2023). Bivariate covariance functions of Pólya type. *J. Multivariate Anal.* **194** Paper No. 105099, 15. MR4524583 https://doi.org/10.1016/j.jmva.2022.105099

MU, W., DAVIS, E. S., REED, K., PHANSTIEL, D., LOVE, M. I. and LI, D. (2024). Gaussian processes for time series with lead-lag effects with applications to biology data. Available at arXiv:2401.07400 [stat].

NEEB, K.-H. (1998). Operator-valued positive definite kernels on tubes. *Monatsh. Math.* **126** 125–160. MR1639387 https://doi.org/10.1007/BF01473583

NYCHKA, D., FURRER, R., PAIGE, J. and SAIN, S. (2021). fields: tools for spatial data. R package version 14.1.

OLVER, F. W. J., OLDE DAALHUIS, A. B., LOZIER, D. W., SCHNEIDER, B. I., BOISVERT, R. F., CLARK, C. W., MILLER, B. R., SAUNDERS, B. V., COHL, H. S. et al., eds. (2021). NIST Digital Library of Mathematical Functions. Available at http://dlmf.nist.gov/. Release 1.1.3 of 2021-09-15.

PACIOREK, C. J. (2003). Nonstationary Gaussian processes for regression and spatial modelling. PhD thesis, Carnegie Mellon Univ. MR2704401

POPESCU, R., DEODATIS, G. and PREVOST, J. H. (1998). Simulation of homogeneous nonGaussian stochastic vector fields. *Probab. Eng. Mech.* **13** 1–13. https://doi.org/10.1016/s0266-8920(97)00001-5

PORCU, E., BEVILACQUA, M., SCHABACK, R. and OATES, C. J. (2024). The Matérn model: A journey through statistics, numerical analysis and machine learning. *Statist. Sci.* **39** 469–492. MR4766860 https://doi.org/10.1214/24-sts923

QADIR, G. A., EUÁN, C. and SUN, Y. (2021). Flexible modeling of variable asymmetries in cross-covariance functions for multivariate random fields. *J. Agric. Biol. Environ. Stat.* **26** 1–22. MR4207471 https://doi.org/10.1007/s13253-020-00414-2

REDFIN (2024). Redfin. Downloaded May 9, 2024. Available at https://www.redfin.com/news/data-center/.

SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional Integrals and Derivatives: Theory and Applications*. Gordon & Breach, Yverdon. Edited and with a foreword by S. M. Nikol'skiĭ, Translated from the 1987 Russian original, Revised by the authors. MR1347689

SCHLATHER, M., FREUDENBERG, A., FURRER, R., KROLL, M., RIPLEY, B. D. and RATCLIFF, J. W. (2022). RandomFieldsUtils: Utilites for the Simulation and Analysis of Random Fields R package version 1.2.5.

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* **44** 522–536. MR1501980 https://doi.org/10.2307/1989894

SHEN, J., STOEV, S. and HSING, T. (2022a). Tangent fields, intrinsic stationarity, and self similarity. *Electron. J. Probab.* **27** 1–56. MR4387842 https://doi.org/10.1214/22-ejp754

SHEN, J., STOEV, S. and HSING, T. (2022b). Supplement to "Tangent fields, intrinsic stationarity, and self similarity". Available at arXiv:2010.14715v3.

SOLAK, E., MURRAY-SMITH, R., LEITHEAD, W., LEITH, D. and RASMUSSEN, C. (2002). Derivative observations in Gaussian process models of dynamic systems. *Adv. Neural Inf. Process. Syst.* **15**.

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

TAFTI, P. D. and UNSER, M. (2010). Fractional Brownian vector fields. *Multiscale Model. Simul.* **8** 1645–1670. MR2728703 https://doi.org/10.1137/090752638

TERRES, M. A., FUENTES, M., HESTERBERG, D. and POLIZZOTTO, M. (2018). Bayesian spectral modeling for multivariate spatial distributions of elemental concentrations in soil. *Bayesian Anal.* **13** 1–28. MR3737941 https://doi.org/10.1214/16-BA1034

TON, J.-F., FLAXMAN, S., SEJDINOVIC, D. and BHATT, S. (2018). Spatial mapping with Gaussian processes and nonstationary Fourier features. *Spat. Stat.* **28** 59–78. MR3887156 https://doi.org/10.1016/j.spasta.2018.02.002

VAN DELFT, A. and EICHLER, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic Process. Appl.* **130** 3687–3710. MR4092417 https://doi.org/10.1016/j.spa.2019.10.006

VU, Q., ZAMMIT-MANGION, A. and CRESSIE, N. (2022). Modeling nonstationary and asymmetric multivariate spatial covariances via deformations. *Statist. Sinica* **32** 2071–2093. MR4478190

WALLIN, J. and BOLIN, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scand. J. Stat.* **42** 872–890. MR3391697 https://doi.org/10.1111/sjos.12141

WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, Cambridge. Macmillan Co., New York. MR0010746

WUERTZ, D. and SETZ, T. (2017). fAsianOptions: Rmetrics—EBM and Asian Option Valuation R package version 3042.82.

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions, Vol. II: Supplementary Notes and References. Springer Series in Statistics*. Springer, New York. MR0915557

YARGER, D., STOEV, S. and HSING, T. (2026). Supplement to "Multivariate Matérn Models—a Spectral Approach." https://doi.org/10.1214/24-STS948SUPPA, https://doi.org/10.1214/24-STS948SUPPB

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

Statistical Science
2026, Vol. 41, No. 1, 99–120
https://doi.org/10.1214/24-STS950

Review of Quasi-Randomization Approaches for Estimation from Nonprobability Samples

Vladislav Beresovsky, Julie Gershunskaya and Terrance D. Savitsky

Abstract. The recent proliferation of computers and the internet have opened new opportunities for collecting and processing data. However, such data are often obtained without a well-planned probability survey design. Such nonprobability based samples cannot be automatically regarded as representative of the population of interest. Several classes of methods for estimation and inferences from nonprobability samples have been developed in recent years. The quasi-randomization methods assume that nonprobability sample selection is governed by an underlying latent random mechanism. The basic idea is to use information collected from a probability ("reference") sample to uncover latent nonprobability survey participation probabilities (also known as "propensity scores") and use them in estimation of target finite population parameters. In this paper, we review and compare theoretical properties of recently developed methods of estimation survey participation probabilities and study their relative performances in simulations.

Key words and phrases: Data combining, nonprobability sample, reference sample, participation probabilities, sample likelihood, variance estimation.

REFERENCES

BEAUMONT, J.-F. and RAO, J. N. K. (2021). Pitfalls of making inferences from non-probability samples: Can data integration through probability samples provide remedies. *Surv. Stat.* **83** 11–22.

BERESOVSKY, V. (2019). On application of a response propensity model to estimation from web samples. In ResearchGate. Available at https://www.researchgate.net/publication/333915871_On_application_of_a_response_propensity_model_to_estimation_from_web_samples.

BERESOVSKY, V., GERSHUNSKAYA, J. and SAVITSKY, T. D. (2026). Supplement to "Review of Quasi-Randomization Approaches for Estimation from Nonprobability Samples." https://doi.org/10.1214/24-STS950SUPP

BETHLEHEM, J. (2010). Selection bias in web surveys. *Int. Stat. Rev.* **78** 161–188.

CHEN, Y., LI, P. and WU, C. (2020). Doubly robust inference with non-probability survey samples. *J. Amer. Statist. Assoc.* **115** 2011–2021. MR4189773 https://doi.org/10.1080/01621459.2019.1677241

ELLIOTT, M. R. (2009). Combining data from probability and non-probability samples using pseudo-weights. *Surv. Pract.* **2** 813–845.

ELLIOTT, M. R. and VALLIANT, R. (2017). Inference for nonprobability samples. *Statist. Sci.* **32** 249–264. MR3648958 https://doi.org/10.1214/16-STS598

FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38. MR1225212 https://doi.org/10.1093/biomet/80.1.27

HEINZE, G. and SCHEMPER, M. (2002). A solution to the problem of separation in logistic regression. *Stat. Med.* **21** 2409–2419.

KIM, J.-K. and TAM, S.-M. (2021). Data integration by combining big data and survey sample data for finite population inference. *Int. Stat. Rev.* **89** 382–401. MR4411911 https://doi.org/10.1111/insr.12434

KRUSKAL, W. and MOSTELLER, F. (1980). Representative sampling. IV. The history of the concept in statistics, 1895–1939. *Int. Stat. Rev.* **48** 169–195. MR0586104 https://doi.org/10.2307/1403151

LIU, A.-C., SCHOLTUS, S. and DE WAAL, T. (2023). Correcting selection bias in big data by pseudo-weighting. *J. Surv. Stat. Methodol.* **11** 1181–1203.

MENG, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *Ann. Appl. Stat.* **12** 685–726. MR3834282 https://doi.org/10.1214/18-AOAS1161SF

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. R. Stat. Soc.* **97** 558–625.

RAO, J. N. K. (2021). On making valid inferences by integrating data from surveys and other sources. *Sankhya B* **83** 242–272. MR4256318 https://doi.org/10.1007/s13571-020-00227-w

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

Vladislav Beresovsky is Research Mathematical Statistician, U.S. Bureau of Labor Statistics, Washington, DC 20212, USA (e-mail: beresovsky.vladislav@bls.gov). Julie Gershunskaya is Mathematical Statistician, U.S. Bureau of Labor Statistics, Washington, DC 20212, USA (e-mail: gershunskaya.julie@bls.gov). Terrance D. Savitsky is Senior Research Mathematical Statistician, U.S. Bureau of Labor Statistics, Washington, DC 20212, USA (e-mail: savitsky.terrance@bls.gov).

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. With comments by R. J. A. Little and a reply by the author. MR0455196 https://doi.org/10.1093/biomet/63.3.581

SAVITSKY, T. D., WILLIAMS, M. R., GERSHUNSKAYA, J. and BERESOVSKY, V. (2023). Methods for combining probability and nonprobability samples under unknown overlaps. *Stat. Transit. New Ser.* **24** 34.

TILLE, Y. and MATEI, A. (2021). sampling: Survey Sampling. Available at https://CRAN.R-project.org/package=sampling.

VALLIANT, R. (2020). Comparing alternatives for estimation from nonprobability samples. *J. Surv. Stat. Methodol.* **8** 231–263.

VANDERWEELE, T. J. and SHPITSER, I. (2011). A new criterion for confounder selection. *Biometrics* **67** 1406–1413. MR2872391 https://doi.org/10.1111/j.1541-0420.2011.01619.x

WANG, L., VALLIANT, R. and LI, Y. (2021). Adjusted logistic propensity weighting methods for population inference using nonprobability volunteer-based epidemiologic cohorts. *Stat. Med.* **40** 5237–5250. MR4330549 https://doi.org/10.1002/sim.9122

WILLIAMS, D. and BRICK, J. M. (2017). Trends in U.S. face-to-face household survey nonresponse and level of effort. *J. Surv. Stat. Methodol.* **6** 186–211. https://doi.org/10.1093/jssam/smx019

WU, C. (2022). A new criterion for confounder selection. *Surv. Methodol.* **48** 283–311.

Statistical Science
2026, Vol. 41, No. 1, 121–142
https://doi.org/10.1214/24-STS952

Randomized and Exchangeable Improvements of Markov's, Chebyshev's and Chernoff's Inequalities

Aaditya Ramdas and Tudor Manole

Abstract. We present simple randomized and exchangeable improvements of Markov's inequality, as well as Chebyshev's inequality and Chernoff bounds. Our variants are never worse and typically strictly more powerful than the original inequalities. The proofs are short and elementary, and can easily yield similarly randomized or exchangeable versions of a host of other inequalities (e.g., martingale inequalities by Doob and Ville) that employ Markov's inequality as an intermediate step. We point out some simple statistical applications involving tests that combine dependent e-values. In particular, we uniformly improve the power of universal inference, and obtain tighter betting-based nonparametric confidence intervals. Simulations reveal nontrivial gains in power (and no losses) in a variety of settings.

Key words and phrases: Randomization, exchangeability, e-values, universal inference, Markov, Chernoff, Ville, Doob, Chebyshev, Bernstein, Hoeffding, martingales.

REFERENCES

BENTKUS, V. (2002). A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Liet. Mat. Rink.* **42** 262–269. MR1947624 https://doi.org/10.1023/A:1020221925664

BENTKUS, V., KALOSHA, N. and VAN ZUIJLEN, M. (2006). On domination of tail probabilities of (super)martingales: Explicit bounds. *Liet. Mat. Rink.* **46** 1–43. MR2251440 https://doi.org/10.1007/s10986-006-0011-5

CANTELLI, F. P. (1929). Sui confini della probabilita. In *Atti del Congresso Internazionale dei Matematici*: *Bologna del* 3 *al* 10 *de Settembre di* 1928 47–60.

CHEN, J. and LI, P. (2009). Hypothesis test for normal mixture models: The EM approach. *Ann. Statist.* **37** 2523–2542. MR2543701 https://doi.org/10.1214/08-AOS651

CHOI, W. and KIM, I. (2023). Averaging p-values under exchangeability. *Statist. Probab. Lett.* **194** Paper No. 109748, 8. MR4521761 https://doi.org/10.1016/j.spl.2022.109748

CHRISTOFIDES, T. C. and SERFLING, R. J. (1990). Maximal inequalities for multidimensionally indexed submartingale arrays. *Ann. Probab.* **18** 630–641. MR1055424

DACUNHA-CASTELLE, D. and GASSIAT, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *Ann. Statist.* **27** 1178–1209. MR1740115 https://doi.org/10.1214/aos/1017938921

DOOB, J. L. (1940). Regularity properties of certain families of chance variables. *Trans. Amer. Math. Soc.* **47** 455–486. MR0002052 https://doi.org/10.2307/1989964

DUNN, R., GANGRADE, A., WASSERMAN, L. and RAMDAS, A. (2024). Universal inference meets random projections: A scalable test for log-concavity. *J. Comput. Graph. Statist.* 1–13.

DUNN, R., RAMDAS, A., BALAKRISHNAN, S. and WASSERMAN, L. (2023). Gaussian universal likelihood ratio testing. *Biometrika* **110** 319–337. MR4588356 https://doi.org/10.1093/biomet/asac064

DURRETT, R. (2019). *Probability—Theory and Examples*. *Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge Univ. Press, Cambridge. Fifth edition of [MR1068527]. MR3930614 https://doi.org/10.1017/9781108591034

GHOSH, B. K. (2002). Probability inequalities related to Markov's theorem. *Amer. Statist.* **56** 186–190. MR1940206 https://doi.org/10.1198/000313002119

GHOSH, J. K. and SEN, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol. II* (*Berkeley*, *Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 789–806. Wadsworth, Belmont, CA. MR0822065

GOFFINET, B., LOISEL, P. and LAURENT, B. (1992). Testing in normal mixture models when the proportions are known. *Biometrika* **79** 842–846. MR1209483 https://doi.org/10.1093/biomet/79.4.842

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. (2024). Safe testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1091–1128. With discussion. MR4825000 https://doi.org/10.1093/jrsssb/qkae011

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

Aaditya Ramdas is Associate Professor, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: aramdas@cmu.edu). Tudor Manole is Postdoctoral Associate, Statistics and Data Science Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA (e-mail: tmanole@mit.edu).

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probab. Surv.* **17** 257–317. MR4100718 https://doi.org/10.1214/18-PS321

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

HUBER, M. (2019). Halving the bounds for the Markov, Chebyshev, and Chernoff inequalities using smoothing. *Amer. Math. Monthly* **126** 915–927. MR4033226 https://doi.org/10.1080/00029890.2019.1656484

IGNATIADIS, N., WANG, R. and RAMDAS, A. (2024). E-values as unnormalized weights in multiple testing. *Biometrika* **111** 417–439. MR4745574 https://doi.org/10.1093/biomet/asad057

LEE, A. J. (1990). *U-Statistics: Theory and Practice*. *Statistics: Textbooks and Monographs* **110**. Dekker, New York. MR1075417

MANOLE, T. and RAMDAS, A. (2023). Martingale methods for sequential estimation of convex functionals and divergences. *IEEE Trans. Inf. Theory* **69** 4641–4658. MR4613565 https://doi.org/10.1109/tit.2023.3250099

RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

RAMDAS, A. and MANOLE, T. (2026). Supplement to "Randomized and exchangeable improvements of Markov's, Chebyshev's and Chernoff's inequalities." https://doi.org/10.1214/24-STS952SUPPA, https://doi.org/10.1214/24-STS952SUPPB

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. Available at arXiv:2009.03167.

ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Stat.* **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

RÜGER, B. (1978). Das maximale Signifikanzniveau des Tests: "Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen". *Metrika* **25** 171–178. MR0526476 https://doi.org/10.1007/BF02204362

RÜSCHENDORF, L. (1982). Random variables with maximum sums. *Adv. in Appl. Probab.* **14** 623–632. MR0665297 https://doi.org/10.2307/1426677

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication (with discussion and response). *J. Roy. Statist. Soc. Ser. A* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

SHAFER, G. and VOVK, V. (2001). *Probability and Finance: It's Only a Game!*. *Wiley Series in Probability and Statistics. Financial Engineering Section*. Wiley-Interscience, New York. MR1852450 https://doi.org/10.1002/0471249696

SHAFER, G. and VOVK, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley-Interscience, New York.

SHEKHAR, S. and RAMDAS, A. (2024). Nonparametric two-sample testing by betting. *IEEE Trans. Inf. Theory* **70** 1178–1203. MR4703709 https://doi.org/10.1109/tit.2023.3305867

TCHEBICHEF, P. (1874). Sur les valeurs limites des intégrales. *J. Math. Pures Appl.* **19** 157–160.

VILLE, J. (1939). *Étude Critique de la Notion de Collectif*. Gauthier-Villars. MR3533075

VOVK, V. and WANG, R. (2020). Combining p-values via averaging. *Biometrika* **107** 791–808. MR4186488 https://doi.org/10.1093/biomet/asaa027

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

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

WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. With discussion. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

Statistical Science
2026, Vol. 41, No. 1, 143–162
https://doi.org/10.1214/24-STS953

What You See Is Not What Is There: Mechanisms, Models and Methods for Point Pattern Deviations

Peter Guttorp, Janine Illian, Joel Kostensalo, Mikko Kuronen, Mari Myllymäki, Aila Särkkä and Thordis L. Thorarinsdottir

Abstract. Many natural systems are observed as point patterns in time, space, or space and time. Examples include plant and cellular systems, animal colonies, earthquakes and wildfires. In practice, the locations of the points are not always observed correctly. However, in the point process literature, there has been relatively scant attention paid to the issue of errors in the location of points. In this paper, we discuss how the observed point pattern may deviate from the actual point pattern and review methods and models that exist to handle such deviations. The discussion is supplemented with several scientific illustrations.

Key words and phrases: Ghost point, measurement error, missing data, point process, thinning.

REFERENCES

ALTWEGG, R. and NICHOLS, J. D. (2019). Occupancy models for citizen-science data. *Methods Ecol*. *Evol*. **10** 8–21.

ANDERSSON, C., GUTTORP, P. and SÄRKKÄ, A. (2016). Discovering early diabetic neuropathy from epidermal nerve fiber patterns. *Stat*. *Med*. **35** 4427–4442. MR3554972 https://doi.org/10.1002/sim.7009

ANSELIN, L. (1995). Local indicator of spatial association—LISA. *Geogr. Anal*. **27** 93–115.

ARCE GUILLEN, R., LINDGREN, F., MUFF, S., GLASS, T. W., BREED, G. A. and SCHLÄGEL, U. E. (2023). Accounting for unobserved spatial variation in step selection analyses of animal movement via spatial random effects. *Methods Ecol*. *Evol*. **14** 2639–2653.

ARMSTRONG, M. P., RUSHTON, G. and ZIMMERMAN, D. L. (1999). Geographically masking health data to preserve confidentiality. *Stat. Med*. **18** 497–525.

ASSUNÇÃO, R. and GUTTORP, P. (1999). Robustness for inhomogeneous Poisson point processes. *Ann*. *Inst*. *Statist*. *Math*. **51** 657–678. MR1735320 https://doi.org/10.1023/A:1004079013014

ASTRUP, R., DUCEY, M. J., GRANHUS, A., RITTER, T. and VON LÜPKE, N. (2014). Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. *Can*. *J*. *For*. *Res*. **44** 666–676.

BACHL, F. E., LINDGREN, F., BORCHERS, D. L. and ILLIAN, J. B. (2019). Inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods Ecol*. *Evol*. **10** 760–766.

BADDELEY, A., CHANG, Y.-M. and SONG, Y. (2013). Leverage and influence diagnostics for spatial point processes. *Scand*. *J*. *Stat*. **40** 86–104. MR3024033 https://doi.org/10.1111/j.1467-9469.2011.00786.x

BADDELEY, A., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns*: *Methodology and Applications with R*. CRC Press, Boca Raton .

BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **67** 617–666. MR2210685 https://doi.org/10.1111/j.1467-9868.2005.00519.x

BAR-HEN, A., CHADŒUF, J., DESSARD, H. and MONESTIEZ, P. (2013). Estimating second order characteristics of point processes with known independent noise. *Stat*. *Comput*. **23** 297–309. MR3041437 https://doi.org/10.1007/s11222-011-9311-7

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

BORCHERS, D. L., ZUCCHINI, W. and FEWSTER, R. M. (1998). Mark-recapture models for line transect surveys. *Biometrics* **54** 1207–1220.

BOWLER, D. E., NILSEN, E. B., BISCHOF, R., O'HARA, R. B., YU, T. T., OO, T., AUNG, M. and LINNELL, J. D. (2019). Integrat-

Peter Guttorp is a Professor II, Norwegian Computing Center, Oslo, Norway (e-mail: peter.guttorp@nr.no). Janine Illian is a Professor, School of Mathematics and Statistics, University of Glasgow, Glasgow, Scotland, UK (e-mail: Janine.Illian@glasgow.ac.uk). Joel Kostensalo is a Senior Scientist, Natural Research Institute Finland (Luke), Joensuu, Finland (e-mail: joel.kostensalo@luke.fi). Mikko Kuronen is a Research Scientist, Natural Research Institute Finland (Luke), Helsinki, Finland (e-mail: mikko.kuronen@luke.fi). Mari Myllymäki is Senior Scientist, Natural Research Institute Finland (Luke), Helsinki, Finland (e-mail: mari.myllymaki@luke.fi). Aila Särkkä is a Professor, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden (e-mail: aila@chalmers.se). Thordis L. Thorarinsdottir is a Professor, Department of Mathematics, University of Oslo, Oslo, Norway (e-mail: thordist@math.uio.no).

ing data from different survey types for population monitoring of an endangered species: The case of the Elås deer. *Sci*. *Rep*. **9** 1–14.

BROCKNER, B., WILLIAMS, K., LUKE, R., SHEEN, D., DOWDY, J., ANDERSON, D. and VEAL, C. (2018). Convolutional neural network based side attack explosive hazard detection in three dimensional voxel radar. In *Detection and Sensing of Mines*, *Explosive Objects and Obscured Targets XXIII* 51.

BUCKLAND, S. T. (2006). Point-transect surveys for songbirds: Robust methodologies. *The Auk* **123** 345–357.

BUCKLAND, S. T., ANDERSON, D. R., BURNHAM, K. P., LAAKE, J. L., BORCHERS, D. L. and THOMAS, L. (2001). *Introduction to Distance Sampling*: *Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

BUCKLAND, S. T., REXSTAD, E. A., MARQUES, T. A. and OEDEKOVEN, C. S. (2015). *Distance Sampling*: *Methods and Applications* **431**. Springer, Berlin.

BYERS, S. and RAFTERY, A. E. (1998). Nearest-neighbor clutter removal for estimating features in spatial point processes. *J*. *Amer*. *Statist*. *Assoc*. **93** 577–584.

CHAKRABORTY, A. and GELFAND, A. E. (2010). Analyzing spatial point patterns subject to measurement error. *Bayesian Anal*. **5** 97–122. MR2596437 https://doi.org/10.1214/10-BA504

CHANDLER, R. B., ROYLE, J. A. and KING, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology* **92** 1429–1435. https://doi.org/10.1890/10-2433.1

CHANG, X., WAAGEPETERSEN, R., YU, H., MA, X., HOLFORD, T. R., WANG, R. and GUAN, Y. (2015). Disease risk estimation by combining case-control data with aggregated information on the population at risk. *Biometrics* **71** 114–121. MR3335355 https://doi.org/10.1111/biom.12256

COEURJOLLY, J.-F. (2017). Median-based estimation of the intensity of a spatial point process. *Ann*. *Inst*. *Statist*. *Math*. **69** 303–331. MR3611522 https://doi.org/10.1007/s10463-015-0536-7

CONNOR, B., CARRIE, I. D., CRAIG, R. and PARSONS, J. (2008). Discriminative imaging using LWIR polarimetric imaging. In *Proceedings of the SPIE Security and Defence Conference* (R. R. Ebert and D. A. Huckridge, eds.) **7113**. SPIE, Bellingham, WA.

CRESSIE, N. and BRANT COLLINS, L. (2001). Analysis of spatial point patterns using bundles of product density LISA functions. *J*. *Agric*. *Biol*. *Environ*. *Stat*. **6** 118–135. MR1812096 https://doi.org/10.1198/108571101300325292

CRESSIE, N. and LAWSON, A. B. (2000). Hierarchical probability models and Bayesian analysis of mine locations. *Adv*. *in Appl*. *Probab*. **32** 315–330. MR1778566 https://doi.org/10.1239/aap/1013540165

CUCALA, L. (2008). Intensity estimation for spatial point processes observed with noise. *Scand*. *J*. *Stat*. **35** 322–334. MR2418744 https://doi.org/10.1111/j.1467-9469.2007.00583.x

D'ANGELO, N., SIINO, M., D'ALESSANDRO, A. and ADELFIO, G. (2022). Local spatial log-Gaussian Cox processes for seismic data. *AStA Adv*. *Stat*. *Anal*. **106** 633–671. MR4510591 https://doi.org/10.1007/s10182-022-00444-w

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*. *Vol*. *I*: *Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

DASGUPTA, A. and RAFTERY, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *J*. *Amer*. *Statist*. *Assoc*. **93** 294–302.

DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **34** 138–147.

DIGGLE, P. J. (1993). Point process modelling in environmental epidemiology. In *Statistics for the Environment* (V. Barnett and F. Turkman, eds.) 89–110. Wiley, New York.

DORAZIO, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Glob*. *Ecol*. *Biogeogr*. **23** 1472–1484.

DRALLE, K. and RUDEMO, M. (1996). Stem number estimation by kernel smoothing of aerial photos. *Can*. *J*. *For*. *Res*. **26** 1228–1236.

DRALLE, K. and RUDEMO, M. (1997). Automatic estimation of individual tree positions from aerial photos. *Can*. *J*. *For*. *Res*. **27** 1728–1736.

DUCEY, M. J. and ASTRUP, R. (2013). Adjusting for nondetection in forest inventories derived from terrestrial laser scanning. *Can*. *J*. *Remote Sens*. **39** 410–425.

ECKERSTORFER, M., BÜHLER, Y., FRAUENFELDER, R. and MALNES, E. (2016). Remote sensing of snow avalanches: Recent advances, potential, and limitations. *Cold Reg*. *Sci*. *Technol*. **121** 126–140.

FANSHAWE, T. R. and DIGGLE, P. J. (2011). Spatial prediction in the presence of positional error. *Environmetrics* **22** 109–122. MR2843341 https://doi.org/10.1002/env.1062

FRONTERRÈ, C., GIORGI, E. and DIGGLE, P. (2018). Geostatistical inference in the presence of geomasking: A composite-likelihood approach. *Spat*. *Stat*. **28** 319–330. MR3887171 https://doi.org/10.1016/j.spasta.2018.06.004

GLENNIE, R., BUCKLAND, S. T., LANGROCK, R., GERRODETTE, T., BALLANCE, L. T., CHIVERS, S. J. and SCOTT, M. D. (2021). Incorporating animal movement into distance sampling. *J*. *Amer*. *Statist*. *Assoc*. **116** 107–115. MR4227678 https://doi.org/10.1080/01621459.2020.1764362

GUTTORP, P. (1995). *Stochastic Modeling of Scientific Data*. *Stochastic Modeling Series*. CRC Press, London. MR1358359 https://doi.org/10.1007/978-1-4899-4449-8

GUTTORP, P., ILLIAN, J., KOSTENSALO, J., KURONEN, M., MYLLYMÄKI, M., SÄRKKÄ, A. and THORARINSDOTTIR, T. L (2026). Supplement to "What you see is not what is there: Mechanisms, models and methods for point pattern deviations." https://doi.org/10.1214/24-STS953SUPP

GUTTORP, P. and THOMPSON, M. L. (1990). A note on point process analysis of Chinese earthquake data. *Pure Appl*. *Geophys*. **133** 541–546.

GUTTORP, P. and THORARINSDOTTIR, T. L. (2012). What happened to discrete chaos, the Quenouille process, and the sharp Markov property? Some history of stochastic point processes. *Int*. *Stat*. *Rev*. **80** 253–268. MR2990354 https://doi.org/10.1111/j.1751-5823.2012.00181.x

HÄBEL, H., BALAZS, A. and MYLLYMÄKI, M. (2021). Spatial analysis of airborne laser scanning point clouds for predicting forest structure. *Math*. *Comput*. *For*. *Nat*. *Res*. *Sci*. **13** 15–28.

HABLE, S., SIGLOCH, K., BARRUOL, G., STÄHLER, S. C. and HADZIIOANNOU, C. (2018). Clock errors in land and ocean bottom seismograms: High-accuracy estimates from multiple-component noise cross-correlations. *Geophys*. *J*. *Int*. **214** 2014–2034.

HAYES, R. and BUCKLAND, S. (1983). Radial-distance models for the line-transect method. *Biometrics* **39** 29–42.

HEDLEY, S. L. and BUCKLAND, S. T. (2004). Spatial models for line transect sampling. *J*. *Agric*. *Biol*. *Environ*. *Stat*. **9** 181–199.

HÖGMANDER, H. (1991). A random field approach to transect counts of wildlife populations. *Biom*. *J*. **33** 1013–1023. MR1201875 https://doi.org/10.1002/bimj.4710330813

HOLMES, Q. A., SCHWARTZ, C. R., SELDIN, J. H., WRIGHT, J. A. and WITTER, L. J. (1995). Adaptive multispectral CFAR detection of land mines. In *Detection Technologies for Mines and Minelike Targets* (A. C. Dubey, I. Cindrich, J. M. Ralston and K. A. Rigano, eds.). *Society of Photo-Optical Instrumentation Engineers* (*SPIE*) *Conference Series* **2496** 421–432.

HUANG, H., MA, X., WAAGEPETERSEN, R., HOLFORD, T. R., WANG, R., RISCH, H., MUELLER, L. and GUAN, Y. (2014). A new estimation approach for combining epidemiological data from multiple sources. *J. Amer. Statist. Assoc.* **109** 11–23. MR3180543 https://doi.org/10.1080/01621459.2013.870904

HUSEN, S., KISSLING, E., DEICHMANN, N., WIEMER, S., GIARDINI, D. and BAER, M. (2003). Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland. *J. Geophys. Res.* **108**.

ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. *Statistics in Practice*. Wiley, Chichester. MR2384630

JACQUEZ, G. M. (2012). A research agenda: Does geocoding positional error matter in health GIS studies? *Spat. Spatiotemporal Epidemiol.* **3** 7–16. https://doi.org/10.1016/j.sste.2012.02.002

JENSEN, L. G., WILLIAMSON, D. J. and HAHN, U. (2022). Semiparametric point process modeling of blinking artifacts in PALM. *Ann. Appl. Stat.* **16** 1500–1523. MR4455890 https://doi.org/10.1214/21-aoas1553

JOHNSON, D. S., HOOTEN, M. B. and KUHN, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *J. Anim. Ecol.* **82** 1155–64.

JONES-TODD, C. M., CAIE, P., ILLIAN, J. B., STEVENSON, B. C., SAVAGE, A., HARRISON, D. J. and BOWN, J. L. (2019). Identifying prognostic structural features in tissue sections of colon cancer patients using point pattern analysis. *Stat. Med.* **38** 1421–1441. MR3920624 https://doi.org/10.1002/sim.8046

JONES-TODD, C. M., SWALLOW, B., ILLIAN, J. B. and TOMS, M. (2018). A spatiotemporal multispecies model of a semicontinuous response. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 705–722. MR3787973 https://doi.org/10.1111/rssc.12250

KAGAN, Y. Y. and KNOPOFF, L. (1980). Spatial distribution of earthquakes: The two point correlation function. *Geophys. J. R. Astron. Soc.* **62** 303–320.

KANSANEN, K., PACKALEN, P., MALTAMO, M. and MEHTÄTALO, L. (2021). Horvitz-Thompson-like estimation with distance-based detection probabilities for circular plot sampling of forests. *Biometrics* **77** 715–728. MR4307667 https://doi.org/10.1111/biom.13312

KANSANEN, K., VAUHKONEN, J., LÄHIVAARA, T. and MEHTÄTALO, L. (2016). Stand density estimators based on individual tree detection and stochastic geometry. *Can. J. For. Res.* **46** 1359–1366.

KING, I., STANFORD, S., SEITZER, P., BERSHADY, M., KEEL, W., KOO, D., WEIR, N., DJORGOVSKI, S. and WINDHORST, R. (1991). The current ability of HST to reveal morphological structure in medium-redshift galaxies. *Astron. J.* **102** 1553–1568.

KONSTANTINOU, K. and SÄRKKÄ, A. (2021). Spatial modeling of epidermal nerve fiber patterns. *Stat. Med.* **40** 6479–6500. MR4352750 https://doi.org/10.1002/sim.9194

KOSTENSALO, J., MEHTÄTALO, L., TUOMINEN, S., PACKALEN, P. and MYLLYMÄKI, M. (2023). Recreating structurally realistic tree maps with airborne laser scanning and ground measurements. *Remote Sens. Environ.* **298** 113782.

KURONEN, M., HENTTONEN, H. M. and MYLLYMÄKI, M. (2019). Correcting for nondetection in estimating forest characteristics from single-scan terrestrial laser measurements. *Can. J. For. Res.* **46** 96–103.

KURONEN, M., MYLLYMÄKI, M., LOAVENBRUCK, A. and SÄRKKÄ, A. (2021). Point process models for sweat gland activation observed with noise. *Stat. Med.* **40** 2055–2072. MR4229838 https://doi.org/10.1002/sim.8891

KURONEN, M., SÄRKKÄ, A., VIHOLA, M. and MYLLYMÄKI, M. (2022). Hierarchical log Gaussian Cox process for regeneration in uneven-aged forests. *Environ. Ecol. Stat.* **40** 185–205.

LAAKE, J. and BORCHERS, D. (2004). Methods for incomplete detection at distance zero. In *Advanced Distance Sampling* 108–189. Oxford Univ. Press, London.

LAHIVAARA, T., SEPPANEN, A., KAIPIO, J. P., VAUHKONEN, J., KORHONEN, L., TOKOLA, T. and MALTAMO, M. (2014). Bayesian approach to tree detection based on airborne laser scanning data. *IEEE Trans. Geosci. Remote Sens.* **52** 2690–2699.

LANGLEY, R. B. (1997). The GPS error budget. *GPS World* **8** 51–56.

LARSEN, M. and RUDEMO, M. (1998). Optimizing templates for finding trees in aerial photographs. *Pattern Recogn. Lett.* **19** 1153–1162.

LAW, R., ILLIAN, J., BURSLEM, D. F., GRATZER, G., GUNATILLEKE, C. and GUNATILLEKE, I. (2009). Ecological information from spatial patterns of plants: Insights from point process theory. *J. Ecol.* **97** 616–628.

LAWSON, A. (1993). A deviance residual for heterogeneous spatial Poisson processes. *Biometrics* **49** 889–897.

LEE, W. H. K. and BRILLINGER, D. R. (1979). On Chinese earthquake history–an attempt to model an incomplete data set by point process analysis. *Pure Appl. Geophys.* **117** 1229–1257.

LUND, J., PENTTINEN, A. and RUDEMO, M. (1999). Bayesian analysis of spatial point patterns from noisy observations. In *Statistical Inference and Perfect Simulation for Point Processes Observed with Noise* (*PhD Thesis of J. Lund*). Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Denmark.

LUND, J. and RUDEMO, M. (2000). Models for point processes observed with noise. *Biometrika* **87** 235–249. MR1782476 https://doi.org/10.1093/biomet/87.2.235

LUND, J. and THÖNNES, E. (2004). Perfect simulation and inference for point processes given noisy observations. *Comput. Statist.* **19** 317–336. MR2066310 https://doi.org/10.1007/BF02892063

MARTENS, A. E. (1968). Errors in measurement and counting of particles using light scattering. *J. Air Pollut. Control Assoc.* **18** 661–663. https://doi.org/10.1080/00022470.1968.10469188

MARTINO, S., PACE, D. S., MORO, S., CASOLI, E., VENTURA, D., FRACHEA, A., SILVESTRI, M., ARCANGELI, A., GIACOMINI, G. et al. (2021). Integration of presence-only data from several sources: A case study on dolphins' spatial distribution. *Ecography* **44** 1533–1543.

MEHTÄTALO, L. (2006). Eliminating the effect of overlapping crowns from aerial inventory estimates. *Can. J. For. Res.* **36** 1649–1660.

MEHTÄTALO, L., YAZIGI, A., KANSANEN, K., PACKALEN, P., LÄHIVAARA, T., MALTAMO, M., MYLLYMÄKI, M. and PENTTINEN, A. (2022). Estimation of forest stand characteristics using individual tree detection, stochastic geometry and a sequential spatial point process model. *Int. J. Appl. Earth Obs. Geoinf.* **112** 102920.

MENKE, W. (1989). *Geophysical Data Analysis*: *Discrete Inverse Theory*. Academic Press, Amsterdam.

MILLER, D. L., FIFIELD, D., WAKEFIELD, E. and SIGOURNEY, D. B. (2021). Extending density surface models to include multiple and double-observer survey data. *PeerJ* **9** e12113. https://doi.org/10.7717/peerj.12113

MILNE, S., MARTIN, J. G., REYNOLDS, G., VAIRAPPAN, C. S., SLADE, E. M., BRODIE, J. F., WICH, S. A., WILLIAMSON, N. and BURSLEM, D. F. (2021). Drivers of Bornean orangutan distribution across a multiple-use tropical landscape. *Remote Sens.* **13** 458.

MOLINA, R. and RIPLEY, B. D. (1989). Using spatial models as priors in astronomical image analysis. *J. Appl. Stat.* **16** 193–206.

MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. *Monographs on Statistics and Applied Probability* **100**. CRC Press/CRC, Boca Raton, FL. MR2004226

MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Stat.* **34** 643–684. MR2392447 https://doi.org/10.1111/j.1467-9469.2007.00569.x

MYLLYMÄKI, M., KURONEN, M. and MRKVIČKA, T. (2021). Testing global and local dependence of point patterns on covariates in parametric models. *Spat. Stat.* **42** Paper No. 100436. MR4233257 https://doi.org/10.1016/j.spasta.2020.100436

MYLLYMÄKI, M. and MRKVIČKA, T. (2023). GET: Global envelopes in R. Available at arXiv:1911.06583 [stat.ME].

MYLLYMÄKI, M., MRKVIČKA, T., GRABARNIK, P., SEIJO, H. and HAHN, U. (2017). Global envelope tests for spatial processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 381–404. MR3611751 https://doi.org/10.1111/rssb.12172

NORDIN, M., ABRAHAMSSON, C., HAMNGREN BLOMQVIST, C., RÖDING, M., OLSSON, E., NYDÉN, M. and RUDEMO, M. (2014). Estimation of mass thickness response of embedded aggregated silica nanospheres from high angle annular dark-field scanning transmission electron micrographs. *J. Microsc.* **253** 166–170.

OGATA, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50** 379–402.

OGATA, Y. and ZHUANG, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics* **413** 13–23.

OLSBO, V., MYLLYMÄKI, M., WALLER, L. A. and SÄRKKÄ, A. (2013). Development and evaluation of spatial point process models for epidermal nerve fibers. *Math. Biosci.* **243** 178–189. MR3065205 https://doi.org/10.1016/j.mbs.2013.03.001

PACKALEN, P., VAUHKONEN, J., KALLIO, E., PEUHKURINEN, J., PITKÄNEN, J., PIPPURI, I., STRUNK, J. and MALTAMO, M. (2013). Predicting the spatial pattern of trees by airborne laser scanning. *Int. J. Remote Sens.* **34** 5154–5165.

PALKA, D. and HAMMOND, P. (2001). Accounting for responsive movement in line transect estimates of abundance. *Can. J. Fish. Aquat. Sci.* **58** 777–787.

PIPPURI, I., KALLIO, E., MALTAMO, M., PELTOLA, H. and PACKALÉN, P. (2012). Exploring horizontal area-based metrics to discriminate the spatial pattern of trees and need for first thinning using airborne laser scanning. *Forestry* **85** 305–314.

PITKÄNEN, T. P., RAUMONEN, P. and KANGAS, A. (2019). Measuring stem diameters with TLS in boreal forests by complementary fitting procedure. *ISPRS J. Photogramm. Remote Sens.* **147** 294–306.

PITKÄNEN, T. P., RAUMONEN, P., LIANG, X., LEHTOMÄKI, M. and KANGAS, A. (2021). Improving TLS-based stem volume estimates by field measurements. *Comput. Electron. Agric.* **180** 105882.

RAJALA, T., REDENBACH, C., SÄRKKÄ, A. and SORMANI, M. (2016). Variational Bayes approach for classification of points in superpositions of point processes. *Spat. Stat.* **15** 85–99. MR3457670 https://doi.org/10.1016/j.spasta.2015.12.001

RANACHER, P., BRUNAUER, R., TRUTSCHNIG, W., DER SPEK, S. V. and REICH, S. (2016). Why GPS makes distances bigger than they are. *Int. J. Geogr. Inf. Sci.* **30** 316–333. https://doi.org/10.1080/13658816.2015.1086924

REDENBACH, C., SÄRKKÄ, A., FREITAG, J. and SCHLADITZ, K. (2009). Anisotropy analysis of pressed point processes. *AStA Adv. Stat. Anal.* **93** 237–261. MR2545694 https://doi.org/10.1007/s10182-009-0106-5

REDENBACH, C., SÄRKKÄ, A. and SORMANI, M. (2015). Classification of points in superpositions of Strauss and Poisson processes. *Spat. Stat.* **12** 81–95. MR3346643 https://doi.org/10.1016/j.spasta.2015.03.003

REGENER, E. (1909). Über Zählung der α-Teilchen durch die Szintillation und über die Größe des elektrischen Elementarquantums. *Sitzungsber. Königlich Preuss. Akad.* 948–965.

RENNER, I. W., ELITH, J., BADDELEY, A., FITHIAN, W., HASTIE, T., PHILLIPS, S. J., POPOVIC, G. and WARTON, D. I. (2015). Point process models for presence-only analysis. *Methods Ecol. Evol.* **6** 366–379.

RENNER, I. W., LOUVRIER, J. and GIMENEZ, O. (2019). Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization. *Methods Ecol. Evol.* **10** 2118–2128.

RODRIGUEZ-ITURBE, I., COX, D. R. and ISHAM, V. (1987). Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. Ser. A* **410** 269–288. MR0887878

ROUSSEL, J.-R. and AUTY, D. (2022). lidR: Airborne LiDAR data manipulation and visualization for forestry applications. R package version 4.0.1.

ROUSSEL, J.-R., AUTY, D., COOPS, N. C., TOMPALSKI, P., GOODBODY, T. R. H., MEADOR, A. S., BOURDON, J.-F., DE BOISSIEU, F. and ACHIM, A. (2020). lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **251** 112061.

ROYLE, J. A. and LINK, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* **87** 835–841. https://doi.org/10.1890/0012-9658(2006)87[835:gsomaf]2.0.co;2

RUTHERFORD, E., GEIGER, H. and BATEMAN, H. (1910). LXXVI. The probability variations in the distribution of α-particles. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **20** 698–707.

SALBERG, A.-B. (2015). Detection of seals in remote sensing images using features extracted from deep convolutional neural networks. In 2015 *IEEE International Geoscience and Remote Sensing Symposium* (*IGARSS*) 1893–1896. IEEE Press, New York.

SCHMIDT, J. H., THOMPSON, W. L., WILSON, T. L. and REYNOLDS, J. H. (2022). Distance sampling surveys: Using components of detection and total error to select among approaches. *Wildl. Monogr.* **210** e1070.

SCHNEIDER, M. (2021). Improving Uncertainty Quantification and Visualization for Spatiotemporal Earthquake Rate Models for the Pacific Northwest. PhD thesis, Dept. Statistics, Univ. Washington, United States.

SICACHA-PARADA, J., STEINSLAND, I., CRETOIS, B. and BORGELT, J. (2021). Accounting for spatial varying sampling effort due to accessibility in citizen science data: A case study of moose in Norway. *Spat. Stat.* **42** Paper No. 100446. MR4233261 https://doi.org/10.1016/j.spasta.2020.100446

STOYAN, D. (1982). A remark on the line transect method. *Biom. J.* **24** 191–195. MR0662894 https://doi.org/10.1002/bimj.4710240210

STOYAN, D. and GRABARNIK, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math. Nachr.* **151** 95–100. MR1121200 https://doi.org/10.1002/mana.19911510108

STOYAN, D. and PENTTINEN, A. (2000). Recent applications of point process methods in forestry statistics. *Statist. Sci.* **15** 61–78. MR1842237 https://doi.org/10.1214/ss/1009212674

STOYAN, D. and STOYAN, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, Chichester. MR1297125

THÀNH, N. T., HÀO, D. N. and SAHLI, H. M. (2009). Infrared thermography for land mine detection. In *Augmented Vision Perception in Infrared: Algorithms and Applied Systems* (R. I. Hammoud, ed.) 3–36. Springer.

VAN LIESHOUT, M. N. M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London. MR1789230 https://doi.org/10.1142/9781860949760

VOSSELMAN, G. and MAAS, H. G. (2010). *Airborne and Terrestrial Laser Scanning*. Whittles Publishing, Dunbeath.

WAAGEPETERSEN, R. and SCHWEDER, T. (2006). Likelihood-based inference for clustered line transect data. *J. Agric. Biol. Environ. Stat.* **11** 264–279.

WARTON, D. I., RENNER, I. W. and RAMP, D. (2013). Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS ONE* **8** e79168. https://doi.org/10.1371/journal.pone.0079168

WATSON, J., JOY, R., TOLLIT, D., THORNTON, S. J. and AUGER-MÉTHÉ, M. (2021). Estimating animal utilization distributions from multiple data types: A joint spatiotemporal point process framework. *Ann. Appl. Stat.* **15** 1872–1896. MR4355080 https://doi.org/10.1214/21-aoas1472

WILLIAMS, D. M., DECHEN QUINN, A. and PORTER, W. F. (2012). Impact of habitat-specific GPS positional error on detection of movement scales by first-passage time analysis. *PLoS ONE* **7** e48439. https://doi.org/10.1371/journal.pone.0048439

WILLIAMSON, L. D., SCOTT, B. E., LAXTON, M., ILLIAN, J. B., TODD, V. L. G., MILLER, P. I. and BROOKES, K. L. (2022). Comparing distribution of harbour porpoise using generalized additive models and hierarchical Bayesian models with integrated nested Laplace approximation. *Ecol. Model.* **470** 110011.

YUAN, Y., BACHL, F. E., LINDGREN, F., BORCHERS, D. L., ILLIAN, J. B., BUCKLAND, S. T., RUE, H. and GERRODETTE, T. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Ann. Appl. Stat.* **11** 2270–2297. MR3743297 https://doi.org/10.1214/17-AOAS1078

ZHANG, H., CHON, C. H., PAN, X. and LI, D. (2009). Methods for counting particles in microfluidic applications. *Microfluid. Nanofluid.* **7** 739–749.

ZHANG, W., CHIPPERFIELD, J. D., ILLIAN, J. B., DUPONT, P., MILLERET, C., DE VALPINE, P. and BISCHOF, R. (2023). A flexible and efficient Bayesian implementation of point process models for spatial capture-recapture data. *Ecology* **104** e3887. https://doi.org/10.1002/ecy.3887

ZIMMERMAN, D. L., LI, J. and FANG, X. (2010). Spatial autocorrelation among automated geocoding errors and its effects on testing for disease clustering. *Stat. Med.* **29** 1025–1036. MR2757131 https://doi.org/10.1002/sim.3836

Statistical Science
2026, Vol. 41, No. 1, 163–187
https://doi.org/10.1214/24-STS957

A Bayesian Practitioner's Guide to Expectation Propagation

Jackson Zhou, Clara Grazian and John T. Ormerod

Abstract. In the field of approximate Bayesian inference, expectation propagation (EP) is an often overlooked counterpart to its older Laplace approximation and variational Bayes cousins, perhaps owing to a lack of theory (especially convergence guarantees) and a higher implementation overhead, where derivations need to be hand-crafted to the specific distributions being approximated. However, when EP is carefully implemented, it is often more accurate than these alternative approaches. With this in mind, the purpose of this review paper is to describe and to consolidate the current state of research on EP at a high level focusing on examples and applications. Our aim is for this broad-based, practical EP guide to encourage its novel application to both existing and new Bayesian problems, where it has the opportunity to dramatically extend the cutting edge in performance.

Key words and phrases: Approximate Bayesian inference, expectation propagation, message passing.

REFERENCES

ANCESCHI, N., FASANO, A., FRANZOLINI, B. and REBAUDO, G. (2024). Scalable expectation propagation for generalized linear models. Preprint. Available at arXiv:2407.02128.

ANDERSON, J. R. and PETERSON, C. (1987). A mean field theory learning algorithm for neural networks. *Complex Systems* **1** 995–1019.

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

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

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

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.

CAKMAK, B. and OPPER, M. (2018). Expectation propagation for approximate inference: Free probability framework. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 1276–1280. IEEE.

CESPEDES, J., OLMOS, P. M., SÁNCHEZ-FERNÁNDEZ, M. and PEREZ-CRUZ, F. (2014). Expectation propagation detection for high-order high-dimensional MIMO systems. *IEEE Trans. Commun.* **62** 2840–2849.

CÉSPEDES, J., OLMOS, P. M., SÁNCHEZ-FERNÁNDEZ, M. and PEREZ-CRUZ, F. (2014). Improved performance of LDPC-coded MIMO systems with EP-based soft-decisions. In 2014 *IEEE International Symposium on Information Theory* 1997–2001. IEEE.

CHOPIN, N. and RIDGWAY, J. (2017). Leave Pima Indians alone: Binary regression as a benchmark for Bayesian computation. *Statist. Sci.* **32** 64–87. MR3634307 https://doi.org/10.1214/16-STS581

CUI, L., WANG, S. and CHENG, S. (2012). Online SNR statistic estimation for LDPC decoding over AWGN channel using Laplace propagation. In 2012 *IEEE Global Communications Conference* (*GLOBECOM*) 3743–3747. IEEE.

CUI, L., WANG, S., CHENG, S., STANKOVIC, L. and STANKOVIC, V. (2012). Adaptive Slepian–Wolf decoding using Laplace propagation. In 2012 *Proceedings of the* 20*th European Signal Processing Conference* (*EUSIPCO*) 564–568. IEEE.

CUNNINGHAM, J. P., HENNIG, P. and LACOSTE-JULIEN, S. (2011). Gaussian probabilities and expectation propagation. Preprint. Available at arXiv:1111.6832.

DEHAENE, G. and BARTHELMÉ, S. (2018). Expectation propagation in the large data limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 199–217. MR3744718 https://doi.org/10.1111/rssb.12241

DEHAENE, G. P. (2016). Expectation propagation performs a smoothed gradient descent. Preprint. Available at arXiv:1612.05053.

Jackson Zhou is a Ph.D. candidate at the School of Mathematics and Statistics, University of Sydney, Sydney, Australia (e-mail: jzho4727@uni.sydney.edu.au). Clara Grazian is a senior lecturer at the School of Mathematics and Statistics, University of Sydney, Sydney, Australia and Chief Investigator of the ARC Training Centre in Data Analytics for Resources and Environments (DARE), Sydney, Australia (e-mail: clara.grazian@sydney.edu.au). John T. Ormerod is an Associate Professor at the School of Mathematics and Statistics, University of Sydney, Sydney, Australia (e-mail: john.ormerod@sydney.edu.au).

DEHAENE, G. P. and BARTHELMÉ, S. (2015). Bounding errors of expectation-propagation. *Adv. Neural Inf. Process. Syst.* **28**.

DRAPER, D. (1993). Challenger USA Space Shuttle O-Ring. UCI Machine Learning Repository. Available at https://doi.org/10.24432/C5PW2T.

ESKIN, E., SMOLA, A. and VISHWANATHAN, S. (2003). Laplace propagation. *Adv. Neural Inf. Process. Syst.* **16**.

ESLAMI, S., TARLOW, D., KOHLI, P. and WINN, J. (2014). Just-in-time learning for fast and flexible inference. *Adv. Neural Inf. Process. Syst.* **27**.

FAES, C., ORMEROD, J. T. and WAND, M. P. (2011). Variational Bayesian inference for parametric and nonparametric regression with missing data. *J. Amer. Statist. Assoc.* **106** 959–971. MR2894756 https://doi.org/10.1198/jasa.2011.tm10301

FREY, B. J. and MACKAY, D. (1997). A revolution: Belief propagation in graphs with cycles. *Adv. Neural Inf. Process. Syst.* **10**.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2020). *Bayesian Data Analysis*, 2nd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR2027492

GHAVAMI, K. and NARAGHI-POUR, M. (2017). Blind channel estimation and symbol detection for multi-cell massive MIMO systems by expectation propagation. *IEEE Trans. Wirel. Commun.* **17** 943–954.

GONZÁLEZ, J., OSBORNE, M. and LAWRENCE, N. (2016). GLASSES: Relieving the myopia of Bayesian optimisation. In *Artificial Intelligence and Statistics* 790–799. PMLR.

GRIFFITHS, R.-R. and HERNÁNDEZ-LOBATO, J. M. (2020). Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science* **11** 577–586.

GUIVER, J. and SNELSON, E. (2009). Bayesian inference for Plackett-Luce ranking models. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 377–384.

HASENCLEVER, L., WEBB, S., LIENART, T., VOLLMER, S., LAKSHMINARAYANAN, B., BLUNDELL, C. and TEH, Y. W. (2017). Distributed Bayesian learning with stochastic natural gradient expectation propagation and the posterior server. *J. Mach. Learn. Res.* **18** Paper No. 106, 37 pp. MR3725445

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

HEESS, N., TARLOW, D. and WINN, J. (2013). Learning to pass expectation propagation messages. *Adv. Neural Inf. Process. Syst.* **26**.

HERNÁNDEZ-LOBATO, D. and HERNÁNDEZ-LOBATO, J. M. (2016). Scalable Gaussian process classification via expectation propagation. In *Artificial Intelligence and Statistics* 168–176. PMLR.

HERNANDEZ-LOBATO, J., LI, Y., ROWLAND, M., BUI, T., HERNÁNDEZ-LOBATO, D. and TURNER, R. (2016). Black-box alpha divergence minimization. In *International Conference on Machine Learning* 1511–1520. PMLR.

HESKES, T., OPPER, M., WIEGERINCK, W., WINTHER, O. and ZOETER, O. (2005). Approximate inference techniques with expectation constraints. *J. Stat. Mech. Theory Exp.* **2005** P11015.

HESKES, T. and ZOETER, O. (2002). Expectation propogation for approximate inference in dynamic Bayesian networks. Preprint. Available at arXiv:1301.0572.

HESTER, J. and VAUGHAN, D. (2021). bench: High precision timing of R expressions. R package version 1.1.2.

ILLIAN, J. B., SØRBYE, S. H. and RUE, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **6** 1499–1530. MR3058673 https://doi.org/10.1214/11-AOAS530

JITKRITTUM, W., GRETTON, A., HEESS, N., ESLAMI, S., LAKSHMINARAYANAN, B., SEJDINOVIC, D. and SZABÓ, Z. (2015). Kernel-based just-in-time learning for passing expectation propagation messages. Preprint. Available at arXiv:1503.02551.

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407. MR1394097 https://doi.org/10.2307/2291420

JYLÄNKI, P., VANHATALO, J. and VEHTARI, A. (2011). Robust Gaussian process regression with a Student-t likelihood. *J. Mach. Learn. Res.* **12** 3227–3257. MR2877599

KIM, A. S. I. and WAND, M. P. (2016). The explicit form of expectation propagation for a simple statistical model. *Electron. J. Stat.* **10** 550–581. MR3471988 https://doi.org/10.1214/16-EJS1114

KIM, A. S. I. and WAND, M. P. (2018). On expectation propagation for generalised, linear and mixed models. *Aust. N. Z. J. Stat.* **60** 75–102. MR3780623 https://doi.org/10.1111/anzs.12199

KINGMA, D. P. and WELLING, M. (2013). Auto-encoding variational Bayes. Preprint. Available at arXiv:1312.6114.

LAURITZEN, S. L. (1992). Propagation of probabilities, means, and variances in mixed graphical association models. *J. Amer. Statist. Assoc.* **87** 1098–1108. MR1209568

LI, X., LI, C., CHI, J. and OUYANG, J. (2022). Approximate posterior inference for Bayesian models: Black-box expectation propagation. *Knowl. Inf. Syst.* **64** 2361–2387.

LI, Y., HERNÁNDEZ-LOBATO, J. M. and TURNER, R. E. (2015). Stochastic expectation propagation. *Adv. Neural Inf. Process. Syst.* **28**.

LINDLEY, D. V. (1980). Approximate Bayesian methods. *Trab. Estad. Investig. Oper.* **31** 223–245.

MANOUCHEHRI, N. and BOUGUILA, N. (2021). Stochastic expectation propagation learning of infinite multivariate beta mixture models for human tissue analysis. In *IECON* 2021–47*th Annual Conference of the IEEE Industrial Electronics Society* 1–6. IEEE.

MAYBECK, P. S. (1982). *Stochastic Models, Estimation, and Control. Vol.* 3. *Mathematics in Science and Engineering* **141**. Academic Press, London. MR0690418

MINKA, T. (2004). Power EP. Technical report, Microsoft Research, Cambridge.

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

MINKA, T. P. (2001a). A family of algorithms for approximate Bayesian inference. Ph.D. thesis, Massachusetts Institute of Technology. MR2717007

MINKA, T. P. (2001b). The EP energy function and minimization schemes. Technical report. Microsoft Research, Cambridge.

MINKA, T. P. (2005). Divergence measures and message passing. Technical report, Microsoft Research, Cambridge.

MINKA, T. P. and LAFFERTY, J. (2012). Expectation-propogation for the generative aspect model. Preprint. Available at arXiv:1301.0588.

MUFF, S., RIEBLER, A., HELD, L., RUE, H. and SANER, P. (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 231–252. MR3302298 https://doi.org/10.1111/rssc.12069

MUÑOZ-GONZÁLEZ, L., LÁZARO-GREDILLA, M. and FIGUEIRAS-VIDAL, A. R. (2011). Heteroscedastic Gaussian process regression using expectation propagation. In 2011 *IEEE International Workshop on Machine Learning for Signal Processing* 1–6. IEEE.

OPPER, M. and WINTHER, O. (2000). Gaussian processes for classification: Mean-field algorithms. *Neural Comput.* **12** 2655–2684. https://doi.org/10.1162/089976600300014881

OPPER, M., WINTHER, O. and JORDAN, M. J. (2005). Expectation consistent approximate inference. *J. Mach. Learn. Res.* **6**.

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

OWEN, D. B. (1980). A table of normal integrals. *Comm. Statist. B—Simulation Comput.* **9** 389–419. MR0570844 https://doi.org/10.1080/03610918008812164

PARK, M., BOHNER, G. and MACKE, J. H. (2015). Unlocking neural population non-stationarities using hierarchical dynamics models. *Adv. Neural Inf. Process. Syst.* **28**.

QI, Y., MINKA, T. P., PICARD, R. W. and GHAHRAMANI, Z. (2004). Predictive automatic relevance determination by expectation propagation. In *Proceedings of the Twenty-First International Conference on Machine Learning* 85.

R CORE TEAM (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

RAYMOND, J., MANOEL, A. and OPPER, M. (2016). Expectation propagation. Preprint. Available at arXiv:1409.6179.

RIBEIRO, F. and OPPER, M. (2011). Expectation propagation with factorizing distributions: A Gaussian approximation and performance results for simple models. *Neural Comput.* **23** 1047–1069. MR2840638 https://doi.org/10.1162/NECO_a_00104

RITTER, H., BOTEV, A. and BARBER, D. (2018). A scalable Laplace approximation for neural networks. In 6*th International Conference on Learning Representations*, *ICLR* 2018-*Conference Track Proceedings* **6**. International Conference on Representation Learning.

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

RUIZ-CÁRDENAS, R., KRAINSKI, E. T. and RUE, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations—INLA. *Comput. Statist. Data Anal.* **56** 1808–1828. MR2892379 https://doi.org/10.1016/j.csda.2011.10.024

SALAMANCA, L., OLMOS, P. M., MURILLO-FUENTES, J. J. and PÉREZ-CRUZ, F. (2013). Tree expectation propagation for ML decoding of LDPC codes over the BEC. *IEEE Trans. Commun.* **61** 465–473.

SANTOS, I., MURILLO-FUENTES, J. J., BOLOIX-TORTOSA, R., ARIAS-DE REYNA, E. and OLMOS, P. M. (2016). Expectation propagation as turbo equalizer in ISI channels. *IEEE Trans. Commun.* **65** 360–370.

SEEGER, M. (2005). Expectation propagation for exponential families. Technical report, LAPMAL.

SEEGER, M. and NICKISCH, H. (2011). Fast convergent algorithms for expectation propagation approximate Bayesian inference. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* 652–660. JMLR Workshop and Conference Proceedings.

SEEGER, M., STEINKE, F. and TSUDA, K. (2007). Bayesian inference and optimal design in the sparse linear model. In *Artificial Intelligence and Statistics* 444–451. PMLR.

SENST, M. and ASCHEID, G. (2011). How the framework of expectation propagation yields an iterative IC-LMMSE MIMO receiver. In 2011 *IEEE Global Telecommunications Conference-GLOBECOM* 2011 1–6. IEEE.

SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690. MR1125725

STAN DEVELOPMENT TEAM (2023). RStan: The R interface to Stan. R package version 2.21.8.

SUN, S. and HE, S. (2019). Generalizing expectation propagation with mixtures of exponential family distributions and an application to Bayesian logistic regression. *Neurocomputing* **337** 180–190.

SWAROOP, S. and TURNER, R. E. (2017). Understanding expectation propagation. In *Advances in Approximate Bayesian Inference Workshop at NIPS*.

TEH, Y. and WELLING, M. (2001). The unified propagation and scaling algorithm. *Adv. Neural Inf. Process. Syst.* **14**.

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

TIPPING, M. (1999). The relevance vector machine. *Adv. Neural Inf. Process. Syst.* **12**.

TITTERINGTON, D. M. (2011). The EM algorithm, variational approximations and expectation propagation for mixtures. In *Mixtures*: *Estimation and Applications*. *Wiley Ser. Probab. Stat.* 1–29. Wiley, Chichester. MR2883348 https://doi.org/10.1002/9781119995678.ch1

TOLVANEN, V., JYLÄNKI, P. and VEHTARI, A. (2014). Expectation propagation for nonstationary heteroscedastic Gaussian process regression. In 2014 *IEEE International Workshop on Machine Learning for Signal Processing* (*MLSP*) 1–6. IEEE.

VEHTARI, A., GELMAN, A., SIVULA, T., JYLÄNKI, P., TRAN, D., SAHAI, S., BLOMSTEDT, P., CUNNINGHAM, J. P., SCHIMINOVICH, D. et al. (2020). Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *J. Mach. Learn. Res.* **21** Paper No. 17, 53 pp. MR4071200

VILLACAMPA-CALVO, C. and HERNANDEZ-LOBATO, D. (2020). Alpha divergence minimization in multi-class Gaussian process classification. *Neurocomputing* **378** 210–227.

WAINWRIGHT, M. J., JORDAN, M. I. et al. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

WIEGERINCK, W. and HESKES, T. (2002). Fractional belief propagation. *Adv. Neural Inf. Process. Syst.* **15**.

WIPF, D. and NAGARAJAN, S. (2007). A new view of automatic relevance determination. *Adv. Neural Inf. Process. Syst.* **20**.

XU, M., LAKSHMINARAYANAN, B., TEH, Y. W., ZHU, J. and ZHANG, B. (2014). Distributed Bayesian posterior sampling via moment sharing. *Adv. Neural Inf. Process. Syst.* **27**.

YEDIDIA, J. S., FREEMAN, W. T. and WEISS, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inf. Theory* **51** 2282–2312. MR2246363 https://doi.org/10.1109/TIT.2005.850085

YPMA, A. and HESKES, T. (2003). Iterated extended Kalman smoothing with expectation-propagation. In 2003 *IEEE XIII Workshop on Neural Networks for Signal Processing* (*IEEE Cat. No.* 03*TH*8718) 219–228. IEEE.

YPMA, A. and HESKES, T. (2005). Novel approximations for inference in nonlinear dynamical systems using expectation propagation. *Neurocomputing* **69** 85–99.

YUILLE, A. L. (2002). CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation. *Neural Comput.* **14** 1691–1722. https://doi.org/10.1162/08997660260028674

ZHANG, D., MENDES, L. L., MATTHÉ, M., GASPAR, I. S., MICHAILOW, N. and FETTWEIS, G. P. (2015). Expectation propagation for near-optimum detection of MIMO-GFDM signals. *IEEE Trans. Wirel. Commun.* **15** 1045–1062.

ZHE, S., LEE, K.-C., ZHANG, K. and NEVILLE, J. (2016). Online spike-and-slab inference with stochastic expectation propagation. NIPS WS.

ZHOU, J., GRAZIAN, C. and ORMEROD, J. T. (2026). Supplement to "A Bayesian Practitioner's Guide to Expectation Propagation." https://doi.org/10.1214/24-STS957SUPP

Statistical Science
2026, Vol. 41, No. 1, 188–200
https://doi.org/10.1214/24-STS958

Markov Chain Monte Carlo Significance Tests

Michael Howes

Abstract. Monte Carlo significance tests are a general tool that produce p-values by generating samples from the null distribution. However, Monte Carlo tests are limited to null hypothesis from which we can exactly sample. Markov chain Monte Carlo (MCMC) significance tests are a way to produce statistical valid p-values for null hypothesis we can only approximately sample from. These methods were first introduced by Besag and Clifford in 1989 and make no assumptions on the mixing time of the MCMC procedure. Here we review the two methods of Besag and Clifford and introduce a new method that unifies the existing procedures. We use simple examples to highlight the difference between MCMC significance tests and standard Monte Carlo tests based on exact sampling. We also survey a range of contemporary applications in the literature including goodness-of-fit testing for the Rasch model, tests for detecting gerrymandering (*Proc. Natl. Acad. Sci. USA* **114** (2017) 2860–2864) and a permutation based test of conditional independence (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** (2020) 175–197).

Key words and phrases: Hypothesis testing, p-values, Markov chain Monte Carlo.

REFERENCES

[1] ANDERSEN, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika* **38** 123–140. MR0311064 https://doi.org/10.1007/BF02291180

[2] BARBER, R. F. and JANSON, L. (2022). Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling. *Ann. Statist.* **50** 2514–2544. MR4500617 https://doi.org/10.1214/22-aos2187

[3] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981 https://doi.org/10.1111/rssb.12340

[4] BESAG, J. and CLIFFORD, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76** 633–642. MR1041408 https://doi.org/10.1093/biomet/76.4.633

[5] BRÉMAUD, P. (2020). *Markov Chains—Gibbs Fields, Monte Carlo Simulation and Queues*. *Texts in Applied Mathematics* **31**. Springer, Cham. MR4174390 https://doi.org/10.1007/978-3-030-45982-6

[6] CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

[7] CHEN, Y., DIACONIS, P., HOLMES, S. P. and LIU, J. S. (2005). Sequential Monte Carlo methods for statistical analysis of tables. *J. Amer. Statist. Assoc.* **100** 109–120. MR2156822 https://doi.org/10.1198/016214504000001303

[8] CHIKINA, M., FRIEZE, A., MATTINGLY, J. C. and PEGDEN, W. (2020). Separating effect from significance in Markov chain tests. *Stat. Public Policy* **7** 101–114. https://doi.org/10.1080/2330443X.2020.1806763

[9] CHIKINA, M., FRIEZE, A. and PEGDEN, W. (2017). Assessing significance in a Markov chain without mixing. *Proc. Natl. Acad. Sci. USA* **114** 2860–2864. MR3628186 https://doi.org/10.1073/pnas.1617540114

[10] CHIKINA, M., FRIEZE, A. and PEGDEN, W. (2019). Understanding our Markov chain significance test: A reply to cho and Rubinstein-salzedo. *Stat. Public Policy* **6** 50–53. https://doi.org/10.1080/2330443X.2019.1615396

[11] CHO, W. K. T. and RUBINSTEIN-SALZEDO, S. (2019). Understanding significance tests from a non-mixing Markov chain for partisan gerrymandering claims. *Stat. Public Policy* **6** 44–49. https://doi.org/10.1080/2330443X.2019.1574687

[12] DIACONIS, P. and STURMFELS, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156 https://doi.org/10.1214/aos/1030563990

[13] DICICCIO, T. J., RITZWOLLER, D. M., ROMANO, J. P. and SHAIKH, A. M. (2022). Confidence intervals for seroprevalence. *Statist. Sci.* **37** 306–321. MR4444369 https://doi.org/10.1214/21-sts844

[14] DUFOUR, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics* **133** 443–477. MR2252905 https://doi.org/10.1016/j.jeconom.2005.06.007

[15] EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**.

Michael Howes is a graduate student in the Department of Statistics, Stanford University, Stanford, California 94305-4020, USA (e-mail: mhowes@stanford.edu).

CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

[16] HOWES, M. (2026). Supplement to "Markov Chain Monte Carlo Significance Tests." https://doi.org/10.1214/24-STS958SUPP

[17] MAGIS, D., BELAND, S., TUERLINCKX, F. and DE BOECK, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behav. Res. Methods* **42** 847–862. https://doi.org/10.3758/BRM.42.3.847

[18] MAIR, P. and HATZINGER, R. (2007). Extended rasch modeling: The eRm package for the application of IRT models in R. *J. Stat. Softw.* **20**. https://doi.org/10.18637/jss.v020.i09

[19] NAIR, Y. and JANSON, L. (2023). Randomization tests for adaptively collected data. https://doi.org/10.48550/arXiv.2301.05365

[20] RAMDAS, A., BARBER, R. F., CANDÈS, E. J. and TIBSHIRANI, R. J. (2023). Permutation tests using arbitrary permutation distributions. *Sankhya A* **85** 1156–1177. MR4619064 https://doi.org/10.1007/s13171-023-00308-8

[21] STEPHENS, M. A. (2012). Goodness-of-fit and sufficiency: Exact and approximate tests. *Methodol. Comput. Appl. Probab.* **14** 785–791. MR2966320 https://doi.org/10.1007/s11009-011-9267-2

[22] WANG, G. (2020). A fast MCMC algorithm for the uniform sampling of binary matrices with fixed margins. *Electron. J. Stat.* **14** 1690–1706. MR4082480 https://doi.org/10.1214/20-EJS1702

Statistical Science
2026, Vol. 41, No. 1, 201–210
https://doi.org/10.1214/24-STS960

From Corrado Gini's Early Contributions to Overdispersion to Modern Models of Voting Behaviour

Antonio Forcina and Jose M. Pavía

Abstract. We review some early contributions by Corrado Gini to modified binomial models and predictive probability and highlight their role, once extended to the multivariate context, for modelling voting behaviour in Ecological Inference. A collection of overdispersed multinomial models are described, their properties investigated and a connection to Gini's results for the corresponding binomial model, when available, is provided. After a concise introduction to Ecological Inference, we discuss recent developments aiming at more realistic models of voting behaviour and some connections to Gini's work.

Key words and phrases: History of statistics, predictive probability, finite exchangeability, compound distributions, association and overdispersion, ecological inference.

REFERENCES

BROWN, P. J. and PAYNE, C. D. (1986). Aggregate data, ecological regression, and voting transitions. *J. Amer. Statist. Assoc.* **81** 452–460. MR0845883

DRAPER, D., HODGES, J. S., MALLOWS, C. L. and PREGIBON, D. (1993). Exchangeability and data analysis (with discussion). *J. Roy. Statist. Soc. Ser. A* **156** 9–37. MR1207984 https://doi.org/10.2307/2982858

FORCINA, A. (2016). Gini's early investigation about the distribution of sexes in human births. *Stat. Appl.* **28** 223–238.

FORCINA, A., GNALDI, M. and BRACALENTE, B. (2012). A revised Brown and Payne model of voting behaviour applied to the 2009 elections in Italy. *Stat. Methods Appl.* **21** 109–119. MR2881632 https://doi.org/10.1007/s10260-011-0184-x

FORCINA, A. and PAVÍA, J. M. (2024). eiCircles: Ecological Inference of RxC Tables by Overdispersed-Multinomial Models. Available at https://CRAN.R-project.org/package=eiCircles.

FORCINA, A. and PELLEGRINO, D. (2019). Estimation of voter transitions and the ecological fallacy. *Qual. Quant.* **53** 1859–1874.

GINI, C. (1908). *Il Sesso Dal Punto di Vista Statistico*. Sandron, Milano, Palermo, Napoli, Roma.

GINI, G. (1911). Considerazioni sulla probabilità a posteriori e applicazioni al rapporto dei sessi nelle nascite umane. Technical Report, Studi economico-giuridici dell'Università di Cagliari, reprinted in Metron, 1949, vol. 15, 133–172.

GOODMAN, L. A. (1953). Ecological regressions and behavior of individuals. *Am. Sociol. Rev.* **18** 351–367.

GOSNELL, H. F. and GILL, N. N. (1935). An analysis of the 1932 presidential vote in Chicago. *Amer. Polit. Sci. Rev.* **29** 967–984.

LEXIS, W. H. R. A. (1877). *Zur Theorie der Massenerscheinungen in der Menschlichen Gesellschaft*. BoD–Books on Demand.

OGBURN, W. F. and GOLTRA, I. (1919). How women vote. *Polit. Sci. Quart.* **34** 413–433.

PATTIE, C. and JOHNSTON, R. (2000). 'People who talk together vote together': An exploration of contextual effects in Great Britain. *Ann. Assoc. Amer. Geogr.* **90** 41–66.

PATTISON, P. and ROBINS, G. (2002). Neighborhood-based models for social networks. *Sociol. Method.* **32** 301–337.

PAVÍA, J. M. and ROMERO, R. (2024). Symmetry estimating R × C vote transfer matrices from aggregate data. *J. Roy. Statist. Soc. Ser. A* **187** 919–943. MR4813931 https://doi.org/10.1093/jrsssa/qnae013

PAVÍA, J. M. and THOMSEN, S. R. (2024). ecolRxC: Ecological inference estimation of R×C tables using latent structure approaches. *Polit. Sci. Res. Methods*. Online available. https://doi.org/10.1017/psrm.2024.57

PLACKETT, R. L. (1977). The marginal totals of a 2×2 table. *Biometrika* **64** 37–42. MR0451507 https://doi.org/10.1093/biomet/64.1.37

ROBINSON, W. S. (1950). Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* **15** 351–357.

ROSEN, O., JIANG, W., KING, G. and TANNER, M. A. (2001). Bayesian and frequentist inference for ecological inference: The $R \times C$ case. *Stat. Neerl.* **55** 134–156. MR1862484 https://doi.org/10.1111/1467-9574.00162

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR0852410

STIGLER, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard Univ. Press, Cambridge, MA. MR1712969

Antonio Forcina is Professor, retired, University of Perugia, Perugia, Italy (e-mail: forcinarosara@gmail.com). Jose M. Pavía is Professor, GIPEyOP, University of Valencia, Valencia, Spain (e-mail: Jose.M.Pavia@uv.es).

WAKEFIELD, J. (2004). Ecological inference for 2×2 tables. *J. Roy. Statist. Soc. Ser. A* **167** 385–445. MR2082057 https://doi.org/10.1111/j.1467-985x.2004.02046_1.x

Statistical Science
2026, Vol. 41, No. 1, 211–217
https://doi.org/10.1214/25-STS978

A Note on Distance Variance for Categorical Variables

Qingyang Zhang

Abstract. This study investigates the extension of distance variance, a validated spread metric for continuous and binary variables, to quantify the spread of general categorical variables. We provide both geometric and algebraic characterizations of distance variance, revealing its connections to some commonly used entropy measures, and the variance-covariance matrix of the one-hot encoded representation. However, we demonstrate that distance variance fails to satisfy the Schur-concavity axiom for categorical variables with more than two categories, leading to counterintuitive results. This limitation hinders its applicability as a universal measure of spread.

Key words and phrases: Measure of spread, categorical data, distance variance, entropy.

REFERENCES

[1] ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100. MR0264805 https://doi.org/10.1214/aoms/1177697287

[2] BICKEL, P. and LEHMANN, E. (2012). *Descriptive Statistics for Nonparametric Models, IV. Spread. Selected Works in Probability and Statistics*. Springer, New York.

[3] BORIAH, S. et al. (2008). Similarity measures for categorical data: A comparative evaluation. In *Proceedings of the* 2008 *SIAM International Conference on Data Mining* 243–254.

[4] CASTRO-PRADO, F. et al. (2024). Tests for categorical data beyond Pearson: a distance covariance and energy distance approach. Available at arXiv:2403.12711.

[5] EDELMANN, D. and GOEMAN, J. (2022). A regression perspective on generalized distance covariance and the Hilbert–Schmidt independence criterion. *Statist. Sci.* **37** 562–579. MR4497232 https://doi.org/10.1214/21-sts841

[6] EDELMANN, D., RICHARDS, D. and VOGEL, D. (2020). The distance standard deviation. *Ann. Statist.* **48** 3395–3416. MR4185813 https://doi.org/10.1214/19-AOS1935

[7] GILULA, Z. and HABERMAN, S. J. (1995). Dispersion of categorical variables and penalty functions: Derivation, estimation, and comparability. *J. Amer. Statist. Assoc.* **90** 1447–1462. MR1379489

[8] HABERMAN, S. J. (1982). Analysis of dispersion of multinomial responses. *J. Amer. Statist. Assoc.* **77** 568–580. MR0675884

[9] LAD, F., SANFILIPPO, G. and AGRÒ, G. (2015). Extropy: Complementary dual of entropy. *Statist. Sci.* **30** 40–58. MR3317753 https://doi.org/10.1214/14-STS430

[10] LIN, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the Fifteenth International Conference on Machine Learning* 296–304. Morgan Kaufmann, San Francisco, CA, USA.

[11] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

[12] MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications. Mathematics in Science and Engineering* **143**. Academic Press, San Deigo, CA. MR0552278

[13] PANDA, S. et al. (2019). High-dimensional and universally consistent k-sample tests. Available at arXiv:1910.08883.

[14] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423, 623–656. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[15] SHEN, C., PANDA, S. and VOGELSTEIN, J. T. (2022). The chi-square test of distance correlation. *J. Comput. Graph. Statist.* **31** 254–262. MR4387225 https://doi.org/10.1080/10618600.2021.1938585

[16] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

[17] TSALLIS, C. (1988). Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52** 479–487. MR0968597 https://doi.org/10.1007/BF01016429

[18] VAN DE VELDEN, M. et al. (2023). A general framework for implementing distances for categorical variables. Available at arXiv:2301.02190.

[19] ZHANG, Q. (2019). Independence test for large sparse contingency tables based on distance correlation. *Statist. Probab. Lett.* **148** 17–22. MR3899082 https://doi.org/10.1016/j.spl.2018.12.010

[20] ZHANG, Q. (2024). On the properties of distance covariance for categorical data: Robustness, sure screening, and approximate null distributions. *Scand. J. Stat.* https://doi.org/10.1111/sjos.12771

Qingyang Zhang is an Associate Professor, Department of Mathematical Sciences, University or Arkansas, Fayetteville, Arkansas 72701, USA (e-mail: qz008@uark.edu).

Statistical Science
2026, Vol. 41, No. 1, 218–221
https://doi.org/10.1214/25-STS982

Conditionality Principle Under Unconstrained Randomness

Vladimir Vovk

Abstract. A very simple example demonstrates that Fisher's application of the conditionality principle to regression ("fixed-x regression"), endorsed by David Sprott and many other followers, makes prediction impossible in the context of statistical learning theory. On the other hand, relaxing the requirement of conditionality makes it possible via, for example, conformal prediction.

Key words and phrases: Assumption of randomness, conditionality principle, machine learning, prediction, regression.

REFERENCES

[1] ALDRICH, J. (2005). Fisher and regression. *Statist. Sci.* **20** 401–417. Original title: "The origins of fixed X regression". MR2210227 https://doi.org/10.1214/088342305000000331

[2] ANGELOPOULOS, A. N., BARBER, R. F. and BATES, S. (2024). Theoretical Foundations of Conformal Prediction. Technical Report No. arXiv:2411.11824 [math.ST] arXiv.org e-Print archive. Pre-publication version of a monograph to be published by Cambridge University Press.

[3] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). The limits of distribution-free conditional predictive inference. *Inf. Inference* **10** 455–482. MR4270755 https://doi.org/10.1093/imaiai/iaaa017

[4] BELLOTTI, A. (2021). Approximation to object conditional validity with inductive conformal predictors. *Proc. Mach. Learn. Res.* **152** 4–23.

[5] BIRNBAUM, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–326. MR0138176

[6] COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. MR0370837

[7] FISHER, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *J. R. Stat. Soc.* **85** 597–612.

[8] FISHER, R. A. (1955). Statistical methods and scientific induction. *J. Roy. Statist. Soc. Ser. B, Methodol.* **17** 69–78. MR0076233

[9] FISHER, R. A. (1990). *Statistical Inference and Analysis*. *Oxford Science Publications*. The Clarendon Press, New York. MR1076366

[10] GUAN, L. (2023). Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika* **110** 33–50. MR4565442 https://doi.org/10.1093/biomet/asac040

[11] HORE, R. and BARBER, R. F. (2025). Conformal prediction with local weights: Randomization enables robust guarantees. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **87** 549–578. MR4896656 https://doi.org/10.1093/jrsssb/qkae103

[12] LEHMANN, E. L. and SCHOLZ, F.-W. (1992). Ancillarity. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu* (M. Ghosh and P. K. Pathak, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 32–51. IMS, Hayward, CA. MR1194408 https://doi.org/10.1214/lnms/1215458837

[13] LEI, J. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika* **106** 749–764. MR4031197 https://doi.org/10.1093/biomet/asz046

[14] LEI, J. and WASSERMAN, L. (2014). Distribution-free prediction bands for non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 71–96. MR3153934 https://doi.org/10.1111/rssb.12021

[15] ROMANO, Y., SESIA, M. and CANDÈS, E. J. (2020). Classification with valid and adaptive coverage. In *Advances in Neural Information Processing Systems* 33 (*NeurIPS* 2020).

[16] THOMPSON, M. (1989). A conversation with David A. Sprott. *Liaison* **3**. Available (in July 2025) on the web.

[17] VAPNIK, V. N. (1998). *Statistical Learning Theory*. (*Adaptive and Learning Systems for Signal Processing*, *Communications*, *and Control*). Wiley, New York. MR1641250

[18] VAPNIK, V. N. and CHERVONENKIS, A. Y. (1974). Теория распознавания образов (*Theory of Pattern Recognition*). Nauka, Moscow. German translation: *Theorie der Zeichenerkennung*, Akademie, Berlin, 1979.

[19] VOVK, V., GAMMERMAN, A. and SHAFER, G. (2022). *Algorithmic Learning in a Random World*, 2nd ed. Springer, Cham.

Vladimir Vovk is Professor, Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, UK (e-mail: v.vovk@rhul.ac.uk).

Statistical Science
2026, Vol. 41, No. 1, 222–226
https://doi.org/10.1214/25-STS983

Choosing Alpha Post Hoc: The Danger of Multiple Standard Significance Thresholds

Jesse Hemerik and Nick Koning

Abstract. A fundamental assumption of classical hypothesis testing is that the significance threshold α is chosen independently from the data. The validity of confidence intervals likewise relies on choosing α beforehand. We point out that the independence of α is guaranteed in practice, because in most fields there exists one standard α that everyone uses—so that α is automatically independent of everything. However, there have been recent calls to decrease α from 0.05 to 0.005. We note that this may lead to multiple accepted standard thresholds within one scientific field. For example, different journals may require different significance thresholds. As a consequence, some researchers may be tempted to conveniently choose their α based on their p-value. We use examples to illustrate that this severely invalidates hypothesis tests, and mention some potential solutions.

Key words and phrases: e-value, hypothesis test, Neyman–Pearson, p-value, significance level.

REFERENCES

[1] AGUINIS, H., VASSAR, M. and WAYANT, C. (2021). On reporting and interpreting statistical significance and p values in medical research. *BMJ Evid.-Based Med*. **26** 39–42.

[2] AMRHEIN, V., GREENLAND, S. and MCSHANE, B. (2019). Scientists rise up against statistical significance. *Nature* **567** 305–307.

[3] BANERJEE, A., CHITNIS, U., JADHAV, S., BHAWALKAR, J. and CHAUDHURY, S. (2009). Hypothesis testing, type I and type II errors. *Ind. Psychiatry J*. **18** 127.

[4] BENJAMIN, D. J., BERGER, J. O., JOHANNESSON, M., NOSEK, B. A., WAGENMAKERS, E.-J., BERK, R., BOLLEN, K. A., BREMBS, B., BROWN, L. et al. (2018). Redefine statistical significance. *Nat. Hum. Behav.* **2** 6–10.

[5] BETENSKY, R. A. (2019). The p-value requires context, not a threshold. *Amer. Statist*. **73** 115–117. MR3925715 https://doi.org/10.1080/00031305.2018.1529624

[6] BORENSTEIN, M. (1994). The case for confidence intervals in controlled clinical trials. *Control. Clin. Trials* **15** 411–428.

[7] BORENSTEIN, M. (2001). *Power and Precision* **1**. Taylor & Francis, London.

[8] DI LEO, G. and SARDANELLI, F. (2020). Statistical significance: P value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *Eur. Radiol. Exp*. **4** 1–8.

[9] FITZPATRICK, B. G., GORMAN, D. M. and TROMBATORE, C. (2024). Impact of redefining statistical significance on P-hacking and false positive rates: An agent-based model. *PLoS ONE* **19** e0303262.

[10] GREENLAND, S., SENN, S. J., ROTHMAN, K. J., CARLIN, J. B., POOLE, C., GOODMAN, S. N. and ALTMAN, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol*. **31** 337–350.

[11] GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. (2024). Safe testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **86** 1091–1128. MR4825000 https://doi.org/10.1093/jrsssb/qkae011

[12] GRÜNWALD, P. D. (2024). Beyond Neyman–Pearson: E-values enable hypothesis testing with a data-driven alpha. *Proc. Natl. Acad. Sci. USA* **121** Paper No. e2302098121. MR4826968 https://doi.org/10.1073/pnas.2302098121

[13] HANSEN, B. (2022). *Econometrics*. Princeton University Press, Princeton, NJ.

[14] HELD, L. (2019). The assessment of intrinsic credibility and a new argument for $p < 0.005$. *R. Soc. Open Sci*. **6** 181534. MR3949248 https://doi.org/10.1098/rsos.181534

[15] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist*. **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

[16] HUBBARD, R. (2019). Will the ASA's efforts to improve statistical practice be successful? Some evidence to the contrary. *Amer. Statist*. **73** 31–35. MR3925706 https://doi.org/10.1080/00031305.2018.1497540

[17] HUBBARD, R. and BAYARRI, M. J. (2003). Confusion over measures of evidence (p's) versus errors (α's) in classical statistical testing. *Amer. Statist*. **57** 171–178. MR2542412 https://doi.org/10.1198/0003130031856

[18] HUBBARD, R. and LINDSAY, R. M. (2008). Why p values are not a useful measure of evidence in statistical significance testing. *Theory Psychol*. **18** 69–88.

Jesse Hemerik is an Assistant Professor at the Econometric Institute, Erasmus University, Rotterdam, The Netherlands (e-mail: hemerik@ese.eur.nl). Nick Koning is an Assistant Professor at the Econometric Institute, Erasmus University, Rotterdam, The Netherlands (e-mail: n.w.koning@ese.eur.nl).

[19] IMBENS, G. W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *J. Econ. Perspect.* **35** 157–174.

[20] IOANNIDIS, J. P. (2005). Why most published research findings are false. *PLoS Med.* **2** e124.

[21] IOANNIDIS, J. P. (2016). Why most clinical research is not useful. *PLoS Med.* **13** e1002049.

[22] IOANNIDIS, J. P. (2018). The proposal to lower P value thresholds to 005. *JAMA* **319** 1429–1430.

[23] JOHNSON, V. E., PAYNE, R. D., WANG, T., ASHER, A. and MANDAL, S. (2017). On the reproducibility of psychological science. *J. Amer. Statist. Assoc.* **112** 1–10. MR3646548 https://doi.org/10.1080/01621459.2016.1240079

[24] KELTER, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP. *BMC Med. Res. Methodol.* **20** 1–12.

[25] KHAN, M. S., IRFAN, S., KHAN, S. U., MEHRA, M. R. and VADUGANATHAN, M. (2020). Transforming the interpretation of significance in heart failure trials. *Eur. J. Heart Fail.* **22** 177.

[26] KIM, J. H. and JI, P. I. (2015). Significance testing in empirical finance: A critical review and assessment. *J. Empir. Finance* **34** 1–14.

[27] KONING, N. W. (2024). Post-hoc alpha Hypothesis Testing and the Post-hoc p-value. arXiv Preprint. Available at arXiv:2312.08040.

[28] KONING, N. W. (2024). Continuous Testing: Unifying Tests and E-values. arXiv Preprint. Available at arXiv:2409.05654.

[29] LAKENS, D., ADOLFI, F. G., ALBERS, C. J., ANVARI, F., APPS, M. A., ARGAMON, S. E., BAGULEY, T., BECKER, R. B., BENNING, S. D. et al. (2018). Justify your alpha. *Nat. Hum. Behav.* **2** 168–171.

[30] LAKENS, D., MCLATCHIE, N., ISAGER, P. M., SCHEEL, A. M. and DIENES, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *J. Gerontol., Ser. B* **75** 45–57.

[31] LARSSON, M., RAMDAS, A. and RUF, J. (2025). The numeraire e-variable and reverse information projection. *Ann. Statist.* **53**. MR4925114 https://doi.org/10.1214/24-aos2487

[32] LEHMANN, E. L. and ROMANO, J. P. (2021). *Testing Statistical Hypotheses*, 4th ed. *Springer Texts in Statistics*. Springer, Cham. MR4489085

[33] MAIER, M. and LAKENS, D. (2022). Justify your alpha: A primer on two practical approaches. *Adv. Methods Pract. Psychol. Sci.* **5** 25152459221080396.

[34] MARTIN, W. E. and BRIDGMON, K. D. (2012). *Quantitative and Statistical Research Methods: From Hypothesis to Results*. Wiley, New York.

[35] MAYO, D. G. and HAND, D. (2022). Statistical significance and its critics: Practicing damaging science, or damaging scientific practice? *Synthese* **200** 220.

[36] MCSHANE, B. B., GAL, D., GELMAN, A., ROBERT, C. and TACKETT, J. L. (2019). Abandon statistical significance. *Amer. Statist.* **73** 235–245. MR3925729 https://doi.org/10.1080/00031305.2018.1527253

[37] MUDGE, J. F., BAKER, L. F., EDGE, C. B. and HOULAHAN, J. E. (2012). Setting an optimal α that minimizes errors in null hypothesis significance tests. *PLoS ONE* **7** e32734.

[38] MUFF, S., NILSEN, E. B., O'HARA, R. B. and NATER, C. R. (2022). Rewriting results sections in the language of evidence. *Trends Ecol. Evol.* **37** 203–210.

[39] RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

[40] SCHEEL, A. M., TIOKHIN, L., ISAGER, P. M. and LAKENS, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspect. Psychol. Sci.* **16** 744–755.

[41] SCHNOG, J.-J. B., SAMSON, M. J., GANS, R. O. and DUITS, A. J. (2021). An urgent call to raise the bar in oncology. *Br. J. Cancer* **125** 1477–1485.

[42] SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J. R. Stat. Soc. A, General* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

[43] SZUCS, D. and IOANNIDIS, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* **15** e2000797.

[44] THAKUR, P. and JHA, V. (2022). Potential effects of lowering the threshold of statistical significance in the field of chronic rhinosinusitis-A meta-research on published randomized controlled trials over last decade. *Braz. J. Otorhinolaryngol.* **88** S83–S89.

[45] THOMPSON, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychol. Sch.* **44** 423–432.

[46] UYGUN TUNÇ, D., TUNÇ, M. N. and LAKENS, D. (2023). The epistemic and pragmatic function of dichotomous claims based on statistical hypothesis tests. *Theory Psychol.* **33** 403–423.

[47] VAN DE SCHOOT, R., DEPAOLI, S., KING, R., KRAMER, B., MÄRTENS, K., TADESSE, M. G., VANNUCCI, M., GELMAN, A., VEEN, D. et al. (2021). Bayesian statistics and modelling. *Nat. Rev. Methods Primers* **1** 1.

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

[49] VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

[50] WASSERSTEIN, R. L. and LAZAR, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose [Editorial]. *Amer. Statist.* **70** 129–133. MR3511040 https://doi.org/10.1080/00031305.2016.1154108

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.
