STATISTICAL SCIENCE

Volume 32, Number 4 November 2017

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 32, Number 4, November 2017. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Statistical Science **Volume 32, Number 4 (487–640) November 2017**

Volume 32 **Number 4** **November 2017**

Statistical Science
2017, Vol. 32, No. 4, 487–500
DOI: 10.1214/17-STS619

Sufficientness Postulates for Gibbs-Type Priors and Hierarchical Generalizations

S. Bacallado, M. Battiston, S. Favaro and L. Trippa

Abstract. A fundamental problem in Bayesian nonparametrics consists of selecting a prior distribution by assuming that the corresponding predictive probabilities obey certain properties. An early discussion of such a problem, although in a parametric framework, dates back to the seminal work by English philosopher W. E. Johnson, who introduced a noteworthy characterization for the predictive probabilities of the symmetric Dirichlet prior distribution. This is typically referred to as Johnson's "sufficientness" postulate. In this paper, we review some nonparametric generalizations of Johnson's postulate for a class of nonparametric priors known as species sampling models. In particular, we revisit and discuss the "sufficientness" postulate for the two parameter Poisson–Dirichlet prior within the more general framework of Gibbs-type priors and their hierarchical generalizations.

Key words and phrases: Bayesian nonparametrics, Dirichlet and two parameter Poisson–Dirichlet process, discovery probability, Gibbs-type species sampling models, hierarchical species sampling models, Johnson's "sufficientness" postulate, Pólya-like urn scheme, predictive probabilities.

REFERENCES

[1] ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969

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

[3] BACALLADO, S., BATTISTON, M., FAVARO, S. and TRIPPA, L. (2017). Supplement to "Sufficientness postulates for Gibbs-type priors and hierarchical generalizations." DOI:10.1214/17-STS619SUPP.

[4] BACALLADO, S., FAVARO, S. and TRIPPA, L. (2013). Bayesian nonparametric analysis of reversible Markov chains. *Ann. Statist.* **41** 870–896. MR3099124

[5] BACALLADO, S., FAVARO, S. and TRIPPA, L. (2015). Looking-backward probabilities for Gibbs-type exchangeable random partitions. *Bernoulli* **21** 1–37. MR3322311

[6] BACALLADO, S., FAVARO, S. and TRIPPA, L. (2015). Bayesian nonparametric inference for shared species richness in multiple populations. *J. Statist. Plann. Inference* **166** 14–23. MR3390130

[7] BATTISTON, M., FAVARO, S., ROY, D. M. and TEH, Y. W. (2016). A characterization of product-form exchangeable feature probability functions. Preprint. Available at arXiv:1607.02066.

[8] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge. MR2253162

[9] BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614

[10] BRODERICK, T., PITMAN, J. and JORDAN, M. I. (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Anal.* **8** 801–836. MR3150470

[11] CARON, F. (2012). Bayesian nonparametric models for bipartite graphs. In *NIPS'*12. *Proc.* 25*th International Conference*

S. Bacallado is University Lecturer, Department of Pure Mathematics and Mathematical Statistics, CB3 0WB Cambridge, United Kingdom (e-mail: sergiobacallado@gmail.com). M. Battiston is Postdoctoral Associate, Department of Statistics, OX1 3LB Oxford, United Kingdom (e-mail: marco.battiston@stats.ox.ac.uk). S. Favaro is Associate Professor, Department of Economics and Statistics, Corso Unione Sovietica 218/bis, 10134 Torino, Italy and also affiliated with Collegio Carlo Alberto, Moncalieri, Italy (e-mail: stefano.favaro@unito.it). L. Trippa is Associate Professor, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 3 Blackfan Circle, Boston, Massachusetts 02215, USA, and Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA (e-mail: ltrippa@jimmy.harvard.edu).

on Neural Information Processing Systems 2051–2059. Curran Associates Inc., Red Hook, NY.

[12] CARON, F. and FOX, E. B. (2015). Sparse graphs with exchangeable random measures. Preprint. Available at arXiv:1401.1137.

[13] CARON, F., TEH, Y. W. and MURPHY, T. B. (2014). Bayesian nonparametric Plackett–Luce models for the analysis of preferences for college degree programmes. *Ann. Appl. Stat.* **8** 1145–1181. MR3262549

[14] CHEN, C., DING, N. and BUNTINE, W. (2012). Dependent hierarchical normalized random measures for dynamic topic modeling. In *Proc. 29th International Conference on Machine Learning* (*ICML*-12) 895–902.

[15] CHEN, C., RAO, V. A., BUNTINE, W. and TEH, Y. W. (2013). Dependent normalized random measures. In *Proc. 30th International Conference on Machine Learning* (*ICML*-13) 969–977.

[16] CRANE, H. (2016). The ubiquitous Ewens sampling formula. *Statist. Sci.* **31** 1–19. MR3458585

[17] DE FINETTI, B. (1938). Sur la condition d'"équivalence partielle." In *VI Colloque Gèneve: "Act. Sc. Ind."* Herman, Paris.

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

[19] ENGEN, S. (1978). *Stochastic Abundance Models: With Emphasis on Biological Communities and Species Diversity*. Chapman & Hall, London. MR0515721

[20] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3** 87–112. Erratum, ibid. **3** 240 (1972); erratum, ibid. **3** 376 (1972). MR0325177

[21] FAVARO, S., LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2009). Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 993–1008. MR2750254

[22] FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2012). A new estimator of the discovery probability. *Biometrics* **68** 1188–1196. MR3040025

[23] FAVARO, S. and WALKER, S. G. (2013). Slice sampling σ-stable Poisson–Kingman mixture models. *J. Comput. Graph. Statist.* **22** 830–847. MR3173745

[24] FENG, S. and HOPPE, F. M. (1998). Large deviation principles for some random combinatorial structures in population genetics and Brownian motion. *Ann. Appl. Probab.* **8** 975–994. MR1661315

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

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

[27] GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* **138** 5674–5685.

[28] GOOD, I. J. (1965). *The Estimation of Probabilities. An Essay on Modern Bayesian Methods. Research Monograph* **30**. MIT Press, Cambridge, MA. MR0185724

[29] HEAUKULANI, C. and ROY, D. M. (2015). Gibbs-type Indian buffet processes. Preprint. Available at arXiv:1512.02543.

[30] HERLAU, T. (2015). Completely random measures for modeling block-structured sparse networks. Preprint. Available at arXiv:1507.02925.

[31] HJORT, N. L. (2000). Bayesian analysis for a generalised Dirichlet process prior Technical report, Matematisk Institutt, Universitetet i Oslo.

[32] HO, M., JAMES, L. F. and LAU, J. W. (2007). Gibbs partitions (EPPF's) derived from a stable subordinator are Fox H and Meijer G transforms. Preprint. Available at arXiv:0708.0619.

[33] HOPPE, F. M. (1984). Pólya-like urns and the Ewens' sampling formula. *J. Math. Biol.* **20** 91–94. MR0758915

[34] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729

[35] JAMES, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Preprint. Available at arXiv:math/0205093.

[36] JAMES, L. F. (2013). Stick-breaking PG(α, ζ)-generalized Gamma processes. Preprint. Available at arXiv:1308.6570.

[37] JO, S., LEE, J., MÜLLER, P., QUINTANA, F. A. and TRIPPA, L. (2017). Dependent species sampling models for spatial density estimation. *Bayesian Anal.* **12** 379–406. MR3620738

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

[39] KINGMAN, J. F. C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* (2) **18** 374–380. MR0509954

[40] KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1** 705–711. MR0350950

[41] LEE, J., QUINTANA, F. A., MÜLLER, P. and TRIPPA, L. (2013). Defining predictive probability functions for species sampling models. *Statist. Sci.* **28** 209–222. MR3112406

[42] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291. MR2236441

[43] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. MR2416792

[44] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 715–740. MR2370077

[45] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics. Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge Univ. Press, Cambridge. MR2730661

[46] LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica* **18** 1653–1668. MR2469329

[47] LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18** 1519–1547. MR2434179

[48] LO, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. MR0733519

[49] LO, A. Y. (1991). A characterization of the Dirichlet process. *Statist. Probab. Lett.* **12** 185–187. MR1130354

[50] LOMELÍ, M., FAVARO, S. and TEH, Y. W. (2017). A marginal sampler for σ-stable Poisson–Kingman mixture models. *J. Comput. Graph. Statist.* **26** 44–53. MR3610406

[51] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448

[52] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249

[53] PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784

[54] PITMAN, J. (2003). Poisson–Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. MR2004330

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

[56] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129

[57] PRÜNSTER, I. (2002). Random probability measures derived from increasing additive processes and their application to Bayesian statistics Ph.D. thesis, Univ. Pavia.

[58] QUINTANA, F. A. and IGLESIAS, P. L. (2003). Bayesian clustering and product partition models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 557–574. MR1983764

[59] RAVEL, J., GAJER, P., ABDO, Z., SCHNEIDER, G. M., KOENIG, S. S., MCCULLE, S. L., KARLEBACH, S., GORLE, R., RUSSELL, J., TACKET, C. O., BROTMAN, R. M., DAVIS, C. C., AULT, K., PERALTA, L. and FORNEY, L. J. (2011). Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. USA* **108** 4680–4687.

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

[61] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31** 560–585. MR1983542

[62] ROLLES, S. W. W. (2003). How edge-reinforced random walk arises naturally. *Probab. Theory Related Fields* **126** 243–260. MR1990056

[63] ROY, D. M. (2014). The continuum-of-urns scheme, generalized beta and Indian buffet processes, and hierarchies thereof. Preprint. Available at arXiv:1501.00208.

[64] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original, revised by the author. MR1739520

[65] TEH, Y. W. (2006). A hierarchical Bayesian language model based on Pitman–Yor processes. In *Proceedings of the* 21*st International Conference on Computational Linguistic*.

[66] TEH, Y. W. and GÖRÜR (2010). Indian buffet processes with power-law behavior. In *NIPS'*09. *Proc.* 22*nd International Conference on Neural Information Processing Systems* 1838–1846. Curran Associates Inc., Red Hook, NY.

[67] TEH, Y. W. and JORDAN, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics. Camb. Ser. Stat. Probab. Math.* **28** 158–207. Cambridge Univ. Press, Cambridge. MR2730663

[68] TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480

[69] WALKER, S. and MULIERE, P. (1999). A characterization of a neutral to the right prior via an extension of Johnson's sufficientness postulate. *Ann. Statist.* **27** 589–599. MR1714716

[70] WATTERSON, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13** 639–651. MR0504014

[71] ZABELL, S. L. (1982). W. E. Johnson's "sufficientness" postulate. *Ann. Statist.* **10** 1090–1099 (1 plate). MR0673645

[72] ZABELL, S. L. (1992). Predicting the unpredictable. *Synthese* **90** 205–232. MR1148566

[73] ZABELL, S. L. (1995). Characterizing Markov exchangeable sequences. *J. Theoret. Probab.* **8** 175–178. MR1308676

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

[75] ZABELL, S. L. (2005). *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge Univ. Press, New York. MR2199124

Statistical Science
2017, Vol. 32, No. 4, 501–513
DOI: 10.1214/17-STS620

Spherical Process Models for Global Spatial Statistics

Jaehong Jeong, Mikyoung Jun and Marc G. Genton

Abstract. Statistical models used in geophysical, environmental, and climate science applications must reflect the curvature of the spatial domain in global data. Over the past few decades, statisticians have developed covariance models that capture the spatial and temporal behavior of these global data sets. Though the geodesic distance is the most natural metric for measuring distance on the surface of a sphere, mathematical limitations have compelled statisticians to use the chordal distance to compute the covariance matrix in many applications instead, which may cause physically unrealistic distortions. Therefore, covariance functions directly defined on a sphere using the geodesic distance are needed. We discuss the issues that arise when dealing with spherical data sets on a global scale and provide references to recent literature. We review the current approaches to building process models on spheres, including the differential operator, the stochastic partial differential equation, the kernel convolution, and the deformation approaches. We illustrate realizations obtained from Gaussian processes with different covariance structures and the use of isotropic and nonstationary covariance models through deformations and geographical indicators for global surface temperature data. To assess the suitability of each method, we compare their log-likelihood values and prediction scores, and we end with a discussion of related research problems.

Key words and phrases: Axial symmetry, chordal distance, geodesic distance, nonstationarity, smoothness, sphere.

REFERENCES

ALEGRÍA, A., CARO, S., BEVILACQUA, M., PORCU, E. and CLARKE, J. (2017a). Estimating covariance functions of multivariate Skew–Gaussian random fields on the sphere. *Spat. Stat.* To appear. Preprint. Available at arXiv:1612.03341.

ALEGRÍA, A., PORCU, E., FURRER, R. and MATEU, J. (2017b). Covariance functions for multivariate Gaussian fields evolving temporally over planet earth. Preprint. Available at arXiv:1701.06010.

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

BANERJEE, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics* **61** 617–625. MR2140936

BOLIN, D. and LINDGREN, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Stat.* **5** 523–550. MR2810408

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2015). Comparing and selecting spatial predictors using local criteria. *TEST* **24** 1–28. MR3314567

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2016). A comparison of spatial predictors when datasets could be very large. *Stat. Surv.* **10** 100–131. MR3527662

CASTRUCCIO, S. and GENTON, M. G. (2014). Beyond axial symmetry: An improved class of models for global data. *Stat* **3** 48–55.

Jaehong Jeong is Postdoctoral Fellow, CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia (e-mail: jaehong.jeong@kaust.edu.sa). Mikyoung Jun is Associate Professor, Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, USA (e-mail: mjun@stat.tamu.edu). Marc G. Genton is Distinguished Professor, CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia (e-mail: marc.genton@kaust.edu.sa).

CASTRUCCIO, S. and GENTON, M. G. (2016). Compressing an ensemble with statistical models: An algorithm for global 3D spatio-temporal temperature. *Technometrics* **58** 319–328. MR3520661

CASTRUCCIO, S. and GUINNESS, J. (2017). An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 329–344. MR3611690

CASTRUCCIO, S. and STEIN, M. L. (2013). Global space–time models for climate ensembles. *Ann. Appl. Stat.* **7** 1593–1611. MR3127960

CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ. MR2848400

DAS, B. (2000). Global covariance modeling: A deformation approach to anisotropy. Ph.D. thesis, Univ. Washington, Seattle, WA. MR2701165

DU, J., MA, C. and LI, Y. (2013). Isotropic variogram matrix functions on spheres. *Math. Geosci.* **45** 341–357. MR3107159

DU, J., LEONENKO, N., MA, C. and SHU, H. (2012). Hyperbolic vector random fields with hyperbolic direct and cross covariance functions. *Stoch. Anal. Appl.* **30** 662–674. MR2946043

FAN, M., PAUL, D., LEE, T. C. M. and MATSUO, T. (2017). Modeling tangential vector fields on a sphere. *J. Amer. Statist. Assoc.* To appear. Preprint. Available at arXiv:1612.0806.

GNEITING, T. (1998). Simple tests for the validity of correlation function models on the circle. *Statist. Probab. Lett.* **39** 119–122. MR1652540

GNEITING, T. (1999a). Isotropic correlation functions on d-dimensional balls. *Adv. in Appl. Probab.* **31** 625–631. MR1742685

GNEITING, T. (1999b). Correlation functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.* **125** 2449–2464.

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554

GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. MR2752612

GUINNESS, J. and FUENTES, M. (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. *J. Multivariate Anal.* **143** 143–152. MR3431424

HEATON, M. J., KATZFUSS, M., BERRETT, C. and NYCHKA, D. W. (2014). Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics* **25** 2–15. MR3233740

HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Stat.* **5** 173–190.

HITCZENKO, M. and STEIN, M. L. (2012). Some theory for anisotropic processes on the sphere. *Stat. Methodol.* **9** 211–227. MR2863609

HUANG, C., ZHANG, H. and ROBESON, S. M. (2011). On the validity of commonly used covariance and variogram functions on the sphere. *Math. Geosci.* **43** 721–733. MR2824128

JEONG, J. and JUN, M. (2015a). A class of Matérn-like covariance functions for smooth processes on a sphere. *Spat. Stat.* **11** 1–18. MR3311853

JEONG, J. and JUN, M. (2015b). Covariance models on the surface of a sphere: When does it matter? *Stat* **4** 167–182. MR3405399

JEONG, J., CASTRUCCIO, S., CRIPPA, P. and GENTON, M. G. (2017). Reducing storage of global wind ensembles with stochastic generators. Preprint. Available at arXiv:1702.01995.

JONES, R. H. (1963). Stochastic processes on a sphere. *Ann. Math. Stat.* **34** 213–218. MR0170378

JUN, M. (2011). Non-stationary cross-covariance models for multivariate processes on a globe. *Scand. J. Stat.* **38** 726–747. MR2859747

JUN, M. (2014). Matérn-based nonstationary cross-covariance models for global processes. *J. Multivariate Anal.* **128** 134–146. MR3199833

JUN, M. and STEIN, M. L. (2007). An approach to producing space–time covariance functions on spheres. *Technometrics* **49** 468–479. MR2394558

JUN, M. and STEIN, M. L. (2008). Nonstationary covariance models for global data. *Ann. Appl. Stat.* **2** 1271–1289. MR2655659

KENT, J. T. (1982). The Fisher–Bingham distribution on the sphere. *J. Roy. Statist. Soc. Ser. B* **44** 71–80. MR0655376

KLEIBER, W. and NYCHKA, D. (2012). Nonstationary modeling for multivariate spatial processes. *J. Multivariate Anal.* **112** 76–91. MR2957287

LANG, A. and SCHWAB, C. (2015). Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.* **25** 3047–3094. MR3404631

LI, T.-H. (1999). Multiscale representation and analysis of spherical data by spherical wavelets. *SIAM J. Sci. Comput.* **21** 924–953. MR1755172

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727

MA, C. (2012). Stationary and isotropic vector random fields on spheres. *Math. Geosci.* **44** 765–778. MR2956272

MA, C. (2013a). Mittag-Leffler vector random fields with Mittag–Leffler direct and cross covariance functions. *Ann. Inst. Statist. Math.* **65** 941–958. MR3105803

MA, C. (2013b). Student's t vector random fields with power-law and log-law decaying direct and cross covariances. *Stoch. Anal. Appl.* **31** 167–182. MR3007889

MA, C. (2015). Isotropic covariance matrix functions on all spheres. *Math. Geosci.* **47** 699–717. MR3368160

MA, C. (2016a). Isotropic covariance matrix polynomials on spheres. *Stoch. Anal. Appl.* **34** 679–706. MR3507186

MA, C. (2016b). Stochastic representations of isotropic vector random fields on spheres. *Stoch. Anal. Appl.* **34** 389–403. MR3488255

MA, C. (2016c). Time varying axially symmetric vector random fields on the sphere. *Random Oper. Stoch. Equ.* **24** 255–266. MR3574938

MA, C. (2017). Time varying isotropic vector random fields on spheres. *J. Theoret. Probab.* To appear.

MARINUCCI, D. and PECCATI, G. (2011). *Random Fields on the Sphere*: *Representation*, *Limit Theorems and Cosmological Applications*. *London Mathematical Society Lecture Note Series* **389**. Cambridge Univ. Press, Cambridge. MR2840154

NORTH, G. R., WANG, J. and GENTON, M. G. (2011). Correlation models for temperature fields. *J. Climate* **24** 5850–5862.

OH, H.-S. and LI, T.-H. (2004). Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 221–238. MR2035768

POPPICK, A. and STEIN, M. L. (2014). Using covariates to model dependence in nonstationary, high-frequency meteorological processes. *Environmetrics* **25** 293–305. MR3258008

PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.* **111** 888–898. MR3538713

PORCU, E. and SCHILLING, R. L. (2011). From Schoenberg to Pick-Nevanlinna: Toward a complete picture of the variogram class. *Bernoulli* **17** 441–455. MR2797998

PORCU, E. and ZASTAVNYI, V. (2011). Characterization theorems for some classes of covariance functions associated to vector valued random fields. *J. Multivariate Anal.* **102** 1293–1301. MR2811618

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

SCHAFFRIN, B. (1993). Biased kriging on the sphere? In *Geostatistics Tróia'* 92 (A. Soares, ed.) 121–131. Springer, New York.

SCHOENBERG, I. J. (1942). Positive definite functions on spheres. *Duke Math. J.* **9** 96–108. MR0005922

SCHREINER, M. (1997). On a new condition for strictly positive definite functions on spheres. *Proc. Amer. Math. Soc.* **125** 531–539. MR1353398

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

STEIN, M. L. (2007). Spatial variation of total column ozone on a global scale. *Ann. Appl. Stat.* **1** 191–210. MR2393847

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16. MR0618629

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

XU, G. and GENTON, M. G. (2017). Tukey g-and-h random fields. *J. Amer. Statist. Assoc.* **112** 1236–1249.

YADRENKO, M. Ĭ. (1983). *Spectral Theory of Random Fields*. Optimization Software, Inc., Publications Division, New York. Translated from the Russian. MR0697386

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Vol. I*: *Basic Results*. Springer, New York. MR0893393

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

Statistical Science
2017, Vol. 32, No. 4, 514–530
DOI: 10.1214/17-STS621

The General Structure of Evidence Factors in Observational Studies

Paul R. Rosenbaum

Abstract. The general structure of evidence factors is examined in terms of the knit product of two permutation groups. An observational or nonrandomized study of treatment effects has two evidence factors if it permits two (nearly) independent tests of the null hypothesis of no treatment effect and two (nearly) independent sensitivity analyses for those tests. Either of the two tests may be biased by nonrandom treatment assignment, but certain biases that would invalidate one test would have no impact on the other, so if the two tests concur, then some aspects of biased treatment assignment have been partially addressed. Expressed in terms of the knit product of two permutation groups, the structure of evidence factors is simpler and less cluttered, but at the same time more general and easier to apply in a new context. The issues are exemplified by an observational study of cigarette smoking as a cause of periodontal disease.

Key words and phrases: Evidence factor, knit product, permutation group, permutation inference, randomization inference, semidirect product, sensitivity analysis, wreath product, Zappa–Szep product.

REFERENCES

ALAM, K. (1974). Some nonparametric tests of randomness. *J. Amer. Statist. Assoc.* **69** 738–739. MR0370907

ATEŞ, F. and ÇEVIK, A. S. (2009). Knit products of some groups and their applications. *Rend. Semin. Mat. Univ. Padova* **121** 1–11. MR2542131

AUSTIN, P. C. and STUART, E. A. (2015). Optimal full matching for survival outcomes: A method that merits more widespread use. *Stat. Med.* **34** 3949–3967. MR3431315

BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalized wreath products of permutation groups. *Proc. Lond. Math. Soc.* (3) **47** 69–82. MR0698928

BELL, C. B. and HALLER, H. S. (1969). Bivariate symmetry tests: Parametric and nonparametric. *Ann. Math. Stat.* **40** 259–269. MR0235658

BRANNATH, W., POSCH, M. and BAUER, P. (2002). Recursive combination tests. *J. Amer. Statist. Assoc.* **97** 236–244. MR1947283

BRIEN, C. J. and BAILEY, R. A. (2006). Multiple randomizations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 571–609. MR2301010

CENTERS FOR DISEASE CONTROL (2016). Smoking, gum disease, and tooth loss. Available at https://www.cdc.gov/tobacco/campaign/tips/diseases/periodontal-gum-disease.html.

COCHRAN, W. G. (1965). The planning of observational studies of human populations (with discussion). *J. Roy. Statist. Soc. Ser. A* **128** 234–266.

CONLON, J. C., LEON, R., PROSCHAN, F. and SETHURAMAN, J. (1977). G-Ordered functions, with applications in statistics. I, II. Technical Report M432, M433, Dept. Statistics, Florida State Univ. Tallahassee, FL. Available at http://www.dtic.mil/get-tr-doc/pdf?AD=ADA049316, http://www.dtic.mil/get-tr-doc/pdf?AD=ADA046584.

CORNFIELD, J., HAENSZEL, W., HAMMOND, E., LILIENFELD, A., SHIMKIN, M. and WYNDER, E. (1959). Smoking and lung cancer. *J. Nat. Cancer Inst.* **22** 173–203. Reprinted in *Internat. J. Epidemiol.* **38** (2009) 1175–1201. With discussion by D. R. Cox, J. Vandenbroucke, M. Zwahlen and J. B. Greenhouse.

COX, D. R. and REID, N. (2000). *The Theory of the Design of Experiments*. Chapman and Hall/CRC Press, London. DOI:10.1002/sim.1089.

DAWID, A. P. (1985). Invariance and independence in multivariate distribution theory. *J. Multivariate Anal.* **17** 304–315. MR0813238

DAWID, A. P. (1988). Symmetry models and hypotheses for structured data layouts. *J. Roy. Statist. Soc. Ser. B* **50** 1–34. MR0954729

DWASS, M. (1960). Some k-sample rank-order tests. In *Contributions to Probability and Statistics* 198–202. Stanford Univ. Press, Stanford, CA. MR0120705

Paul R. Rosenbaum is Robert G. Putzel Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6340, USA (e-mail: rosenbaum@wharton.upenn.edu).

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* **10** 11–43. MR0642717

EATON, M. L. and PERLMAN, M. D. (1977). Reflection groups, generalized Schur functions, and the geometry of majorization. *Ann. Probab.* **5** 829–860. MR0444864

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660

FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.

GASTWIRTH, J. L. (1992). Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables. *Jurimetrics* **33** 19–34.

GILBERT, N. D. and WAZZAN, S. (2008). Zappa–Szép products of bands and groups. *Semigroup Forum* **77** 438–455. MR2457329

HAMMOND, E. C. (1964). Smoking in relation to mortality and morbidity: Findings in first thirty-four months of follow-up in a prospective study started in 1959. *J. Natl. Cancer Inst.* **32** 1161–1188.

HANSEN, B. B. and KLOPFER, S. O. (2006). Optimal full matching and related designs via network flows. *J. Comput. Graph. Statist.* **15** 609–627. MR2280151

HOSMAN, C. A., HANSEN, B. B. and HOLLAND, P. W. (2010). The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *Ann. Appl. Stat.* **4** 849–870. MR2758424

HSU, J. Y., SMALL, D. S. and ROSENBAUM, P. R. (2013). Effect modification and design sensitivity in observational studies. *J. Amer. Statist. Assoc.* **108** 135–148. MR3174608

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York. MR0606374

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

ISAACS, I. M. (2009). *Algebra*: *A Graduate Course*. *Graduate Studies in Mathematics* **100**. Amer. Math. Soc., Providence, RI. Reprint of the 1994 original. MR2472787

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

LIU, W., KURAMOTO, S. J. and STUART, E. A. (2013). An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prev. Sci.* **14** 570–580.

MARDEN, J. I. (1992). Use of nested orthogonal contrasts in analyzing rank data. *J. Amer. Statist. Assoc.* **87** 307–318. MR1173802

MARITZ, J. S. (1979). A note on exact robust confidence intervals for location. *Biometrika* **66** 163–166. MR0529161

MCCANDLESS, L. C., GUSTAFSON, P. and LEVY, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat. Med.* **26** 2331–2347. MR2368419

PEIRCE, C. S. (1868). Some consequences of four incapacities. *J. Specul. Philos.* **2** 140–157. Reprinted in R. B. Talisse and S. F. Aikin, eds. (2011). *The Pragmatism Reader: From Peirce through the Present*. Harvard Univ. Press, Cambridge, MA.

PIMENTEL, S. D., YOON, F. and KEELE, L. (2015). Variable-ratio matching with fine balance in a study of the Peer Health Exchange. *Stat. Med.* **34** 4070–4082. MR3431322

RANDLES, R. H. and HOGG, R. V. (1971). Certain uncorrelated statistics and independent rank statistics. *J. Amer. Statist. Assoc.* **66** 569–574.

ROMAN, S. (2012). *Fundamentals of Group Theory. An Advanced Approach*. Birkhäuser/Springer, New York. MR2866265

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

ROSENBAUM, P. R. (1991). A characterization of optimal designs for observational studies. *J. Roy. Statist. Soc. Ser. B* **53** 597–610. MR1125717

ROSENBAUM, P. R. (1993). Hodges–Lehmann point estimates of treatment effect in observational studies. *J. Amer. Statist. Assoc.* **88** 1250–1253. MR1245357

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

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

ROSENBAUM, P. R. (2007). Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. *Biometrics* **63** 456–464. MR2370804

ROSENBAUM, P. R. (2010a). Evidence factors in observational studies. *Biometrika* **97** 333–345. MR2650742

ROSENBAUM, P. R. (2010b). Design sensitivity and efficiency in observational studies. *J. Amer. Statist. Assoc.* **105** 692–702. MR2724853

ROSENBAUM, P. R. (2011). Some approximate evidence factors in observational studies. *J. Amer. Statist. Assoc.* **106** 285–295. MR2816721

ROSENBAUM, P. R. (2013). Impact of multiple matched controls on design sensitivity in observational studies. *Biometrics* **69** 118–127. MR3058058

ROSENBAUM, P. R. (2015a). How to see more in observational studies: Some new quasi-experimental devices. *Ann. Rev. Statist. App.* **2** 21–48.

ROSENBAUM, P. R. (2015b). Two R packages for sensitivity analysis in observational studies. *Observ. Stud.* **1** 1–17.

ROSENBAUM, P. R. (2016a). Using Scheffé projections for multiple outcomes in an observational study of smoking and periodontal disease. *Ann. Appl. Stat.* **10** 1447–1471. MR3553231

ROSENBAUM, P. R. (2016b). The cross-cut statistic and its sensitivity to bias in observational studies with ordered doses of treatment. *Biometrics* **72** 175–183. MR3500586

ROSENBAUM, P. R. (2017). *Observation and Experiment*. Harvard Univ. Press, Cambridge, MA.

ROSENBAUM, P. R. and SILBER, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *J. Amer. Statist. Assoc.* **104** 1398–1405. MR2750570

ROSENBAUM, P. R. and SMALL, D. S. (2017). An adaptive Mantel–Haenszel test for sensitivity analysis in observational studies. *Biometrics* **73** 422–430. DOI:10.1111/biom.12591.

ROTMAN, J. J. (1995). *An Introduction to the Theory of Groups*, 4th ed. *Graduate Texts in Mathematics* **148**. Springer, New York. MR1307623

SHEPHERD, B. E., GILBERT, P. B., JEMIAI, Y. and ROTNITZKY, A. (2006). Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics* **62** 332–342. MR2236845

STUART, E. A. and GREEN, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Dev. Psychol.* **44** 395–406.

SUSSER, M. (1973). *Causal Thinking in the Health Sciences*: *Concepts and Strategies in Epidemiology*. Oxford Univ. Press, New York.

SUSSER, M. (1987). Falsification, verification and causal inference in epidemiology: Reconsideration in the light of Sir Karl Popper's philosophy. In *Epidemiology, Health and Society: Selected Papers* (M. Susser, ed.) 82–93. Oxford Univ. Press, New York.

SZÉP, J. (1950). On the structure of groups which can be represented as the product of two subgroups. *Acta Sci. Math.* (*Szeged*) **12** 57–61. MR0037296

TOMAR, S. L. and ASMA, S. (2000). Smoking-attributable periodontitis in the United States: Findings from NHANES III. *J. Periodont.* **71** 743–751.

WERFEL, U., LANGEN, V., EICKHOFF, I., SCHOONBROOD, J., VAHRENHOLZ, C., BRAUKSIEPE, A., POPP, W. and NORPOTH, K. (1998). Elevated DNA strand breakage frequencies in lymphocytes of welders exposed to chromium and nickel. *Carcinogenesis* **19** 413–418.

WOLFE, D. A. (1973). Some general results about uncorrelated statistics. *J. Amer. Statist. Assoc.* **68** 1013–1018. MR0418333

YANG, D., SMALL, D. S., SILBER, J. H. and ROSENBAUM, P. R. (2012). Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. *Biometrics* **68** 628–636. MR2959630

YU, B. B. and GASTWIRTH, J. L. (2005). Sensitivity analysis for trend tests: Application to the risk of radiation exposure. *Biostatistics* **6** 201–209.

ZAYKIN, D. V., ZHIVOTOVSKY, L. A., WESTFALL, P. H. and WEIR, B. S. (2002). Truncated product method for combining P-values. *Genet. Epidemiol.* **22** 170–185. DOI:10.1002/gepi.0042.

ZHANG, K., SMALL, D. S., LORCH, S., SRINIVAS, S. and ROSENBAUM, P. R. (2011). Using split samples and evidence factors in an observational study of neonatal outcomes. *J. Amer. Statist. Assoc.* **106** 511–524. MR2847966

ZUBIZARRETA, J. R., NEUMAN, M., SILBER, J. H. and ROSENBAUM, P. R. (2012). Contrasting evidence within and between institutions that provide treatment in an observational study of alternative forms of anesthesia. *J. Amer. Statist. Assoc.* **107** 901–915. MR3010879

Statistical Science
2017, Vol. 32, No. 4, 531–560
DOI: 10.1214/17-STS622

Hierarchical Sparse Modeling: A Choice of Two Group Lasso Formulations

Xiaohan Yan and Jacob Bien

Abstract. Demanding sparsity in estimated models has become a routine practice in statistics. In many situations, we wish to require that the sparsity patterns attained honor certain problem-specific constraints. *Hierarchical sparse modeling* (HSM) refers to situations in which these constraints specify that one set of parameters be set to zero whenever another is set to zero. In recent years, numerous papers have developed convex regularizers for this form of sparsity structure, which arises in many areas of statistics including interaction modeling, time series analysis, and covariance estimation. In this paper, we observe that these methods fall into two frameworks, the *group lasso* (GL) and *latent overlapping group lasso* (LOG), which have not been systematically compared in the context of HSM. The purpose of this paper is to provide a side-by-side comparison of these two frameworks for HSM in terms of their statistical properties and computational efficiency. We call special attention to GL's more aggressive shrinkage of parameters deep in the hierarchy, a property not shared by LOG. In terms of computation, we introduce a finite-step algorithm that exactly solves the proximal operator of LOG for a certain simple HSM structure; we later exploit this to develop a novel path-based block coordinate descent scheme for general HSM structures. Both algorithms greatly improve the computational performance of LOG. Finally, we compare the two methods in the context of covariance estimation, where we introduce a new sparsely-banded estimator using LOG, which we show achieves the statistical advantages of an existing GL-based method but is simpler to express and more efficient to compute.

Key words and phrases: Hierarchical sparsity, convex regularization, group lasso, latent overlapping group lasso.

REFERENCES

BACH, F. (2008). Exploring large feature spaces with hierarchical multiple kernel learning. In *Proceedings of the* 21*st International Conference on Neural Information Processing Systems*. *NIPS'*08 105–112. Curran Associates Inc., Red Hook, NY.

BACH, F., JENATTON, R., MAIRAL, J. and OBOZINSKI, G. (2012). Structured sparsity through convex optimization. *Statist. Sci.* **27** 450–468. MR3025128

BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. MR2486527

BERTSEKAS, D. P. (1999). *Nonlinear Programming*, 2nd ed. Athena Scientific, Belmont, MA. MR3444832

BIEN, J., BUNEA, F. and XIAO, L. (2016). Convex banding of the covariance matrix. *J. Amer. Statist. Assoc.* **111** 834–845. MR3538709

BIEN, J., TAYLOR, J. and TIBSHIRANI, R. (2013). A LASSO for hierarchical interactions. *Ann. Statist.* **41** 1111–1141. MR3113805

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

CHOI, N. H., LI, W. and ZHU, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *J. Amer.*

Xiaohan Yan is Ph.D. Candidate, Department of Statistical Science, Cornell University, Comstock Hall, Ithaca, New York 14853, USA (e-mail: xy257@cornell.edu). Jacob Bien is Assistant Professor, Department of Data Sciences and Operations, University of Southern California, Bridge Hall, Los Angeles, California 90089, USA (e-mail: jbien@usc.edu).

Statist. Assoc. **105** 354–364. MR2656056

CHOULDECHOVA, A. and HASTIE, T. (2015). Generalized additive model selection. Available at arXiv:1506.03850.

HARIS, A., WITTEN, D. and SIMON, N. (2016). Convex modeling of interactions with strong heredity. *J. Comput. Graph. Statist.* **25** 981–1004. MR3572025

JACOB, L., OBOZINSKI, G. and VERT, J. (2009). Group lasso with overlap and graph lasso. In *Proceedings of the* 26*th Annual International Conference on Machine Learning. ICML*'09 433–440. ACM, New York.

JENATTON, R., AUDIBERT, J.-Y. and BACH, F. (2011). Structured variable selection with sparsity-inducing norms. *J. Mach. Learn. Res.* **12** 2777–2824. MR2854347

JENATTON, R., MAIRAL, J., OBOZINSKI, G. and BACH, F. (2010). Proximal methods for sparse hierarchical dictionary learning. In *Proceedings of the* 27*th International Conference on International Conference on Machine Learning. ICML*'10 487–494. Omnipress, Madison, WI.

JENATTON, R., MAIRAL, J., OBOZINSKI, G. and BACH, F. (2011). Proximal methods for hierarchical sparse coding. *J. Mach. Learn. Res.* **12** 2297–2334. MR2825428

LEVINA, E., ROTHMAN, A. and ZHU, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Stat.* **2** 245–263. MR2415602

LIM, M. and HASTIE, T. (2015). Learning interactions via hierarchical group-lasso regularization. *J. Comput. Graph. Statist.* **24** 627–654. MR3397226

LOU, Y., BIEN, J., CARUANA, R. and GEHRKE, J. (2016). Sparse partially linear additive models. *J. Comput. Graph. Statist.* **25** 1026–1040. MR3572032

NELDER, J. A. (1977). A reformulation of linear models. *J. Roy. Statist. Soc. Ser. A* **140** 48–76. MR0458743

NESTEROV, YU. (2013). Gradient methods for minimizing composite functions. *Math. Program.* **140** 125–161. MR3071865

NICHOLSON, W. B., BIEN, J. and MATTESON, D. S. (2014). Hierarchical vector autoregression. Available at arXiv:1412.5250.

OBOZINSKI, G., JACOB, L. and VERT, J.-P. (2011). Group lasso with overlaps: The latent group lasso approach. Research report. Available at https://hal.inria.fr/inria-00628498.

RADCHENKO, P. and JAMES, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *J. Amer. Statist. Assoc.* **105** 1541–1553. MR2796570

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539–550. MR2672482

SCHMIDT, M. and MURPHY, K. (2010). Convex structure learning in log-linear models: Beyond pairwise potentials. In *Proceedings of the* 13*th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research* **9** 709–716.

SHE, Y., WANG, Z. and JIANG, H. (2017). Group regularized estimation under structural hierarchy. *J. Amer. Statist. Assoc.* To appear. DOI:10.1080/01621459.2016.1260470.

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2013). A sparse-group lasso. *J. Comput. Graph. Statist.* **22** 231–245. MR3173712

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

TSENG, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** 475–494. MR1835069

TURLACH, B. A., VENABLES, W. N. and WRIGHT, S. J. (2005). Simultaneous variable selection. *Technometrics* **47** 349–363. MR2164706

VILLA, S., ROSASCO, L., MOSCI, S. and VERRI, A. (2014). Proximal methods for the latent group lasso penalty. *Comput. Optim. Appl.* **58** 381–407. MR3201966

YUAN, M., JOSEPH, V. R. and ZOU, H. (2009). Structured variable selection and estimation. *Ann. Appl. Stat.* **3** 1738–1757. MR2752156

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. Ser. B* **68** 49–67. MR2212574

ZHAO, P., ROCHA, G. and YU, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.* **37** 3468–3497. MR2549566

Statistical Science
2017, Vol. 32, No. 4, 561–579
DOI: 10.1214/17-STS623

Instrumental Variable Estimation with a Stochastic Monotonicity Assumption

Dylan S. Small, Zhiqiang Tan, Roland R. Ramsahai, Scott A. Lorch and M. Alan Brookhart

Abstract. The instrumental variables (IV) method provides a way to estimate the causal effect of a treatment when there are unmeasured confounding variables. The method requires a valid IV, a variable that is independent of the unmeasured confounding variables and is associated with the treatment but which has no effect on the outcome beyond its effect on the treatment. An additional assumption often made is deterministic monotonicity, which says that for each subject, the level of the treatment that a subject would take is a monotonic increasing function of the level of the IV. However, deterministic monotonicity is sometimes not realistic. We introduce a stochastic monotonicity assumption, a relaxation that only requires a monotonic increasing relationship to hold across subjects between the IV and the treatments conditionally on a set of (possibly unmeasured) covariates. We show that under stochastic monotonicity, the IV method identifies a weighted average of treatment effects with greater weight on subgroups of subjects on whom the IV has a stronger effect. We provide bounds on the global average treatment effect under stochastic monotonicity and a sensitivity analysis for violations of stochastic monotonicity. We apply the methods to a study of the effect of premature babies being delivered in a high technology neonatal intensive care unit (NICU) vs. a low technology unit.

Key words and phrases: Causal inference, observational study, instrumental variable, two stage least squares.

REFERENCES

ABADIE, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *J. Econometrics* **113** 231–263. MR1960380

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

ANGRIST, J. D. and PISCHKE, J.-S. (2009). *Mostly Harmless Econometrics*: *An Empiricist's Companion*. Princeton Univ. Press, Princeton, NJ.

BAGNOLI, M. and BERGSTROM, T. (2005). Log-concave probability and its applications. *Econom. Theory* **26** 445–469. MR2213177

BAIOCCHI, M., SMALL, D. S., LORCH, S. and ROSENBAUM, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *J. Amer. Statist. Assoc.* **105** 1285–1296. MR2796550

BALKE, A. and PEARL, J. (1997). Bounds on treatment effects for studies with imperfect compliance. *J. Amer. Statist. Assoc.* **92** 1171–1176.

BOEF, A. G. C., GUSSEKLOO, J., DEKKERS, O. M., FREY, P., KEARNEY, P. M., KERSE, N., MALLEN, C. D., MCCARTHY, V. J. C., MOOIJAART, S. P., MUTH, C., RODONDI, N., ROSEMANN, T., RUSSELL, A., SCHERS, H., VIRGINI, V., DE WAAL, M. W. M., WARNER, A., LE CESSIE, S. and DEN ELZEN, W. P. J. (2016). Physician's prescribing preference as an instrumental variable: Exploring as-

Dylan S. Small is Class of 1965 Wharton Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu). Zhiqiang Tan is Professor, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854, USA (e-mail: ztan@stat.rutgers.edu). Roland R. Ramsahai is Vice President at Reinsurance Division of Berkshire Hathaway Group, Stamford, CT 06902 (e-mail: roland.ramsahai@hotmail.com). Scott A. Lorch is Associate Professor, Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: LORCH@email.chop.edu). M. Alan Brookhart is Professor, Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27599, USA (e-mail: mabrook@email.unc.edu).

sumptions using survey data. *Epidemiology* **27** 276–283.

BROOKHART, M. A. and SCHNEEWEISS, S. (2007). Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int. J. Biostat.* **3** Art. 14, 25. MR2383610

BROOKHART, M. A., WANG, P., SOLOMON, D. H. and SCHNEEWEISS, S. (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* **17** 268–275.

CAVOLINA, M. J. F., KELLY, M. A. T., STONE, J. A. J. and DAVIS, R. G. M. (2000). *Growing up Catholic*: *The Millennium Edition*: *An Infinitely Funny Guide for the Faithful*, *the Fallen and Everyone in-Between*. Image.

CHALAK, K. (2017). Instrumental variables methods with heterogeneity and mismeasured instruments. *Econometric Theory* **33** 69–104. MR3574861

CHENG, J. and SMALL, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 815–836. MR2301296

COX, D. R. (1958). *Planning of Experiments*. Wiley, New York. MR0095561

DE CHAISEMARTIN, C. (2017). Tolerating defiance: Local average treatment effects without monotonicity. *Quant. Econ.* **8** 367–396.

DEATON, A. (2010). Instruments, randomization and learning about development. *J. Econ. Lit.* **48** 424–455.

DIJKSTRA, E. W. (1959). A note on two problems in connexion with graphs. *Numer. Math.* **1** 269–271. MR0107609

DINARDO, J. and LEE, D. S. (2011). Program evaluation and research designs. In *Handbook of Labor Economics*, *Vol.* 4 463–536. Elsevier, Amsterdam.

FINKELSTEIN, A., TAUBMAN, S., WRIGHT, B., BERNSTEIN, M., GRUBER, J., NEWHOUSE, J. P., ALLEN, H., BAICKER, K. and GROUP OREGON HEALTH STUDY (2012). The Oregon health insurance experiment: Evidence from the first year. *Q. J. Econ.* **127** 1057–1106.

GUO, Z., CHENG, J., LORCH, S. A. and SMALL, D. S. (2014). Using an instrumental variable to test for unmeasured confounding. *Stat. Med.* **33** 3528–3546. MR3260644

HERNÁN, M. A. and ROBINS, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology* **17** 360–372.

HOROWITZ, J. L. and MANSKI, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Amer. Statist. Assoc.* **95** 77–88. MR1803142

HUBER, M. and MELLACE, G. (2012). Relaxing monotonicity in the identification of local average treatment effects. Working paper.

IMBENS, G. W. (2014). Instrumental variables: An econometrician's perspective. *Statist. Sci.* **29** 323–358. MR3264545

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **61** 467–476.

KLEIN, T. J. (2010). Heterogeneous treatment effects: Instrumental variables without monotonicity? *J. Econometrics* **155** 99–116. MR2607188

KORN, E. L. and BAUMRIND, S. (1998). Clinician preferences and the estimation of causal treatment differences. *Statist. Sci.* **13** 209–235. MR1665709

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153. MR0202228

LORCH, S. A., BAIOCCHI, M., AHLBERG, C. E. and SMALL, D. S. (2012a). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics* **130** 270–278.

LORCH, S. A., KROELINGER, C. D., AHLBERG, C. and BARFIELD, W. D. (2012b). Factors that mediate racial/ethnic disparities in US fetal death rates. *Am. J. Publ. Health* **102** 1902–1910.

MANSKI, C. F. (1990). Non-parametric bounds on treatment effects. *Am. Econ. Rev.* **80** 351–374.

MCCLELLAN, M., MCNEIL, B. J. and NEWHOUSE, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *J. Am. Med. Dir. Assoc.* **272** 859–866.

NEAL, D. (1997). The effects of Catholic secondary schooling on educational achievement. *J. Labor Econ.* **14** 98–123.

OGBURN, E. L., ROTNITZKY, A. and ROBINS, J. M. (2015). Doubly robust estimation of the local average treatment effect curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 373–396. MR3310531

PEARL, J. (2011). Principal stratification—a goal or a tool? *Int. J. Biostat.* **7** Art. 20, 15. MR2787410

PHIBBS, C. S., MARK, D. H., LUFT, H. S., PELTZMAN-RENNIE, D. J., GARNICK, D. W., LICHTENBERG, E. and MCPHEE, S. J. (1993). Choice of hospital for delivery: A comparison of high-risk and low-risk women. *Health Serv. Res.* **28** 201.

RAMSAHAI, R. R. (2012). Causal bounds and observable constraints for non-deterministic models. *J. Mach. Learn. Res.* **13** 829–848. MR2913720

ROBINS, J. M. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology*: *A Focus on AIDS* (L. Sechrest, H. Freeman and A. Mulley, eds.) 113–159.

ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185

ROSENBAUM, P. R. (2010). *Design of Observational Studies*. Springer, New York. MR2561612

RUBIN, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *J. Amer. Statist. Assoc.* **81** 961–962.

SMALL, D. S., TAN, Z., RAMSAHAI, R. R., LORCH, S. A. and BROOKHART, M. A. (2017). Supplement to "Instrumental variable estimation with a stochastic monotonicity assumption." DOI:10.1214/17-STS623SUPP.

STOCK, J. H. (2001). Instrumental variables in economics and statistics. In *International Encyclopedia of the Social & Behavioral Sciences* (N. J. Smelser and P. B. Baltes, eds.) 7577–7582. Elsevier, Amsterdam.

SWANSON, S. A. and HERNÁN, M. A. (2014). Think globally, act globally: An epidemiologist's perspective on instrumental variable estimation [discussion of MR3264545]. *Statist. Sci.* **29** 371–374. MR3264549

SWANSON, S. A., MILLER, M., ROBINS, J. M. and HERNÁN, M. A. (2015). Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology* **26** 414–420.

TAN, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *J. Amer. Statist. Assoc.* **101** 1607–1618. MR2279483

TAN, Z. (2010). Marginal and nested structural models using instrumental variables. *J. Amer. Statist. Assoc.* **105** 157–169. MR2757199

VANDERWEELE, T. J. and SHPITSER, I. (2013). On the definition of a confounder. *Ann. Statist.* **41** 196–220. MR3059415

YANG, F., LORCH, S. A. and SMALL, D. S. (2014). Estimation of causal effects using instrumental variables with nonignorable missing covariates: Application to effect of type of delivery NICU on premature infants. *Ann. Appl. Stat.* **8** 48–73. MR3191982

Statistical Science
2017, Vol. 32, No. 4, 580–599
DOI: 10.1214/17-STS624

The Coordinate-Based Meta-Analysis of Neuroimaging Data

Pantelis Samartsidis, Silvia Montagna, Timothy D. Johnson and Thomas E. Nichols

Abstract. Neuroimaging meta-analysis is an area of growing interest in statistics. The special characteristics of neuroimaging data render classical meta-analysis methods inapplicable and therefore new methods have been developed. We review existing methodologies, explaining the benefits and drawbacks of each. A demonstration on a real dataset of emotion studies is included. We discuss some still-open problems in the field to highlight the need for future research.

Key words and phrases: Meta-analysis, neuroimaging, functional magnetic resonance imaging.

REFERENCES

ARMINGER, G. and MUTHÉN, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63** 271–300.

BAILEY, D. L., TOWNSEND, D. W., VALK, P. E. and MAISEY, M. N., eds. (2006). *Positron Emission Tomography*: *Basic Sciences*. Springer, Berlin.

BARTELS, A. and ZEKI, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage* **21** 1155–1166.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.* **57** 289–300. MR1325392

BUTTON, K. S., IOANNIDIS, J. P. A., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. J. and MUNAFÒ, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev., Neurosci.* **14** 365–76.

CALHOUN, V. D. and PEARLSON, G. D. (2012). A selective review of simulated driving studies: Combining naturalistic and hybrid paradigms, analysis approaches, and future directions. *NeuroImage* **59** 25–35.

CARP, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage* **63** 289–300.

CASPERS, J., ZILLES, K., BEIERLE, C., ROTTSCHY, C. and EICKHOFF, S. B. (2014). A novel meta-analytic approach: Mining frequent co-activation patterns in neuroimaging databases. *Neuroimage* **90** 390–402.

COLE, D. M., SMITH, S. M. and BECKMANN, C. F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Frontiers in Systems Neuroscience* **4** 8.

COSTAFREDA, S. G., DAVID, A. S. and BRAMMER, M. J. (2009). A parametric approach to voxel-based meta-analysis. *NeuroImage* **46** 115–122.

COSTAFREDA, S., BRAMMER, M., DAVID, A. and FU, C. (2008). Predictors of amygdala activation during the processing of emotional stimuli. *Brains Res. Rev.* **58** 57–70.

DAVID, S. P., WARE, J. J., CHU, I. M., LOFTUS, P. D., FUSAR-POLI, P., RADUA, J., MUNAFÒ, M. R. and IOANNIDIS, J. P. A. (2013). Potential reporting bias in fMRI studies of the brain. *PLoS ONE* **8**.

DELVECCHIO, G., FOSSATI, P., BOYER, P., BRAMBILLA, P., FALKAI, P., GRUBER, O., HIETALA, J., LAWRIE, S. M., MARTINOT, J.-L., MCINTOSH, A. M., MEISENZAHL, E. and FRANGOU, S. (2012). Common and distinct neural correlates of emotional processing in bipolar disorder and major depressive disorder: A voxel-based meta-analysis of functional magnetic resonance imaging studies. *Eur. Neuropsychopharmacol.* **22** 100–113.

EICKHOFF, S. B., LAIRD, A. R., GREFKES, C., WANG, L. E., ZILLES, K. and FOX, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* **30** 2907–2926.

EICKHOFF, S. B., BZDOK, D., LAIRD, A. R., KURTH, F. and FOX, P. T. (2012). Activation likelihood estimation meta-

Pantelis Samartsidis is Investigator Statistician, MRC Biostatistics Unit, University of Cambridge, Robinson Way, Cambridge, CB2 0SR, United Kingdom (e-mail: pantelis.samartsidis@mrc-bsu.cam.ac.uk). Silvia Montagna is Lecturer, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7FS, United Kingdom (e-mail: s.montagna@kent.ac.uk). Timothy D. Johnson is Professor, Biostatistics Department, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: tdjtdj@umich.edu). Thomas E. Nichols is Professor, Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, OX3 7LF, United Kingdom (e-mail: thomas.nichols@bdi.ox.ac.uk).

analysis revisited. *NeuroImage* **59** 2349–2361.

EICKHOFF, S. B., NICHOLS, T. E., LAIRD, A. R., HOFFSTAEDTER, F., AMUNTS, K., FOX, P. T., BZDOK, D. and EICKHOFF, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage* **137** 70–85.

EICKHOFF, S. B., LAIRD, A. R., FOX, P. M., LANCASTER, J. L. and FOX, P. T. (2017). Implementation errors in the GingerALE software: Description and recommendations. *Hum. Brain Mapp.* **38** 7–11.

ETKIN, A. and WAGER, T. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164** 1476–1488.

FARAH, M. J. (2014). Brain images, babies, and bathwater: Critiquing critiques of functional neuroimaging. *Hastings Cent. Rep.* **44** S19–30.

FOX, P. T., PARSONS, L. M. and LANCASTER, J. L. (1998). Beyond the single study: Function/location metanalysis in cognitive neuroimaging. *Curr. Opin. Neurobiol.* **8** 178–187.

FOX, P. T., LANCASTER, J. L., PARSONS, L. M., XIONG, J. H. and ZAMARRIPA, F. (1997). Functional volumes modeling: Theory and preliminary assessment. *Hum. Brain Mapp.* **5** 306–311.

FRISTON, K. J., PENNY, W., PHILLIPS, C., KIEBEL, S., HINTON, G. and ASHBURNER, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage* **16** 465–483.

FUSAR-POLI, P. (2012). Voxel-wise meta-analysis of fMRI studies in patients at clinical high risk for psychosis. *J. Psychiatry Neurosci.* **37** 106–112.

GENOVESE, C., LAZAR, N. and NICHOLS, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15** 870–878.

GORGOLEWSKI, K. J., VAROQUAUX, G., RIVERA, G., SCHWARZ, Y., GHOSH, S. S., MAUMET, C., SOCHAT, V. V., NICHOLS, T. E., POLDRACK, R. A., POLINE, J.-B., YARKONI, T. and MARGULIES, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinform* **9** 8.

GREENLAND, S. (1994). Invited commentary: A critical look at some popular meta-analytic methods. *Am. J. Epidemiol.* **140** 290–296.

HARTUNG, J., KNAPP, G. and SINHA, B. K. (2008). *Statistical Meta-Analysis with Applications*. Wiley, Hoboken, NJ. MR2435836

HEDGES, L. V. and OLKIN, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Inc., Orlando, FL. MR0798597

HUETTEL, S. A., SONG, A. W. and MCCARTHY, G. (2009). *Functional Magnetic Resonance Imaging Second Edition*. Sinauer Associates, Inc, Sunderland, MA.

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

KANG, J., JOHNSON, T. D., NICHOLS, T. E. and WAGER, T. D. (2011). Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *J. Amer. Statist. Assoc.* **106** 124–134. MR2816707

KANG, J., NICHOLS, T. E., WAGER, T. D. and JOHNSON, T. D. (2014). A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. *Ann. Appl. Stat.* **8** 1800–1824. MR3271354

KIM, S.-G. and OGAWA, S. (2012). Biophysical and physiological origins of blood oxygenation level-dependent fMRI signals. *J. Cereb. Blood Flow Metab.* **32** 1188–1206.

KOBER, H., BARRETT, L. F., JOSEPH, J., BLISS-MOREAU, E., LINDQUIST, K. and WAGER, T. D. (2008). Functional grouping and cortical and subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage* **42** 998–1031.

KONOVA, A. B., MOELLER, S. J. and GOLDSTEIN, R. Z. (2013). Common and distinct neural targets of treatment: Changing brain function in substance addiction. *Neurosci. Biobehav. Rev.* **37** 2806–2817.

LAIRD, A. R., LANCASTER, J. J. and FOX, P. T. (2005). BrainMap: The social evolution of a human brain mapping database. *Neuroinformatics* **3** 65–77.

LAIRD, A. R., FOX, P. M., PRICE, C. J., GLAHN, D. C., UECKER, A. M., LANCASTER, J. L., TURKELTAUB, P. E., KOCHUNOV, P. and FOX, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Hum. Brain Mapp.* **25** 155–164.

LAZAR, N. A., LUNA, B., SWEENEY, J. A. and EDDY, W. F. (2002). Combining brains: A survey of methods for statistical pooling of information. *NeuroImage* **16** 538–550.

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545

MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. **100**. Chapman & Hall/CRC, Boca Raton, FL. MR2004226

MONTAGNA, S., TOKDAR, S. T., NEELON, B. and DUNSON, D. B. (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics* **68** 1064–1073. MR3040013

MONTAGNA, S., WAGER, T. D., BARRETT, L. F., JOHNSON, T. D. and NICHOLS, T. E. (2017). Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data. *Biometrics*. To appear.

MUMFORD, J. A. and NICHOLS, T. (2006). Modeling and inference of multisubject fMRI data. *IEEE Eng. Med. Biol. Mag.* **25** 42–51.

MUMFORD, J. A. and NICHOLS, T. (2009). Simple group fMRI modeling and inference. *NeuroImage* **47** 1469–1475.

NICHOLS, T. and HAYASAKA, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Stat. Methods Med. Res.* **12** 419–446. MR2005445

NICHOLS, T. E. and HOLMES, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.* **15** 1–25.

NIELSEN, F. Å. and HANSEN, L. K. (2002). Modeling of activation data in the BrainMap™ database: Detection of outliers. *Hum. Brain Mapp.* **15** 146–156.

PHELPS, E. A. and LEDOUX, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron* **48** 175–187.

POLDRACK, R. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron* **72** 692–697.

POLDRACK, R. A., MUMFORD, J. A. and NICHOLS, T. E. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge Univ. Press, Cambridge. MR2839490

POLDRACK, R. A., BARCH, D. M., MITCHELL, J. P., WAGER, T. D., WAGNER, A. D., DEVLIN, J. T., CUMBA, C., KOYEJO, O. and MILHAM, M. P. (2013). Toward open sharing of task-based fMRI data: The OpenfMRI project. *Front. Neuroinform* **7** 12.

POLINE, J.-B., BREEZE, J. L., GHOSH, S., GORGOLEWSKI, K., HALCHENKO, Y. O., HANKE, M., HASELGROVE, C., HELMER, K. G., KEATOR, D. B., MARCUS, D. S., POLDRACK, R. A., SCHWARTZ, Y., ASHBURNER, J. and KENNEDY, D. N. (2012). Data sharing in neuroimaging research. *Front. Neuroinform* **6** 9.

RADUA, J. and MATAIX-COLS, D. (2009). Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *Br. J. Psychiatry* **195** 393–402.

RADUA, J. and MATAIX-COLS, D. (2012). Meta-analytic methods for neuroimaging data explained. *Biol. Mood Anxiety Disord.* **2** 6.

RADUA, J., MATAIX-COLS, D., PHILLIPS, M. L., EL-HAGE, W., KRONHAUS, D. M., CARDONER, N. and SURGULADZE, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur. Psychiatry* **27** 605–611.

RADUA, J., RUBIA, K., CANALES-RODRÍGUEZ, E. J., POMAROL-CLOTET, E., FUSAR-POLI, P. and MATAIX-COLS, D. (2014). Anisotropic kernels for coordinate-based meta-analyses of neuroimaging studies. *Front. Psychiatry* **5** 13.

RAEMAEKERS, M., VINK, M., ZANDBELT, B., VAN WEZEL, R. J. A., KAHN, R. S. and RAMSEY, N. F. (2007). Test-retest reliability of fMRI activation during prosaccades and antisaccades. *NeuroImage* **36** 532–542.

RICHLAN, F., KRONBICHLER, M. and WIMMER, H. (2011). Meta-analyzing brain dysfunctions in dyslexic children and adults. *NeuroImage* **56** 1735–1742.

SALIMI-KHORSHIDI, G., SMITH, S. M., KELTNER, J. R., WAGER, T. D. and NICHOLS, T. E. (2009). Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. *NeuroImage* **45** 810–823.

SAMARTSIDIS, P., MONTAGNA, S., NICHOLS, T. E. and JOHNSON, T. D. (2017). Supplement to "The coordinate-based meta-analysis of neuroimaging data." DOI:10.1214/17-STS624SUPP.

SHERMER, M. (2008). Why you should be skeptical of brain scans. *Sci. Am. Mind* **19** 66–71.

TALAIRACH, J. and TOURNOUX, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme, Stuttgart.

THIRION, B., PINEL, P., MERIAUX, S., ROCHE, A., DEHAENE, S. and POLINE, J.-B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage* **35** 105–120.

TURKELTAUB, P. E., EDEN, G. F., JONES, K. M. and ZEFFIRO, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage* **16** 765–780.

TURKELTAUB, P. E., EICKHOFF, S. B., LAIRD, A. R., FOX, M., WIENER, M. and FOX, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* **33** 1–13.

VUL, E., HARRIS, C., WINKIELMAN, P. and PASHLER, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* **4** 274–290.

WAGER, T. D., JONIDES, J. and READING, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *NeuroImage* **22** 1679–1693.

WAGER, T. D., LINDQUIST, M. and KAPLAN, L. (2007). Meta-analysis of functional neuroimaging data: Current and future directions. *Soc. Cogn. Affect Neurosci.* **2** 150–158.

WAGER, T. D., PHAN, K. L., LIBERZON, I. and TAYLOR, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: A meta-analysis of findings from neuroimaging. *NeuroImage* **19** 513–531.

WAGER, T. D., BARRETT, L. F., BLISS-MOREAU, E., LINDQUIST, K., DUNCAN, S., KOBER, H., JOSEPH, J., DAVIDSON, M. and MIZE, J. (2008). The neuroimaging of emotion. In *Handbook of Emotions*, 3rd ed. (M. Lewis, J. M. Haviland-Jones and L. F. Barrett, eds.) 249–271. Guilford Press, New York, NY.

WOLPERT, R. L. and ICKSTADT, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267. MR1649114

XUE, W., KANG, J., BOWMAN, F. D., WAGER, T. D. and GUO, J. (2014). Identifying functional co-activation patterns in neuroimaging studies via Poisson graphical models. *Biometrics* **70** 812–822. MR3295742

YARKONI, T., POLDRACK, R. A., NICHOLS, T. E., VAN ESSEN, D. C. and WAGER, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8** 665–670.

YUE, Y. R., LINDQUIST, M. A. and LOH, J. M. (2012). Meta-analysis of functional neuroimaging data using Bayesian nonparametric binary regression. *Ann. Appl. Stat.* **6** 697–718. MR2976488

YUE, Y. and SPECKMAN, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *J. Comput. Graph. Statist.* **19** 96–116. MR2654402

Statistical Science
2017, Vol. 32, No. 4, 600–615
DOI: 10.1214/17-STS627

Contemporary Frequentist Views of the 2 x 2 Binomial Trial

Enrico Ripamonti, Chris Lloyd and Piero Quatto

Abstract. The 2×2 table is the simplest of data structures yet it is of immense practical importance. It is also just complex enough to provide a theoretical testing ground for general frequentist methods. Yet after 70 years of debate, its correct analysis is still not settled. Rather than recount the entire history, our review is motivated by contemporary developments in likelihood and testing theory as well as computational advances. We will look at both conditional and unconditional tests. Within the conditional framework, we explain the relationship of Fisher's test with variants such as mid-p and Liebermeister's test, as well as modern developments in likelihood theory, such as p^* and approximate conditioning. Within an unconditional framework, we consider four modern methods of correcting approximate tests to properly control size by accounting for the unknown value of the nuisance parameter: maximisation (M), partial maximisation (B), estimation (E) and estimation followed by maximisation (E + M). Under the conditional model, we recommend Fisher's test. For the unconditional model, amongst standard approximate methods, Liebermeister's tests come closest to controlling size. However, our best recommendation is the E procedure applied to the signed root likelihood statistic, as this performs very well in terms of size and power and is easily computed. We support our assertions with a numerical study.

Key words and phrases: Approximate conditioning, binomial trial, conditional test, exact tests, Fisher test, Liebermeister test, mid-p test, parametric bootstrap, unconditional test.

REFERENCES

AGRESTI, A. (1992). A survey of exact inference for contingency tables. *Statist. Sci.* **7** 131–177. With comments and a rejoinder by the author. MR1173420

AGRESTI, A. (2001). Exact inference for categorica data: Recent advances and continuing controversies. *Stat. Med.* **20** 2709–2722.

AGRESTI, A. (2002). *Categorical Data Analysis*. Wiley, Hoboken, NJ. MR3087436

BARNARD, G. A. (1945). A new test for 2×2 tables. *Nature* **156** 177. MR0013274

BARNARD, G. A. (1947). Significance tests for 2×2 tables. *Biometrika* **34** 123–138. MR0019285

BARNDORFF-NIELSEN, O. (1973). On M-ancillarity. *Biometrika* **60** 447–455. MR0345255

BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023

BASU, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* **72** 355–366. MR0451477

BERGER, R. L. and BOOS, D. D. (1994). P values maximized over a confidence set for the nuisance parameter. *J. Amer. Statist. Assoc.* **89** 1012–1016. MR1294746

BERGER, R. L. and SIDIK, K. (2003). Exact unconditional tests for a 2×2 matched-pairs design. *Stat. Methods Med. Res.* **12** 91–108. MR1963335

BERKSON, J. (1978). In dispraise of the exact test: Do the marginal totals of the 2×2 table contain relevant information respecting

Enrico Ripamonti is Postdoc Researcher, Department of Economics, Management, and Statistics, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy (e-mail: enrico.ripamonti@unimib.it). Chris Lloyd is Professor, Melbourne Business School, The University of Melbourne, 200 Leicester Street, Carlton Victoria 3053, Australia (e-mail: c.lloyd@mbs.edu). Piero Quatto is Professor, Department of Economics, Management, and Statistics, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126, Milan, Italy (e-mail: piero.quatto@unimib.it).

the table proportions? *J. Statist. Plann. Inference* **2** 27–42.
BOSCHLOO, R. D. (1970). Raised conditional level of significance for the 2×2-table when testing the equality of two probabilities. *Stat. Neerl.* **24** 1–35. MR0264827
BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. MR1892660
BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30** 160–201. MR1892660
CHOI, L., BLUME, J. D. and DUPONT, W. D. (2015). Elucidating the foundations of statistical inference with 2×2 tables. *PLoS ONE* **10** e0121263.
COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286. MR0581725
COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London. MR0370837
DAVISON, A. C., FRASER, D. A. S. and REID, N. (2006). Improved likelihood inference for discrete data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 495–508. MR2278337
FINNER, H. and STRASSBURGER, K. (2002). Structural properties of UMPU-tests for 2×2 tables and some applications. *J. Statist. Plann. Inference* **104** 103–120. MR1900521
FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, London.
GAIL, M. H. and GART, J. J. (1973). The determination of sample sizes for use with the exact conditional test in 2×2 comparative trials. *Biometrics* **29** 441–448.
GART, J. J. (1969). An exact test for comparing matched proportions in crossover designs. *Biometrika* **56** 75–80.
GODAMBE, V. P. (1980). On sufficiency and ancillarity in the presence of a nuisance parameter. *Biometrika* **67** 155–162. MR0570517
HARRIS, B. and SOMS, A. P. (1991). Theory and counterexamples for confidence limits on system reliability. *Statist. Probab. Lett.* **11** 411–417. MR1114531
HIRJI, K. F. (2006). *Exact Analysis of Discrete Data*. Chapman & Hall/CRC, Boca Raton, FL. MR2193238
HIRJI, K. F., MEHTA, C. R. and PATEL, N. R. (1988). Exact inference for matched case-control studies. *Biometrics* **44** 803–814. MR0963915
HIRJI, K. F., TAN, S. J. and ELASHOFF, R. M. (1991). A quasi-exact test for comparing two binomial proportions. *Stat. Med.* **10** 1137–1153.
HOWARD, J. V. (1998). The 2×2 table: A discussion from a Bayesian viewpoint. *Statist. Sci.* **13** 351–367. MR1705267
HWANG, J. T. G. and YANG, M.-C. (2001). An optimality theory for mid p-values in 2×2 contingency tables. *Statist. Sinica* **11** 807–826. MR1863164
KABAILA, P. (2005). Computation of exact confidence limits from discrete data. *Comput. Statist.* **20** 401–414. MR2236616
KABAILA, P. and LLOYD, C. J. (2006). Improved Buehler limits based on refined designated statistics. *J. Statist. Plann. Inference* **136** 3145–3155. MR2256221
LANCASTER, H. O. (1961). Significance tests in discrete distributions. *J. Amer. Statist. Assoc.* **56** 223–234. MR0124107
LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. MR0107933
LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927
LIEBERMEISTER, C. (1877). *Über Wahrscheinlichkeitsrechnung in Anwendung Auf Therapeutische Statistik*. Breitkiof and Härtel.
LLOYD, C. J. (2008a). A new exact and more powerful unconditional test of no treatment effect from binary matched pairs. *Biometrics* **64** 716–723. MR2526621
LLOYD, C. J. (2008b). Exact P-values for discrete models obtained by estimation and maximization. *Aust. N. Z. J. Stat.* **50** 329–345. MR2474195
LLOYD, C. J. (2010a). P-values based on approximate conditioning and p^*. *J. Statist. Plann. Inference* **140** 1073–1081. MR2574669
LLOYD, C. J. (2010b). Bootstrap and second-order tests of risk difference. *Biometrics* **66** 975–982. MR2758234
LLOYD, C. J. (2012). Computing highly accurate or exact P-values using importance sampling. *Comput. Statist. Data Anal.* **56** 1784–1794. MR2892377
LYDERSEN, S., FAGERLAND, M. W. and LAAKE, P. (2009). Recommended tests for association in 2×2 tables. *Stat. Med.* **28** 1159–1175. MR2662203
MARTIN ANDRES, A. (1991). A review of classic non-asymptotic methods for comparing two proportions by means of independent samples. *Comm. Statist. Simulation Comput.* **20** 551–583.
MCDONALD, L. L., DAVIS, B. M. and MILLIKEN, G. A. (1977). A nonrandomized unconditional test for comparing two proportions in 2×2 contingency tables. *Technometrics* **19** 145–158.
MEHROTRA, D. V., CHAN, I. S. F. and BERGER, R. L. (2003). A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics* **59** 441–450. MR1982586
MEHTA, C. R. and HILTON, J. F. (1993). Exact power of conditional and unconditional tests: Going beyond the 2×2 contingency table. *Amer. Statist.* **47** 91–98.
PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **5** 157–175.
PIERCE, D. A. and PETERS, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 701–737. MR1185218
PIERCE, D. A. and PETERS, D. (1999). Improving on exact tests by approximate conditioning. *Biometrika* **86** 265–277. MR1705363
RIPAMONTI, E., LLOYD, C. and QUATTO, P. (2017). Supplement to "Contemporary frequentist views of the 2×2 binomial trial." DOI:10.1214/17-STS627SUPP.
RÖHMEL, J. (2005). Problems with existing procedures to calculate exact unconditional p-values for non-inferiority/superiority and confidence intervals for two binomials and how to resolve them. *Biom. J.* **47** 37–47. MR2135888
RÖHMEL, J. and MANSMANN, U. (1999). Unconditional non-asymptotic one-sided tests for independent binomial proportions when the interest lies in showing non-inferiority and/or superiority. *Biom. J.* **41** 149–170. MR1693980
SENETA, E. and PHIPPS, M. C. (2001). On the comparison of two observed frequencies. *Biom. J.* **43** 23–43. MR1820037
SKIPKA, G., MUNK, A. and FREITAG, G. (2004). Unconditional exact tests for the difference of binomial probabilities—contrasted and compared. *Comput. Statist. Data Anal.* **47** 757–773. MR2101550

STORER, B. E. and KIM, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *J. Amer. Statist. Assoc.* **85** 146–155.

TOCHER, K. D. (1950). Extension of the Neyman–Pearson theory of tests to discontinuous variates. *Biometrika* **37** 130–144. MR0036972

WELLS, M. T. (2010). Optimality results for mid p-values. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. (IMS) Collect.* **6** 184–198. IMS, Beachwood, OH. MR2798519

YATES, F. (1984). Tests of significance for 2×2 contingency tables. *J. Roy. Statist. Soc. Ser. A* **147** 426–463. MR0769998

Statistical Science
2017, Vol. 32, No. 4, 616–629
DOI: 10.1214/17-STS628

Elo Ratings and the Sports Model: A Neglected Topic in Applied Probability?

David Aldous

Abstract. In a simple model for sports, the probability A beats B is a specified function of their difference in strength. One might think this would be a staple topic in Applied Probability textbooks (like the Galton–Watson branching process model, for instance) but it is curiously absent. Our first purpose is to point out that the model suggests a wide range of questions, suitable for "undergraduate research" via simulation but also challenging as professional research. Our second, more specific, purpose concerns Elo-type rating algorithms for tracking changing strengths. There has been little foundational research on their accuracy, despite a much-copied "30 matches suffice" claim, which our simulation study casts doubt upon.

Key words and phrases: Elo rating, Bradley–Terry model, dynamic ratings, sports forecasting.

REFERENCES

ADLER, I., CAO, Y., KARP, R., PEKOZ, E. and ROSS, S. M. (2016). Random knockout tournaments. *Oper. Res.* To appear. Available at arXiv:1612.04448.

ALDOUS, D. J. (2013). Using prediction market data to illustrate undergraduate probability. *Amer. Math. Monthly* **120** 583–593. MR3096465

ALDOUS, D. (2017). Mathematical probability foundations of dynamic sports ratings: Overview and open problems. To appear.

ALDOUS, D. and HAN, W. (2017). Introducing Nash equilibria via an online casual game that people actually play. *Amer. Math. Monthly* **124** 506–517. MR3654829

BAYER, D. and DIACONIS, P. (1992). Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** 294–313. MR1161056

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925

CATTELAN, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statist. Sci.* **27** 412–433. MR3012434

CATTELAN, M., VARIN, C. and FIRTH, D. (2013). Dynamic Bradley–Terry modelling of sports tournaments. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **62** 135–150. MR3042315

CHETRITE, R., DIEL, R. and LERASLE, M. (2017). The number of potential winners in Bradley–Terry model in random environment. *Ann. Appl. Probab.* **27** 1372–1394. MR3678473

COX, D. R. and SNELL, E. J. (1989). *Analysis of Binary Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **32**. Chapman & Hall, London. MR1014891

CURIEL, R. S. DA S. (2017). World Football Elo Ratings. Available at http://www.eloratings.net.

DAVID, H. A. (1988). *The Method of Paired Comparisons*, 2nd ed. *Griffin's Statistical Monographs & Courses* **41**. Charles Griffin & Co., Ltd., London. MR0947340

GARDINER, C. W. (1983). *Handbook of Stochastic Methods*: *For Physics*, *Chemistry and the Natural Sciences*. *Springer Series in Synergetics* **13**. Springer, Berlin. MR0714148

GLICKMAN, M. E. (2001). Dynamic paired comparison models with stochastic variances. *J. Appl. Stat.* **28** 673–689. MR1862491

HVATTUM, L. M. and ARNTZEN, H. (2010). Using Elo ratings for match result prediction in association football. *Int. J. Forecast.* **26** 460–470.

JABIN, P.-E. and JUNCA, S. (2015). A continuous model for ratings. *SIAM J. Appl. Math.* **75** 420–442. MR3323555

KIRÁLY, F. J. and QIAN, Z. (2017). Modelling competitive sports: Bradley–Terry–Elo models for supervised and online learning of paired competition outcomes. Available at arXiv:1701.08055.

KNORR-HELD, L. (2000). Dynamic rating of sports teams. *Statistician* **49** 261–276.

KOVALCHIK, S. (2016). Searching for the GOAT of tennis win prediction. *J. Quant. Anal. Sports* **12** 127–138.

LANGE, K. (2010). *Applied Probability*, 2nd ed. Springer, New York. MR2680838

LANGVILLE, A. N. and MEYER, C. D. (2012). *Who's #1? The Science of Rating and Ranking*. Princeton Univ. Press, Princeton, NJ. MR2894870

David Aldous is Professor, Department of Statistics, 367 Evans Hall # 3860, University of California, Berkeley, Berkeley, California 94720, USA (e-mail: aldous@stat.berkeley.edu).

LASEK, J., SZLÁVIK, Z. and BHULAI, S. (2013). The predictive power of ranking systems in association football. *Int. J. Appl. Pattern Recogn.* **1** 27–46.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London. MR3223057

MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

RESNICK, S. I. (1987). *Extreme Values*, *Regular Variation*, *and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810

TETLOCK, P. E. (2006). *Expert Political Judgment*: *How Good Is It? How Can We Know?* Princeton Univ. Press, Princeton, NJ.

UNITED, O. (2017). Weekly Tennis ELO Rankings. Available at https://tenniseloranking.blogspot.com.

WIKIPEDIA (2014a). Elo rating system. Wikipedia, the free encyclopedia. [Online; accessed, 31-October-2014].

WIKIPEDIA (2014b). Tournament. Wikipedia, the free encyclopedia. [Online; accessed. 4-December-2014].

WIKIPEDIA (2017). Promotion and relegation. Wikipedia, the free encyclopedia. [Online; accessed 19-February-2017].

Statistical Science
2017, Vol. 32, No. 4, 630–639
DOI: 10.1214/17-STS625

On a General Definition of Depth for Functional Data

Irène Gijbels and Stanislav Nagy

Abstract. In this paper, we provide an elaboration on the desirable properties of statistical depths for functional data. Although a formal definition has been put forward in the literature, there are still several unclarities to be tackled, and further insights to be gained. Herein, a few interesting connections between the wanted properties are found. In particular, it is demonstrated that the conditions needed for some desirable properties to hold are extremely demanding, and virtually impossible to be met for common depths. We establish adaptations of these properties which prove to be still sensible, and more easily met by common functional depths.

Key words and phrases: Data depth, functional data, multivariate statistics, robustness.

REFERENCES

CHAKRABORTY, A. and CHAUDHURI, P. (2014a). On data depth in infinite dimensional spaces. *Ann. Inst. Statist. Math.* **66** 303–324. MR3171407

CHAKRABORTY, A. and CHAUDHURI, P. (2014b). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286

CLAESKENS, G., HUBERT, M., SLAETS, L. and VAKILI, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423. MR3180573

CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496. MR2336349

CUEVAS, A. and FRAIMAN, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivariate Anal.* **100** 753–766. MR2478196

DUTTA, S., GHOSH, A. K. and CHAUDHURI, P. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli* **17** 1420–1434. MR2854779

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. Springer, New York. MR2229687

FLEMING, R. J. and JAMISON, J. E. (2003). *Isometries on Banach Spaces*: *Function Spaces*. *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **129**. Chapman & Hall/CRC, Boca Raton, FL. MR1957004

FRAIMAN, R. and MUNIZ, G. (2001). Trimmed means for functional data. *TEST* **10** 419–440. MR1881149

GIJBELS, I. and NAGY, S. (2017) Supplement to "On a general definition of depth for functional data." DOI:10.1214/17-STS625SUPP.

KUELBS, J. and ZINN, J. (2013). Concerns with functional depth. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 831–855. MR3125749

KUELBS, J. and ZINN, J. (2015). Half-region depth for stochastic processes. *J. Multivariate Anal.* **142** 86–105. MR3412741

LONG, J. P. and HUANG, J. Z. (2016). A study of functional depths. Preprint. Available at arXiv:1506.01332.

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590

LÓPEZ-PINTADO, S. and ROMO, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.* **55** 1679–1695. MR2748671

MOSLER, K. (2013). Depth statistics. In *Robustness and Complex Data Structures* 17–34. Springer, Heidelberg. MR3135871

NAGY, S., GIJBELS, I., OMELKA, M. and HLUBINKA, D. (2016). Integrated depth for functional data: Statistical properties and consistency. *ESAIM Probab. Stat.* **20** 95–130. MR3528619

NIETO-REYES, A. and BATTEY, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.* **31** 61–79. MR3458593

Irène Gijbels is Professor, Department of Mathematics and Leuven Statistics Research Centre (LStat), KU Leuven, Celestijnenlaan 200b—box 2400, 3001 Leuven, Belgium (e-mail: irene.gijbels@kuleuven.be). Stanislav Nagy is Research Assistant, Department of Mathematics and Leuven Statistics Research Centre (LStat), KU Leuven, Celestijnenlaan 200b—box 2400, 3001 Leuven, Belgium, and Research Assistant, Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic (e-mail: stanislav.nagy@kuleuven.be; nagy@karlin.mff.cuni.cz).

PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth: A focus on centrality. *J. Amer. Statist. Assoc.* **108** 1105–1119. MR3174687

RUDIN, W. (1991). *Functional Analysis*, 2nd ed. McGraw-Hill, New York. MR1157815

SERFLING, R. (2006). Multivariate symmetry and asymmetry. In *Encyclopedia of Statistical Sciences*, *Vol.* 8, 2nd ed. 5338–5345. Wiley, New York.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver, B. C.*, 1974), *Vol.* 2 523–531. MR0426989

VÄISÄLÄ, J. (2003). A proof of the Mazur–Ulam theorem. *Amer. Math. Monthly* **110** 633–635. MR2001155

VALENTINE, F. A. (1964). *Convex Sets*. McGraw-Hill, New York. MR0170264

ZUO, Y. and SERFLING, R. (2000a). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005

ZUO, Y. and SERFLING, R. (2000b). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inference* **84** 55–79. MR1747497

Statistical Science
2017, Vol. 32, No. 4, 640
DOI: 10.1214/17-STS626

Correction to "A Topologically Valid Definition of Depth for Functional Data"

Alicia Nieto-Reyes and Heather Battey

Alicia Nieto-Reyes is Associate Professor, Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain (e-mail: alicia.nieto@unican.es). Heather Battey is Lecturer, Imperial College London, Department of Mathematics, South Kensington Campus, London, England SW7 2AZ (e-mail: h.battey@imperial.ac.uk).

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Alison Etheridge, Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom

President-Elect: Xiao-Li Meng, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

Past President: Jon Wellner, Department of Statistics, University of Washington, Seattle, Washington 98195-4322, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisonsin, Madison, Wisconin 53706-1510, USA

Program Secretary: Judith Rousseau, Université Paris Dauphine, Place du Maréchal DeLattre de Tassigny, 75016 Paris, France

IMS EDITORS

The Annals of Statistics. *Editors*: Edward I. George, Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; Tailen Hsing, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109-1107, USA

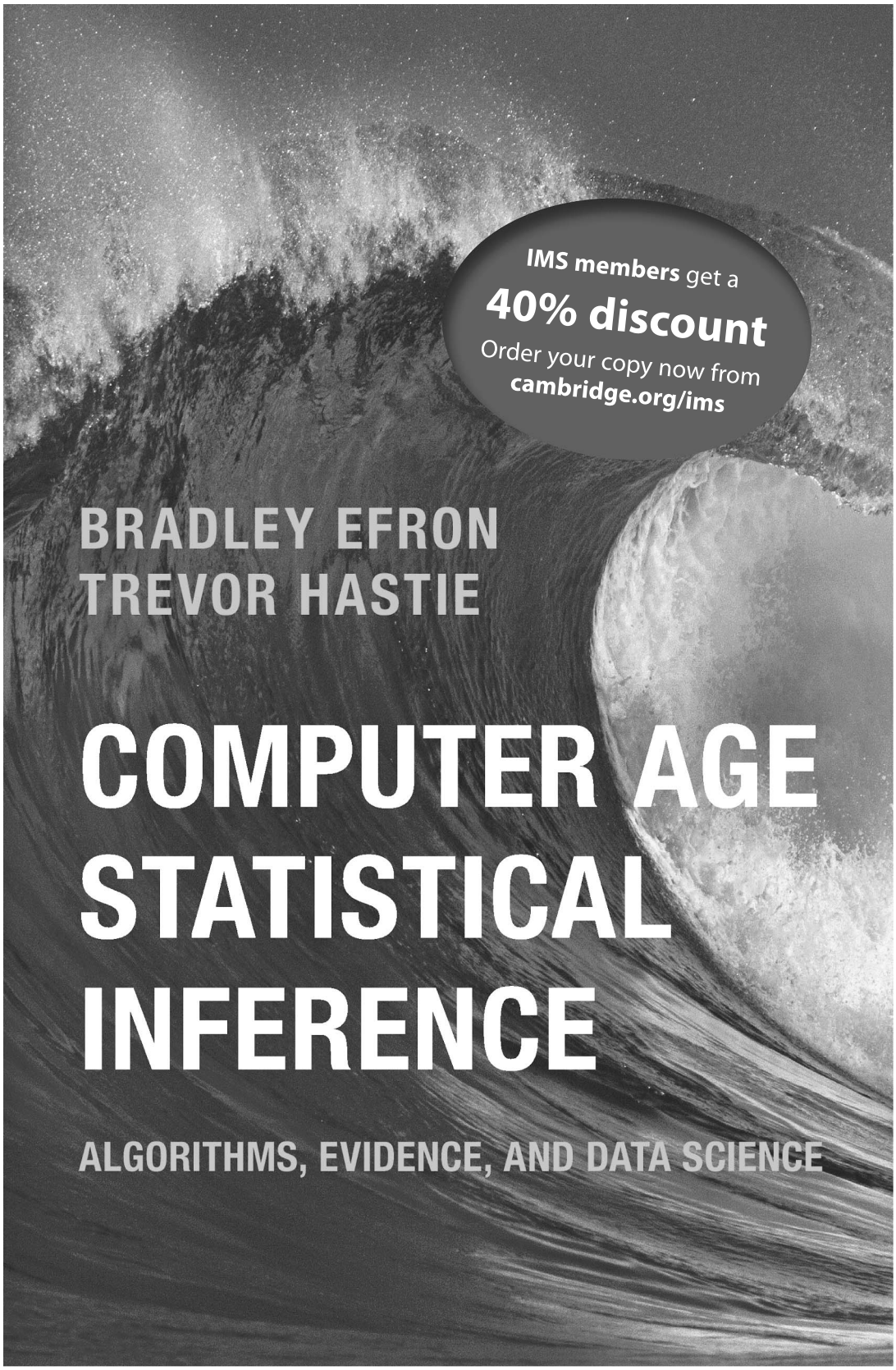
The Annals of Applied Statistics. *Editor-in-Chief*: Tilmann Gneiting, Heidelberg Institute for Theoretical Studies, HITS gGmbH, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

The Annals of Probability. *Editor*: Maria Eulália Vares, Instituto de Matemática, Universidade Federal do Rio de Janeiro, 21941-909 Rio de Janeiro, RJ, Brazil

The Annals of Applied Probability. *Editor*: Bálint Tóth, School of Mathematics, University of Bristol, University Walk, BS8 1TW, Bristol, United Kingdom, and Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

Statistical Science. *Editor*: Cun-Hui Zhang, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854, USA

The IMS Bulletin. *Editor*: Vlada Limic, DR2, CNRS, Université Paris Sud 11, UMR 8628, Département de Mathématiques, 91405 Orsay, France


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE