STATISTICAL SCIENCE

Volume 40, Number 2 May 2025

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 40, Number 2, May 2025. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices. **POSTMASTER:** Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Statistical Science Volume 40, Number 2 (167–345) May 2025

Volume 40 Number 2 May 2025

Statistical Science
2025, Vol. 40, No. 2, 167–194
https://doi.org/10.1214/23-STS896

Item Response Theory—A Statistical Framework for Educational and Psychological Measurement

Yunxiao Chen, Xiaoou Li, Jingchen Liu and Zhiliang Ying

Abstract. Item response theory (IRT) has become one of the most popular statistical models for psychometrics, a field of study concerned with the theory and techniques of psychological measurement. The IRT models are latent factor models tailored to the analysis, interpretation and prediction of individuals' behaviors in answering a set of measurement items that typically involve categorical response data. Many important questions of measurement are directly or indirectly answered through the use of IRT models, including scoring individuals' test performances, validating a test scale, linking two tests, among others. This paper provides a review of item response theory, including its statistical framework and psychometric applications. We establish connections between item response theory and related topics in statistics, including empirical Bayes, nonparametric methods, matrix completion, regularized estimation and sequential analysis. Possible future directions of IRT are discussed from the perspective of statistical learning.

Key words and phrases: Psychometrics, measurement theory, factor analysis, item response theory, latent trait, validity, reliability.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **AC-19** 716–723. System identification and time-series analysis. MR0423716 https://doi.org/10.1109/tac.1974.1100705

ANDERSEN, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc. Ser. B* **32** 283–301. MR0273723

ANDERSEN, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika* **42** 69–81. MR0483255 https://doi.org/10.1007/BF02293746

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. V* 111–150. Univ. California Press, Berkeley–Los Angeles, CA. MR0084943

ANDRICH, D. (2010). Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika* **75** 292–308. MR2719929 https://doi.org/10.1007/s11336-010-9154-8

AMERICAN PSYCHIATRIC ASSOCIATION (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn., (DSM-5®). American Psychiatric Association, Arlington, VA.

BAFUMI, J., GELMAN, A., PARK, D. K. and KAPLAN, N. (2005). Practical issues in implementing and understanding Bayesian ideal point estimation. *Polit. Anal.* **13** 171–187.

BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313 https://doi.org/10.1214/11-AOS966

BALSIS, S., RUCHENSKY, J. R. and BUSCH, A. J. (2017). Item response theory applications in personality disorder research. *Personality Disorders: Theory, Research, and Treatment* **8** 298–308.

BAROCAS, S., HARDT, M. and NARAYANAN, A. (2018). Fairness and machine learning. fairmlbook.org. http://www.fairmlbook.org.

BARTHOLOMEW, D., KNOTT, M. and MOUSTAKI, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2849614 https://doi.org/10.1002/9781119970583

BARTHOLOMEW, D. J. (1980). Factor analysis for categorical data. *J. Roy. Statist. Soc. Ser. B* **42** 293–321. With discussion. MR0596159

BARTHOLOMEW, D. J. (1984). The foundations of factor analysis. *Biometrika* **71** 221–232. MR0767150 https://doi.org/10.1093/biomet/71.2.221

BARTHOLOMEW, D. J. and LEUNG, S. O. (2002). A goodness of fit test for sparse 2^p contingency tables. *Br. J. Math. Stat. Psychol.* **55** 1–15. MR1930152 https://doi.org/10.1348/000711002159617

BARTHOLOMEW, D. J. and TZAMOURANI, P. (1999). The goodness of fit of latent trait models in attitude measurement. *Sociol. Methods*

Yunxiao Chen is Associate Professor, Department of Statistics, London School of Economics and Political Science, London, UK (e-mail: y.chen186@lse.ac.uk). Xiaoou Li is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, USA. Jingchen Liu is Professor, Department of Statistics, Columbia University, New York, New York 10027, USA. Zhiliang Ying is Professor, Department of Statistics, Columbia University, New York, New York 10027, USA.

Res. **27** 525–546.

BARTROFF, J., LAI, T. L. and SHIH, M.-C. (2013). *Sequential Experimentation in Clinical Trials*: *Design and Analysis*. *Springer Series in Statistics*. Springer, New York. MR2987767 https://doi.org/10.1007/978-1-4614-6114-2

BELZAK, W. C. M. and BAUER, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychol*. *Methods* **25** 673–690. https://doi.org/10.1037/met0000253

BERRIDGE, D. M. and CROUCHLEY, R. (2011). *Multivariate Generalized Linear Mixed Models Using R*. CRC Press/CRC, Boca Raton, FL.

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

BHASKAR, S. A. (2016). Probabilistic low-rank matrix completion from quantized measurements. *J*. *Mach*. *Learn*. *Res*. **17** Paper No. 60, 34. MR3504620

BHASKAR, S. A. and JAVANMARD, A. (2015). 1-bit matrix completion under exact low-rank constraint. In 2015 49*th Annual Conference on Information Sciences and Systems* (*CISS*) 1–6. IEEE.

BIRDSALL, M. (2011). *Implementing Computer Adaptive Testing to Improve Achievement Opportunities*. Office of Qualifications and Examinations Regulation, Coventry, UK.

BIRNBAUM, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores* (F. M. Lord and M. R. Novick, eds.) 397–479. Addison-Wesley, Reading, MA.

BOCK, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37** 29–51.

BOCK, R. D. and AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46** 443–459. MR0668311 https://doi.org/10.1007/BF02293801

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR0996025 https://doi.org/10.1002/9781118619179

BRADLOW, E. T., WAINER, H. and WANG, X. (1999). A Bayesian random effects model for testlets. *Psychometrika* **64** 153–168.

BRAEKEN, J. (2011). A boundary mixture approach to violations of conditional independence. *Psychometrika* **76** 57–76. MR2783872 https://doi.org/10.1007/s11336-010-9190-4

BRAEKEN, J., TUERLINCKX, F. and DE BOECK, P. (2007). Copula functions for residual dependency. *Psychometrika* **72** 393–411. MR2361964 https://doi.org/10.1007/s11336-007-9005-4

BROWNE, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivar*. *Behav*. *Res*. **36** 111–150.

CAI, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika* **75** 33–57. MR2609729 https://doi.org/10.1007/s11336-009-9136-x

CAI, L. (2010b). Metropolis–Hastings Robbins–Monro algorithm for confirmatory item factor analysis. *J*. *Educ*. *Behav*. *Stat*. **35** 307–335.

CAI, L., YANG, J. S. and HANSEN, M. (2011). Generalized full-information item bifactor analysis. *Psychol*. *Methods* **16** 221–248.

CAI, T. and ZHOU, W.-X. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *J*. *Mach*. *Learn*. *Res*. **14** 3619–3647. MR3159403

CARLSON, J. E. and VON DAVIER, M. (2017). Item response theory. In *Advancing Human Assessment*: *A Synthesis over Seven Decades* (R. E. Bennett and M. von Davier, eds.) 133–178. Springer, New York, NY.

CATTELL, R. B. (1966). The scree test for the number of factors. *Multivar*. *Behav*. *Res*. **1** 245–276. https://doi.org/10.1207/s15327906mbr0102_10

CELLA, D., CHANG, C.-H. and HEINEMANN, A. W. (2002). Item response theory (IRT): Applications in quality of life measurement, analysis and interpretation. In *Statistical Methods for Quality of Life Studies* (M. Mesbah, B. F. Cole and M.-L. T. Lee, eds.) 169–185. Springer, New York, NY.

CELLA, D., YOUNT, S., ROTHROCK, N., GERSHON, R., COOK, K., REEVE, B., ADER, D., FRIES, J. F., BRUCE, B. et al. (2007). The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Med*. *Care* **45** S3–S11. https://doi.org/10.1097/01.mlr.0000258615.42478.55

CEN, H., KOEDINGER, K. and JUNKER, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International Conference on Intelligent Tutoring Systems* (M. Ikeda, K. D. Ashley and T.-W. Chan, eds.) 164–175. Springer.

CHANG, H.-H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika* **80** 1–20. MR3325843 https://doi.org/10.1007/s11336-014-9401-5

CHANG, H.-H., MAZZEO, J. and ROUSSOS, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *J*. *Educ*. *Meas*. **33** 333–353.

CHANG, H.-H. and YING, Z. (1996). A global information approach to computerized adaptive testing. *Appl*. *Psychol*. *Meas*. **20** 213–229.

CHANG, H.-H. and YING, Z. (1999). A-stratified multistage computerized adaptive testing. *Appl*. *Psychol*. *Meas*. **23** 211–222.

CHANG, H.-H. and YING, Z. (2009). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *Ann*. *Statist*. **37** 1466–1488. MR2509080 https://doi.org/10.1214/08-AOS614

CHANG, Y.-C. I. (2005). Application of sequential interval estimation to adaptive mastery testing. *Psychometrika* **70** 685–713. MR2272511 https://doi.org/10.1007/s11336-005-1140-9

CHEN, W.-H. and THISSEN, D. (1997). Local dependence indexes for item pairs using item response theory. *J*. *Educ*. *Behav*. *Stat*. **22** 265–289.

CHEN, Y. (2020). A continuous-time dynamic choice measurement model for problem-solving process data. *Psychometrika* **85** 1052–1075. MR4204752 https://doi.org/10.1007/s11336-020-09734-1

CHEN, Y., CULPEPPER, S. A., CHEN, Y. and DOUGLAS, J. (2018a). Bayesian estimation of the DINA Q matrix. *Psychometrika* **83** 89–108. MR3767014 https://doi.org/10.1007/s11336-017-9579-4

CHEN, Y., CULPEPPER, S. A., WANG, S. and DOUGLAS, J. (2018b). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Appl*. *Psychol*. *Meas*. **42** 5–23.

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023). Statistical inference for noisy incomplete binary matrix. *J*. *Mach*. *Learn*. *Res*. **24** Paper No. [95], 66. MR4582517

CHEN, Y. and LI, X. (2022). Determining the number of factors in high-dimensional generalized latent factor models. *Biometrika* **109** 769–782. MR4472847 https://doi.org/10.1093/biomet/asab044

CHEN, Y., LI, X., LIU, J. and YING, Z. (2018a). Recommendation system for adaptive learning. *Appl*. *Psychol*. *Meas*. **42** 24–41.

CHEN, Y., LI, X., LIU, J. and YING, Z. (2018b). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika* **83** 538–562. MR3851946 https://doi.org/10.1007/s11336-018-9610-4

CHEN, Y., LI, X. and ZHANG, S. (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika* **84** 124–146. MR3910911 https://doi.org/10.1007/s11336-018-9646-5

CHEN, Y., LI, X. and ZHANG, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *J. Amer. Statist. Assoc.* **115** 1756–1770. MR4189755 https://doi.org/10.1080/01621459.2019.1635485

CHEN, Y., LIU, J., XU, G. and YING, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *J. Amer. Statist. Assoc.* **110** 850–866. MR3367269 https://doi.org/10.1080/01621459.2014.934827

CHEN, Y., MOUSTAKI, I. and ZHANG, H. (2020). A note on likelihood ratio tests for models with latent variables. *Psychometrika* **85** 996–1012. MR4204748 https://doi.org/10.1007/s11336-020-09735-0

CHENG, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika* **74** 619–632. MR2565330 https://doi.org/10.1007/s11336-009-9123-2

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.* **25** 573–578. MR0065087 https://doi.org/10.1214/aoms/1177728725

CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Stat.* **30** 755–770. MR0108874 https://doi.org/10.1214/aoms/1177706205

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

CHIU, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Appl. Psychol. Meas.* **37** 598–618.

CHIU, C.-Y., DOUGLAS, J. A. and LI, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika* **74** 633–665. MR2565331 https://doi.org/10.1007/s11336-009-9125-0

CHRISTOFFERSSON, A. (1975). Factor analysis of dichotomized variables. *Psychometrika* **40** 5–32. MR0408121 https://doi.org/10.1007/BF02291477

CHUNG, K. L. (1954). On a stochastic approximation method. *Ann. Math. Stat.* **25** 463–483. MR0064365 https://doi.org/10.1214/aoms/1177728716

COHEN, A. S. and CHO, S.-J. (2016). Information criteria. In *Handbook of Item Response Theory: Models, Statistical Tools, and Applications* (W. van der Linden, ed.). Chapman & Hall/CRC Press, Boca Raton, FL.

COLLINS, L. M., FIDLER, P. L., WUGALTER, S. E. and LONG, J. D. (1993). Goodness-of-fit testing for latent class models. *Multivar. Behav. Res.* **28** 375–389. https://doi.org/10.1207/s15327906mbr2803_4

COX, D. R. and REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91** 729–737. MR2090633 https://doi.org/10.1093/biomet/91.3.729

CRESSIE, N. and HOLLAND, P. W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika* **48** 129–141. MR0699244 https://doi.org/10.1007/BF02314681

CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* **16** 297–334.

CULPEPPER, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *J. Educ. Behav. Stat.* **40** 454–476.

CULPEPPER, S. A. (2019). Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. *Psychometrika* **84** 333–357. MR3947367 https://doi.org/10.1007/s11336-018-9643-8

DAVENPORT, M. A., PLAN, Y., VAN DEN BERG, E. and WOOTTERS, M. (2014). 1-bit matrix completion. *Inf. Inference* **3** 189–223. MR3311452 https://doi.org/10.1093/imaiai/iau006

DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001

DE JONG, M. G., STEENKAMP, J.-B. E., FOX, J.-P. and BAUMGARTNER, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *J. Mark. Res.* **45** 104–115.

DE LEEUW, J. and VERHELST, N. (1986). Maximum likelihood estimation in generalized Rasch models. *J. Educ. Stat.* **11** 183–196.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. With discussion. MR0501537

DE BOECK, P. and WILSON, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. Springer, New York, NY.

DE LA TORRE, J. (2011). The generalized DINA model framework. *Psychometrika* **76** 179–199. MR2788881 https://doi.org/10.1007/s11336-011-9207-7

DE LA TORRE, J. and CHIU, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika* **81** 253–273. MR3505366 https://doi.org/10.1007/s11336-015-9467-8

DIBELLO, L. V., STOUT, W. F. and ROUSSOS, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In *Cognitively Diagnostic Assessment* (P. D. Nichols, S. Chipman and R. L. Brennan, eds.) 361–389. Routledge, New York, NY.

DIEBOLT, J. and IP, E. H. S. (1996). Stochastic EM: Method and application. In *Markov Chain Monte Carlo in Practice*. *Interdiscip. Statist.* 259–273. CRC Press, London. MR1397971

DOIGNON, J.-P. and FALMAGNE, J.-C. (2012). *Knowledge Spaces*. Springer, New York. MR1672475 https://doi.org/10.1007/978-3-642-58625-5

DOUGLAS, J. (1997). Joint consistency of nonparametric item characteristic curve and ability estimation. *Psychometrika* **62** 7–28. MR1439472 https://doi.org/10.1007/BF02294778

DOUGLAS, J. A. (2001). Asymptotic identifiability of nonparametric item response models. *Psychometrika* **66** 531–540. MR1961912 https://doi.org/10.1007/BF02296194

DOUGLAS, J. A., STOUT, W. and DIBELLO, L. V. (1996). A kernel-smoothed version of SIBTEST with applications to local DIF inference and function estimation. *J. Educ. Behav. Stat.* **21** 333–363.

DRTON, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. MR2502658 https://doi.org/10.1214/07-AOS571

EDWARDS, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika* **75** 474–497. MR2719939 https://doi.org/10.1007/s11336-010-9161-9

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

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

EMBRETSON, S. E. and REISE, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FANG, G., GUO, J., XU, X., YING, Z. and ZHANG, S. (2021). Identifiability of bifactor models. *Statist. Sinica* **31** 2309–2330. MR4338090 https://doi.org/10.5705/ss.20

FANG, G., LIU, J. and YING, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika* **84** 19–40. MR3910906 https://doi.org/10.1007/s11336-018-09658-x

FERGUSON, G. A. (1942). Item selection by the constant process. *Psychometrika* **7** 19–29.

FEUERVERGER, A., HE, Y. and KHATRI, S. (2012). Statistical significance of the Netflix challenge. *Statist. Sci.* **27** 202–231. MR2963993 https://doi.org/10.1214/11-STS368

FIENBERG, S. E. (2000). Contingency tables and log-linear models: Basic results and new developments. *J. Amer. Statist. Assoc.* **95** 643–647.

FIENBERG, S. E. and MEYER, M. M. (1983). Loglinear models and categorical data analysis with psychometric and econometric applications. *J. Econometrics* **22** 191–214. MR0719131 https://doi.org/10.1016/0304-4076(83)90099-4

GHOSH, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statist. Probab. Lett.* **23** 165–170. MR1341360 https://doi.org/10.1016/0167-7152(94)00109-L

GIBBONS, R. D. and HEDEKER, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika* **57** 423–436.

GOLDBERGER, A. S. (1972). Structural equation methods in the social sciences. *Econometrica* **40** 979–1001. MR0327267 https://doi.org/10.2307/1913851

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GOODMAN, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part IA modified latent structure approach. *Am. J. Sociol.* **79** 1179–1259.

GOODMAN, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231.

GU, M. G. and KONG, F. H. (1998). A stochastic approximation algorithm with Markov chain Monte-Carlo method for incomplete data estimation problems. *Proc. Natl. Acad. Sci. USA* **95** 7270–7274.

GU, Y. and XU, G. (2020). Partial identifiability of restricted latent class models. *Ann. Statist.* **48** 2082–2107. MR4134787 https://doi.org/10.1214/19-AOS1878

GUO, H. and SINHARAY, S. (2011). Nonparametric item response curve estimation with correction for measurement error. *J. Educ. Behav. Stat.* **36** 755–778.

HABERMAN, S. J. (1977a). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841. MR0501540

HABERMAN, S. J. (1977b). Product models for frequency tables involving indirect observation. *Ann. Statist.* **5** 1124–1147. MR0445687

HABERMAN, S. J. (2005). Identifiability of parameters in item response models with unconstrained ability distributions (ETS research report series RR-05-24). Technical report, Educational Testing Service, Princeton, NJ.

HABERMAN, S. J. and SINHARAY, S. (2013). Generalized residuals for general models for contingency tables with application to item response theory. *J. Amer. Statist. Assoc.* **108** 1435–1444.

HABERMAN, S. J., SINHARAY, S. and CHON, K. H. (2013). Assessing item fit for unidimensional item response theory models using residuals from estimated item response functions. *Psychometrika* **78** 417–440. https://doi.org/10.1007/s11336-012-9305-1

HAERTEL, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *J. Educ. Meas.* **26** 301–321.

HAMBLETON, R. K., SWAMINATHAN, H. and ROGERS, H. J. (1991). *Fundamentals of Item Response Theory*. Sage, Newbury Park, CA.

HAYS, R. D., MORALES, L. S. and REISE, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Med. Care* **38** II28–II42. https://doi.org/10.1097/00005650-200009002-00007

HENSON, R. A., TEMPLIN, J. L. and WILLSE, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika* **74** 191–210.

HOLLAND, P. W. (1990a). The Dutch identity: A new tool for the study of item response models. *Psychometrika* **55** 5–18. MR1060262 https://doi.org/10.1007/BF02294739

HOLLAND, P. W. (1990b). On the sampling theory foundations of item response theory models. *Psychometrika* **55** 577–601. MR1094332 https://doi.org/10.1007/BF02294609

HOLLAND, P. W. and HOSKENS, M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika* **68** 123–149. MR2272374 https://doi.org/10.1007/BF02296657

HOLLAND, P. W. and THAYER, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In *Test Validity* (H. Wainer and H. I. Braun, eds.) 129–145. Lawrence Erlbaum Associates, Hillsdale, NJ.

HOLLAND, P. W. and WAINER, H. (1993). *Differential Item Functioning*. Lawrence Erlbaum Associates, Hillsdale, NJ.

HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30** 179–185. https://doi.org/10.1007/BF02289447

HOSKENS, M. and DE BOECK, P. (1997). A parametric model for local dependence among test items. *Psychol. Methods* **2** 261–275.

HUANG, P.-H. (2018). A penalized likelihood method for multi-group structural equation modelling. *Br. J. Math. Stat. Psychol.* **71** 499–522. https://doi.org/10.1111/bmsp.12130

HUBER, P., RONCHETTI, E. and VICTORIA-FESER, M.-P. (2004). Estimation of generalized linear latent variable models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 893–908.

INKSTER, B., STILLWELL, D., KOSINSKI, M. and JONES, P. (2016). A decade into Facebook: Where is psychiatry in the digital age? *Lancet Psychiatry* **3** 1087–1090.

IP, E. H. (2002a). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika* **67** 367–386. MR1961323 https://doi.org/10.1007/BF02294990

IP, E. H. (2002b). On single versus multiple imputation for a class of stochastic algorithms estimating maximum likelihood. *Comput. Statist.* **17** 517–524. MR1952695 https://doi.org/10.1007/s001800200124

IP, E. H., WANG, Y. J., DE BOECK, P. and MEULDERS, M. (2004). Locally dependent latent trait model for polytomous responses with application to inventory of hostility. *Psychometrika* **69** 191–216.

ISING, E. (1925). Beitrag zur theorie des ferromagnetismus. *Z. Phys. Hadrons Nucl.* **31** 253–258.

JENNRICH, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika* **69** 257–273.

JENNRICH, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika* **71** 173–191.

JENNRICH, R. I. and BENTLER, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika* **76** 537–549.

JENNRICH, R. I. and SAMPSON, P. F. (1966). Rotation for simple loadings. *Psychometrika* **31** 313–323. MR0203869 https://doi.org/10.1007/BF02289465

JIANG, H. and STOUT, W. (1998). Improved type I error control and reduced estimation bias for DIF detection using SIBTEST. *J. Educ. Behav. Stat.* **23** 291–322.

JIANG, Z. and TEMPLIN, J. (2019). Gibbs samplers for logistic item response models via the Pólya-gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika* **84** 358–374. https://doi.org/10.1007/s11336-018-9641-x

JOHNSON, M. S. (2006). Nonparametric estimation of item and respondent locations from unfolding-type items. *Psychometrika* **71** 257–279. MR2259175 https://doi.org/10.1007/s11336-003-1098-9

JOHNSON, M. S. (2007). Modeling dichotomous item responses with free-knot splines. *Comput. Statist. Data Anal.* **51** 4178–4192. MR2364438 https://doi.org/10.1016/j.csda.2006.04.021

JÖRESKOG, K. G. and MOUSTAKI, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivar. Behav. Res.* **36** 347–387. https://doi.org/10.1207/S15327906347-387

JUNKER, B. W. and SIJTSMA, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Appl. Psychol. Meas.* **25** 258–272.

KAISER, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23** 187–200.

KAISER, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20** 141–151.

KAPCZINSKI, F., MWANGI, B. and PASSOS, I. C., eds. (2019). *Personalized Psychiatry: Big Data Analytics in Mental Health*. Springer, New York, NY.

KATSIKATSOU, M., MOUSTAKI, I., YANG-WALLENTIN, F. and JÖRESKOG, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Comput. Statist. Data Anal.* **56** 4243–4258. MR2957868 https://doi.org/10.1016/j.csda.2012.04.010

KHAJAH, M., WING, R., LINDSEY, R. and MOZER, M. (2014a). Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Proceedings of the 7th International Conference on Educational Data Mining* (J. Stamper, Z. Pardos, M. Mavrikis and B. McLaren, eds.) 99–106, London, United Kingdom.

KHAJAH, M. M., HUANG, Y., GONZÁLEZ-BRENES, J. P., MOZER, M. C. and BRUSILOVSKY, P. (2014b). Integrating knowledge tracing and item response theory: A tale of two frameworks. In *Proceedings of the 4th International Workshop on Personalization Approaches in Learning Environments (PALE 2014) at the 22th International Conference on User Modeling, Adaptation, and Personalization* (M. Kravcik, O. Santos and J. Boticario, eds.) 7–12, Aalborg, Denmark.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 887–906. MR0086464 https://doi.org/10.1214/aoms/1177728066

KIM, S. (2012). A note on the reliability coefficients for item response model-based ability estimates. *Psychometrika* **77** 153–162.

KINGSTON, N. and BROADDUS, A. (2017). The use of learning map systems to support the formative assessment in mathematics. *Educ. Sci.* **7** 41.

KNIGHT, J. (2013). *High-Impact Instruction: A Framework for Great Teaching*. Corwin Press, Thousand Oaks, CA.

KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA.

KOREN, Y. and BELL, R. (2015). Advances in collaborative filtering. In *Recommender Systems Handbook* 77–118.

KOTOV, R., KRUEGER, R. F., WATSON, D., ACHENBACH, T. M., ALTHOFF, R. R., BAGBY, R. M., BROWN, T. A., CARPENTER, W. T., CASPI, A. et al. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126** 454–477.

LAI, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statist. Sinica* **11** 303–351.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

LAWLEY, D. N. (1943). On problems connected with item selection and test construction. *Proc. Roy. Soc. Edinburgh Sect. A* **61** 273–287. MR0007966

LAWLEY, D. N. (1944). The factorial analysis of multiple item tests. *Proc. Roy. Soc. Edinburgh Sect. A* **62** 74–82. MR0010946

LAZARSFELD, P. and HENRY, N. (1968). *Latent Structure Analysis*. Houghton Mifflin, New York, NY.

LAZARSFELD, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In *Studies in Social Psychology in World War II, Vol. 4: Measurement and Prediction* (S. A. Stouffer, ed.) 362–412. Princeton Univ. Press, Princeton, NJ.

LAZARSFELD, P. F. (1959). Latent structure analysis. *Psychology: A Study of a Science* **3** 476–543.

LEE, K.-F. (2018). *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt, New York, NY.

LEE, S.-Y., POON, W.-Y. and BENTLER, P. M. (1990). A three-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika* **55** 45–51.

LEE, S.-Y., POON, W.-Y. and BENTLER, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika* **57** 89–105.

LEE, S. Y., POON, W. Y. and BENTLER, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *Br. J. Math. Stat. Psychol.* **48 (Pt 2)** 339–358. https://doi.org/10.1111/j.2044-8317.1995.tb01067.x

LI, H.-H. and STOUT, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika* **61** 647–677.

LINDSAY, B., CLOGG, C. C. and GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86** 96–107.

LINDSAY, B. G. (1988). Composite likelihood methods. *Sovrem. Mat.* **80** 221–239.

LIU, J., XU, G. and YING, Z. (2012). Data-driven learning of Q-matrix. *Appl. Psychol. Meas.* **36** 548–564. https://doi.org/10.1177/0146621612456591

LIU, J., XU, G. and YING, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli* **19** 1790–1817.

LIU, J., YING, Z. and ZHANG, S. (2015). A rate function approach to computerized adaptive testing for cognitive diagnosis. *Psychometrika* **80** 468–490. MR3353968 https://doi.org/10.1007/s11336-013-9395-4

LIU, Y. and MAYDEU-OLIVARES, A. (2013). Local dependence diagnostics in IRT modeling of binary data. *Educ. Psychol. Meas.* **73** 254–274.

LORD, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educ. Psychol. Meas.* **28** 989–1020.

LORD, F. M. (1971). Robbins-Monro procedures for tailored testing. *Educ. Psychol. Meas.* **31** 3–31.

LORD, F. M. (1977). A broad-range tailored test of verbal ability. *Appl. Psychol. Meas.* **1** 95–100.

LORD, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Routledge, New York, NY.

LORD, F. M. and NOVICK, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.

MACREADY, G. B. and DAYTON, C. M. (1977). The use of probabilistic models in the assessment of mastery. *J. Educ. Stat.* **2** 99–120.

MAGIS, D., TUERLINCKX, F. and DE BOECK, P. (2015). Detection of differential item functioning using the lasso approach. *J. Educ. Behav. Stat.* **40** 111–135.

MANTEL, N. and HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22** 719–748.

MASTERS, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47** 149–174.

MAYDEU-OLIVARES, A. and JOE, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *J. Amer. Statist. Assoc.* **100** 1009–1020. MR2201027 https://doi.org/10.1198/016214504000002069

MAYDEU-OLIVARES, A. and JOE, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika* **71** 713–732.

MCCULLOCH, C. E. and SEARLE, S. R. (2001). *Generalized*, *Linear*, *and Mixed Models*. *Wiley Series in Probability and Statistics*: *Texts*, *References*, *and Pocketbooks Section*. Wiley Interscience, New York. MR1884506

MENG, X.-L. and SCHILLING, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *J*. *Amer. Statist*. *Assoc*. **91** 1254–1267.

MISLEVY, R. J. (1984). Estimating latent distributions. *Psychometrika* **49** 359–381.

MISLEVY, R. J. (1985). Estimation of latent group effects. *J*. *Amer. Statist*. *Assoc*. **80** 993–997.

MISLEVY, R. J. and STOCKING, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Appl*. *Psychol*. *Meas*. **13** 57–75.

MISLEVY, R. J. and WU, P.-K. (1996). Missing responses and irt ability estimation: Omits, choice, time limits, and adaptive testing. ETS Research Report Series, 1996:i–36.

MOKKAN, R. J. and LEWIS, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Appl*. *Psychol*. *Meas*. **6** 417–430.

MOKKEN, R. (1971). *A Theory and Procedure of Scale Analysis*. Mouton, The Hague, The Netherlands.

MOUSTAKI, I. and KNOTT, M. (2000). Generalized latent trait models. *Psychometrika* **65** 391–411.

MURAKI, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Appl*. *Psychol*. *Meas*. **16** 159–176.

MUTHÉN, B. (1978a). Contributions to factor analysis of dichotomous variables. *Psychometrika* **43** 551–560. MR0521904 https://doi.org/10.1007/BF02293813

MUTHÉN, B. (1978b). Contributions to factor analysis of dichotomous variables. *Psychometrika* **43** 551–560.

MUTHÉN, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49** 115–132.

MUTHÉN, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *J*. *Educ*. *Stat*. **10** 121–132.

MUTHÉN, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In *Test Validity* (W. Howard and H. I. Braun, eds.) 213–238. Routledge, New York, NY.

NANDAKUMAR, R. and ROUSSOS, L. (2004). Evaluation of the CAT-SIB DIF procedure in a pretest setting. *J*. *Educ*. *Behav. Stat*. **29** 177–199.

NANDAKUMAR, R. and STOUT, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *J*. *Educ*. *Stat*. **18** 41–68.

NAVE, G., MINXHA, J., GREENBERG, D. M., KOSINSKI, M., STILLWELL, D. and RENTFROW, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychol*. *Sci*. **29** 1145–1158. https://doi.org/10.1177/0956797618761659

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32. MR0025113 https://doi.org/10.2307/1914288

NOVICK, M. R. (1966). The axioms and principal results of classical test theory. *J*. *Math*. *Psych*. **3** 1–18. MR0192593 https://doi.org/10.1016/0022-2496(66)90002-2

OECD (2018). *The Future of Education and Skills*: *Education* 2030. Organisation for Economic Co-operation and Development (OECD), Paris, France.

PARIKH, N. and BOYD, S. (2014). Proximal algorithms. *Found*. *Trends Optim*. **1** 127–239.

PEARSON, K. (1900a). I. Mathematical contributions to the theory of evolution.—VII. On the correlation of characters not quantitatively measurable. *Philos*. *Trans*. *R*. *Soc*. *Lond*. *Ser. A*, *Math*. *Phys*. *Sci*. **195** 1–47.

PEARSON, K. (1900b). 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*. **50** 157–175.

POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J*. *Control Optim*. **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

PUTERMAN, M. L. (2014). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Hoboken, NJ. MR1270015

RAMSAY, J. O. and ABRAHAMOWICZ, M. (1989). Binomial regression with monotone splines: A psychometric application. *J*. *Amer. Statist*. *Assoc*. **84** 906–915.

RAMSAY, J. O. and WINSBERG, S. (1991). Maximum marginal likelihood estimation for semiparametric item analysis. *Psychometrika* **56** 365–379. MR1131766 https://doi.org/10.1007/BF02294480

RASCH, G. (1960). *Probabilistic Models for Some Intelligence and Achievement Tests*. Nielsen and Lydiche, Copenhagen, Denmark.

RECKASE, M. (2009). *Multidimensional Item Response Theory*. Springer, New York, NY.

REISE, S. P. and WALLER, N. G. (2009). Item response theory and clinical measurement. *Annu*. *Rev*. *Clin*. *Psychol*. **5** 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

REISER, M. (1996). Analysis of residuals for the multionmial item response model. *Psychometrika* **61** 509–528.

RICHARDSON, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika* **1** 33–49.

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

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

ROBIN, F., STEFFEN, M. and LIANG, L. (2014). The multistage test implementation of the GRE revised general test. In *Computerized Multistage Testing*: *Theory and Applications* (D. Yan, A. A. von Davier and C. Lewis, eds.) 325–341. CRC Press/CRC, Boca Raton, FL.

ROBINSON, P. M. (1974). Identification, estimation and large-sample theory for regressions containing unobservable variables. *Internat*. *Econom*. *Rev*. **15** 680–692. MR0356376 https://doi.org/10.2307/2525734

ROUSSOS, L. A. and STOUT, W. (2004). Differential item functioning analysis. In *The Sage Handbook of Quantitative Methodology for the Social Sciences* (D. Kaplan, ed.) 107–116. Sage, Thousand Oaks, CA.

RUPP, A. A., TEMPLIN, J. and HENSON, R. A. (2010). *Diagnostic Measurement*: *Theory*, *Methods*, *and Applications*. Guilford Press, New York, NY.

RUPPERT, D. (1988). Efficient estimators from a slowly convergent Robbins–Monro procedure. School of Oper. Res. and Ind. Eng., Cornell Univ., Ithaca, NY, Tech. Rep No. 781.

RUTKOWSKI, L., VON DAVIER, M. and RUTKOWSKI, D. (2013). *Handbook of International Large-Scale Assessment*: *Background*, *Technical Issues*, *and Methods of Data Analysis*. CRC Press, Boca Raton, FL.

SAMEJIMA, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement* **17** 1–68.

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

SHEALY, R. and STOUT, W. (1993a). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika* **58** 159–194.

SHEALY, R. T. and STOUT, W. F. (1993b). An item response theory model for test bias and differential test functioning. In *Differential Test Functioning* (P. Holland and H. Wainer, eds.) 197–239. Lawrence Erlbaum Associates, Hillsdale, NJ.

SHMUELI, G. (2010). To explain or to predict? *Statist. Sci*. **25** 289–310. MR2791669 https://doi.org/10.1214/10-STS330

SIJTSMA, K. and MOLENAAR, I. W. (2002). *Introduction to Nonparametric Item Response Theory*. Sage, Thousand Oaks, CA.

SIJTSMA, K. and VAN DER ARK, L. A. (2017). A tutorial on how to do a mokken scale analysis on your test and questionnaire data. *Br. J. Math. Stat. Psychol*. **70** 137–158.

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling*: *Multilevel*, *Longitudinal*, *and Structural Equation Models*. *Interdisciplinary Statistics*. CRC Press/CRC, Boca Raton, FL. MR2059021 https://doi.org/10.1201/9780203489437

SPEARMAN, C. (1904a). "General intelligence," objectively determined and measured. *Amer. J. Psychol*. **15** 201–292.

SPEARMAN, C. (1904b). The proof and measurement of association between two things. *Amer. J. Psychol*. **15** 72–101.

STOUT, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika* **52** 589–617. MR0936600 https://doi.org/10.1007/BF02294821

STOUT, W., LI, H.-H., NANDAKUMAR, R. and BOLT, D. (1997). MULTISIB: A procedure to investigate DIF when a test is intentionally two-dimensional. *Appl. Psychol. Meas*. **21** 195–213.

STOUT, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika* **55** 293–325. MR1075575 https://doi.org/10.1007/BF02295289

SUN, J., CHEN, Y., LIU, J., YING, Z. and XIN, T. (2016). Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika* **81** 921–939. MR3583913 https://doi.org/10.1007/s11336-016-9529-6

TANG, X., CHEN, Y., LI, X., LIU, J. and YING, Z. (2019). A reinforcement learning approach to personalized learning recommendation systems. *Br. J. Math. Stat. Psychol*. **72** 108–135.

TANG, X., WANG, Z., HE, Q., LIU, J. and YING, Z. (2020). Latent feature extraction for process data via multidimensional scaling. *Psychometrika* **85** 378–397. MR4128067 https://doi.org/10.1007/s11336-020-09708-3

TANG, X., WANG, Z., LIU, J. and YING, Z. (2021). An exploratory analysis of the latent structure of process data via action sequence autoencoders. *Br. J. Math. Stat. Psychol*. **74** 1–33. https://doi.org/10.1111/bmsp.12203

TATSUOKA, C. and FERGUSON, T. (2003). Sequential classification on partially ordered sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **65** 143–157. MR1959819 https://doi.org/10.1111/1467-9868.00377

TATSUOKA, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *J. Educ. Meas*. **20** 345–354.

TEMPLIN, J. L. and HENSON, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychol. Methods* **11** 287–305. https://doi.org/10.1037/1082-989X.11.3.287

THISSEN, D. and STEINBERG, L. (2009). Item response theory. In *The Sage Handbook of Quantitative Methods in Psychology* (Milsap, ed.) 148–177.

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

TJUR, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scand. J. Stat*. **9** 23–30. MR0651857

TUTZ, G. and SCHAUBERGER, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika* **80** 21–43. MR3325844 https://doi.org/10.1007/s11336-013-9377-6

VAN DER LINDEN, W. J. (2018). *Handbook of Item Response Theory*. Chapman and Hall/CRC, New York, NY.

VAN DER LINDEN, W. J. and GLAS, C. A. W., eds. (2010). *Elements of Adaptive Testing*. *Statistics for Social and Behavioral Sciences*. Springer, New York. MR2675773 https://doi.org/10.1007/978-0-387-85461-8

VASDEKIS, V. G. S., CAGNONE, S. and MOUSTAKI, I. (2012). A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika* **77** 425–441. MR2943106 https://doi.org/10.1007/s11336-012-9264-6

VASDEKIS, V. G. S., RIZOPOULOS, D. and MOUSTAKI, I. (2014). Weighted pairwise likelihood estimation for a general class of random effects models. *Biostatistics* **15** 677–689. https://doi.org/10.1093/biostatistics/kxu018

VON DAVIER, M. (2008). A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol*. **61** 287–307. MR2649038 https://doi.org/10.1348/000711007X193957

VON DAVIER, M. and LEE, Y.-S., eds. (2019). *Handbook of Diagnostic Classification Models*. Springer, New York, NY.

WAINER, H., DORANS, N. J., FLAUGHER, R., GREEN, B. F. and MISLEVY, R. J. (2000). *Computerized Adaptive Testing*: *A Primer*. Routledge, New York, NY.

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

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat*. **16** 117–186. MR0013275 https://doi.org/10.1214/aoms/1177731118

WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Stat*. **19** 326–339. MR0026779 https://doi.org/10.1214/aoms/1177730197

WANG, S., YANG, Y., CULPEPPER, S. A. and DOUGLAS, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *J. Educ. Behav. Stat*. **43** 57–87.

WANG, Y. and KOSINSKI, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J. Pers. Soc. Psychol*. **114** 246–257. https://doi.org/10.1037/pspa0000098

WANG, Z., TANG, X., LIU, J. and YING, Z. (2023). Subtask analysis of process data through a predictive model. *Br. J. Math. Stat. Psychol*. **76** 211–235.

WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

WEISS, D. J. (1974). Strategies of adaptive ability measurement (research report 74–5). Technical report, Univ. Minnesota, Dept. Psychology, Psychometric Methods Program, Minneapolis, MN.

WEISS, D. J., ed. (1978). *Proceedings of the* 1977 *Computerized Adaptive Testing Conference*. Univ. Minnesota, Dept. Psychology, Psychometric Methods Program, Minneapolis, MN.

WEISS, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Appl. Psychol. Meas*. **6** 473–492.

WEISS, D. J. and KINGSBURY, G. G. (1984). Application of computerized adaptive testing to educational problems. *J. Educ. Meas*. **21** 361–375.

WILSON, K. and NICHOLS, Z. (2015). *The Knewton Platform*: *A General-Purpose Adaptive Learning Infrastructure*. Knewton, New York, NY. White paper.

WINSBERG, S., THISSEN, D. and WAINER, H. (1984). Fitting item characteristic curves with spline functions. *ETS Res*. *Rep*. *Ser.* **84** 1–14.

WIRTH, R. J. and EDWARDS, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychol*. *Methods* **12** 58–79. https://doi.org/10.1037/1082-989X.12.1.58

WOOD, R. L., WINGERSKY, M. S. and LORD, F. M. (1978). LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Educational Testing Service.

WORLD HEALTH ORGANIZATION (1992). *The ICD*-10 *Classification of Mental and Behavioural Disorders*. World Health Organization, Geneva, Switzerland.

XU, G. (2017). Identifiability of restricted latent class models with binary responses. *Ann*. *Statist*. **45** 675–707. MR3650397 https://doi.org/10.1214/16-AOS1464

XU, G. and SHANG, Z. (2018). Identifying latent structures in restricted latent class models. *J*. *Amer*. *Statist*. *Assoc*. **113** 1284–1295. MR3862357 https://doi.org/10.1080/01621459.2017.1340889

XU, G. and ZHANG, S. (2016). Identifiability of diagnostic classification models. *Psychometrika* **81** 625–649. MR3535051 https://doi.org/10.1007/s11336-015-9471-z

XU, X., CHANG, H. and DOUGLAS, J. (2003). A simulation study to compare cat strategies for cognitive diagnosis. In *Annual Meeting of the American Educational Research Association*, Chicago, IL.

YAN, D., VON DAVIER, A. A. and LEWIS, C. (2014). *Computerized Multistage Testing*: *Theory and Applications*. CRC Press, Boca Raton, FL.

YEN, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Appl*. *Psychol*. *Meas*. **8** 125–145.

ZELLNER, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Internat*. *Econom*. *Rev.* **11** 441–454.

ZHANG, C.-H. (2003). Compound decision theory and empirical Bayes methods. *Ann*. *Statist*. **31** 379–390. MR1983534 https://doi.org/10.1214/aos/1051027872

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann*. *Statist*. **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZHANG, J. and STOUT, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika* **64** 213–249. MR1700709 https://doi.org/10.1007/BF02294536

ZHANG, S. and CHANG, H.-H. (2016). From smart testing to smart learning: How testing technology can assist the new generation of education. *International Journal of Smart Technology and Learning* **1** 67–92.

ZHANG, S. and CHEN, Y. (2022). Computation for latent variable model estimation: A unified stochastic proximal framework. *Psychometrika* **87** 1473–1502. MR4504999 https://doi.org/10.1007/s11336-022-09863-9

ZHANG, S., CHEN, Y. and LIU, Y. (2020). An improved stochastic EM algorithm for large-scale full-information item factor analysis. *Br. J. Math*. *Stat*. *Psychol*. **73** 44–71.

ZHANG, S., WANG, Z., QI, J., LIU, J. and YING, Z. (2023). Accurate assessment via process data. *Psychometrika* **88** 76–97. MR4554873 https://doi.org/10.1007/s11336-022-09880-8

ZHU, Y., SHEN, X. and YE, C. (2016). Personalized prediction and sparsity pursuit in latent factor models. *J*. *Amer*. *Statist*. *Assoc*. **111** 241–252. MR3494656 https://doi.org/10.1080/01621459.2014.999158

Statistical Science
2025, Vol. 40, No. 2, 195–197
https://doi.org/10.1214/24-STS921

Comment: Item Response Theory—Some Minutiae

David Thissen

REFERENCES

[1] ALBERT, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *J. Educ. Stat.* **17** 251–269. https://doi.org/10.2307/1165149

[2] BAUER, D. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychol. Methods* **22** 507–526. https://doi.org/10.1037/met0000077

[3] BELZAK, W. C. M. and BAUER, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychol. Methods* **25** 673–690. https://doi.org/10.1037/met0000253

[4] BIRNBAUM, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores* (F. M. Lord and M. R. Novick, eds.) 395–479. Addison-Wesley, Reading, MA.

[5] BOCK, R. D. (1997). A brief history of item response theory. *Educ. Meas., Issues Pract.* **16** 21–33. https://doi.org/10.1111/j.1745-3992.1997.tb00605.x

[6] BOCK, R. D. and GIBBONS, R. D. (2021). *Item Response Theory*. Wiley, Hoboken, NJ. https://doi.org/10.1002/9781119716723

[7] BOCK, R. D., ZIMOWSKI, M. F. and THISSEN, D. (1997). IRT estimation of domain scores. *J. Educ. Meas.* **34** 197–211. https://doi.org/10.1111/j.1745-3984.1997.tb00515.x

[8] CARLSON, J. and VON DAVIER, M. (2017). Item response theory. In *Advancing Human Assessment: The Methodological, Psychological, and Policy Contributions of ETS* (R. Bennett and M. Von Davier, eds.) 133–178. Springer, Cham. https://doi.org/10.1007/978-3-319-58689-2_5

[9] CULPEPPER, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *J. Educ. Behav. Stat.* **40** 454–476. https://doi.org/10.3102/1076998615595403

[10] EDWARDS, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika* **75** 474–497. MR2719939 https://doi.org/10.1007/s11336-010-9161-9

[11] EMBRETSON, S. (1984). A general latent trait model for response processes. *Psychometrika* **49** 175–186. https://doi.org/10.1007/BF02294171

[12] EMBRETSON, S. E. and REISE, S. P. (2000). *Item Response Theory for Psychologists*. Erlbaum, Hillsdale, NJ.

[13] FOX, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications. Statistics for Social and Behavioral Sciences*. Springer, New York. MR2657265 https://doi.org/10.1007/978-1-4419-0742-4

[14] HANSEN, M., CAI, L., MONROE, S. and LI, Z. (2016). Limited-information goodness-of-fit testing of diagnostic classification item response models. *Br. J. Math. Stat. Psychol.* **69** 225–252. https://doi.org/10.1111/bmsp.12074

[15] HOLLAND, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika* **55** 577–601. MR1094332 https://doi.org/10.1007/BF02294609

[16] HOLLAND, P. W. and WAINER, H., eds. (1993). *Differential Item Functioning*. Erlbaum, Hillsdale, NJ.

[17] JIANG, Z. and TEMPLIN, J. (2019). Gibbs samplers for logistic item response models via the Pólya-gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika* **84** 358–374. MR3947368 https://doi.org/10.1007/s11336-018-9641-x

[18] LAWLEY, D. N. (1943). On problems connected with item selection and test construction. *Proc. Roy. Soc. Edinburgh Sect. A* **61** 273–287. MR0007966

[19] LEIGHTON, J. and GIERL, M. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9780511611186

[20] PATZ, R. and JUNKER, B. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *J. Educ. Behav. Stat.* **24** 146–178. https://doi.org/10.3102/10769986024002146

[21] PATZ, R. and JUNKER, B. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *J. Educ. Behav. Stat.* **24** 342–366. https://doi.org/10.3102/10769986024004342

[22] PETKEWICH, M. O. (2016). The impact of the choice of the item response theory model used in the analysis of student response data from statewide educational assessments. Ph.D. thesis, University of South Carolina.

[23] REISE, S. P. and WALLER, N. G. (2009). Item response theory and clinical measurement. *Annu. Rev. Clin. Psychol.* **5** 27–48. PMID: 18976138. https://doi.org/10.1146/annurev.clinpsy.032408.153553

[24] ROUSSOS, L. A. and STOUT, W. (2004). Differential item functioning analysis. In *The Sage Handbook of Quantitative Methodology for the Social Sciences* (D. Kaplan, ed.) 107–116, Sage, Thousand Oaks, CA.

[25] RUPP, A. A., TEMPLIN, J. and HENSON, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Guilford, New York, NY. https://doi.org/10.1080/15366367.2018.1434349

[26] STEINBERG, L. and THISSEN, D. (1995). Item response theory in personality research. In *Personality Research, Methods & Theory: A Festschrift Honoring Donald W. Fiske* (P. Shrout and S. Fiske, eds.) 161–181. Erlbaum, Hillsdale.

[27] STEINBERG, L. and THISSEN, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychol. Methods* **1** 81–97. https://doi.org/10.1037/1082-989X.1.1.81

[28] TATSUOKA, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *J. Educ. Meas.* **20** 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

David Thissen is Professor Emeritus, Department of Psychology and Neuroscience, The University of North Carolina, Chapel Hill, North Carolina 27599, USA (e-mail: dthissen@email.unc.edu).

[29] TATSUOKA, K. K. (1987). Toward an integration of item-response theory and cognitive error diagnosis Technical Report No. AD-A183 189, Univ. Illinois at Urbana Computer-Based Education Research Lab.

[30] TATSUOKA, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In *Diagnostic Monitoring of Skill and Knowledge Acquisition* (N. Frederiksen, R. Glaser, A. Lesgold and M. G. Shafto, eds.) 453–488. Erlbaum, Hillsdale.

[31] THISSEN, D. and ORLANDO, M. (2001). Item response theory for items scored in two categories. In *Test Scoring* (D. Thissen and H. Wainer, eds.) 73–140. Erlbaum, Mahwah, NJ. https://doi.org/10.4324/9781410604729-8

[32] THISSEN, D. and STEINBERG, L. (2009). Item response theory. In *The Sage Handbook of Quantitative Methods in Psychology* (R. Millsap and A. Maydeu-Olivares, eds.) 148–177. Sage, Thousand Oaks.

[33] VAN DER LINDEN, W. J. (2018). *Handbook of Item Response Theory*, *Three Volume Set*. *Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences*. CRC Press https://doi.org/10.1201/9781315119144.

[34] WIRTH, R. J. and EDWARDS, M. (2007). Item factor analysis: Current approaches and future directions. *Psychol*. *Methods* **12** 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Statistical Science
2025, Vol. 40, No. 2, 198–200
https://doi.org/10.1214/24-STS946

Comment: Wondering and Mind-Wandering Thoughts on IRT

Paul De Boeck

Abstract. I would like to congratulate Chen, Li, Liu and Ying for their excellent review paper. Based on their inspiring review, my comments focus on fixed vs. random effects for persons and items, the reinterpretation of latent variables as random effects instead of causal entities, the reevaluation of local dependence, and potential biases of DIF detection methods. In conclusion, like Chen et al., I see a broader potential for IRT than measurement.

Key words and phrases: Random and fixed effects, latent variables, local dependence, DIF.

REFERENCES

BECHGER, T. M. and MARIS, G. (2015). A statistical test for differential item pair functioning. *Psychometrika* **80** 317–340. MR3353960 https://doi.org/10.1007/s11336-014-9408-y

BOLSINOVA, M., DE BOECK, P. and TIJMSTRA, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika* **82** 1126–1148. MR3736344 https://doi.org/10.1007/s11336-016-9537-6

CAI, L. (2020). flexMIRT. R version 3.62: *Flexible multilevel multidimensional item analysis and test scoring* [Computer software]. Vector Psychometric Group.

CHEN, Y., LI, X., LIU, J. and YING, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika* **83** 538–562. https://doi.org/10.1007/s11336-018-9610-4

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023). DIF statistical inference without knowing anchoring items. *Psychometrika* **88** 1097–1122. MR4668563 https://doi.org/10.1007/s11336-023-09930-9

DE BOECK, P. (2008). Random item IRT models. *Psychometrika* **73** 533–559. MR2469780 https://doi.org/10.1007/s11336-008-9092-x

HOLLAND, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika* **55** 577–601. MR1094332 https://doi.org/10.1007/BF02294609

KANG, I., DE BOECK, P. and RATCLIFF, R. (2022). Modeling conditional dependence of response accuracy and response time with the diffusion item response theory model. *Psychometrika* **87** 725–748. MR4434013 https://doi.org/10.1007/s11336-021-09819-5

MAGIS, D., TUERLINCKX, F. and DE BOECK, P. (2015). Detection of differential item functioning using the lasso approach. *J. Educ. Behav. Stat.* **40** 111–135. https://doi.org/10.3102/1076998614559747

MUTHÉN, B. and ASPAROUHOV, T. (2014). IRT studies of many groups: The alignment method. *Front. Psychol.* **5** 978. https://doi.org/10.3389/fpsyg.2014.00978

SINHARAY, S., MATTHEW, S., JOHNSON, M. S. and WILLIAMSON, D. M. (2003). Calibrating item families and summarizing the results using family expected response functions. *J. Educ. Behav. Stat.* **28** 295–313. https://doi.org/10.3102/10769986028004295

TUTZ, G. and SCHAUBERGER, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika* **80** 21–43. MR3325844 https://doi.org/10.1007/s11336-013-9377-6

VAN DER LINDEN, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika* **72** 287–308. MR2361958 https://doi.org/10.1007/s11336-006-1478-z

WAINER, H. (2010). 14 conversations ayout three things. *Journal of Educational and Behavioral Statistics* **35** 5–25. https://doi.org/10.3102/1076998609355124

WANG, S., DE BOECK, P. and YOTEBIENG, M. (2023). Heywood cases in unidimensional factor models and item response models for binary data. *Appl. Psychol. Meas.* **47** 141–154. https://doi.org/10.1177/01466216231151701

WANG, W.-C. (2004). Effects of anchor item methods on the detection of differential item functioning within the family of Rasch models. *J. Exp. Educ.* **72** 221–261. https://doi.org/10.3200/JEXE.72.3.221-261

WOODS, C. M., CAI, L. and WANG, M. (2013). The langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educ. Psychol. Meas.* **73** 532–547. https://doi.org/10.1177/0013164412464875

WU, H. and BROWNE, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika* **80** 571–600. MR3392016 https://doi.org/10.1007/s11336-015-9451-3

Paul De Boeck is Emeritus Professor, Department of Psychology, The Ohio State University, Columbus, Ohio, USA (e-mail: deboeck.2@osu.edu).

Statistical Science
2025, Vol. 40, No. 2, 201–204
https://doi.org/10.1214/24-STS947

Comment: The Evolution of Items and Responses in IRT

Matthew S. Johnson

REFERENCES

[1] BENNETT, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educ. Assess.* **28** 83–104. https://doi.org/10.1080/10627197.2023.2202312

[2] DEVLIN, J., CHANG, M.-W., LEE, K. and TOUTANOVA, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies* 4171–4186.

[3] DOBRESCU, L. I., HOLDEN, R., MOTTA, A., PICCOLI, A., ROBERTS, P. and WALKER, S. (2021). Cultural context in standardized tests Technical Report No. 2021-80, UNSW Economics Working Paper. Available at https://ssrn.com/abstract=3983663 or https://doi.org/10.2139/ssrn.3983663.

[4] FERNANDEZ, N., GHOSH, A., LIU, N., WANG, Z., CHOFFIN, B., BARANIUK, R. and LAN, A. (2023). Automated scoring for reading comprehension via in-context BERT tuning. arXiv preprint. Available at arXiv:2205.09864.

[5] HOOD, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *J. Negro Educ.* **67** 187–196.

[6] JOHNSON, M. S., LIU, X. and MCCAFFREY, D. F. (2022). Psychometric methods to evaluate measurement and algorithmic bias in automated scoring. *J. Educ. Meas.* **59** 338–361. https://doi.org/10.1111/jedm.12335

[7] RANDALL, J. (2021). "Color-neutral" is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educ. Meas.*, *Issues Pract.* **40** 82–90. https://doi.org/10.1111/emip.12429

[8] SINHARAY, S. and JOHNSON, M. S. (2024). Computation and accuracy evaluation of comparable scores on culturally responsive assessments. *J. Educ. Meas.* **61** 5–46. https://doi.org/10.1111/jedm.12381

[9] SIRECI, S. G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educ. Meas.*, *Issues Pract.* **39** 100–105. https://doi.org/10.1111/emip.12377

[10] WHITMER, J., BEITING-PARRISH, M., BLANKENSHIP, C., FOLWER-DAWSON, A. and PITCHER, M. (2023). NAEP Math Item Automated Scoring Data Challenge Results: High Accuracy and Potential for Additional Insights. https://doi.org/10.35542/osf.io/eyzgd

[11] WHITMER, J., DENG, E., BLANKENSHIP, C., BEITING-PARRISH, M., ZHANG, T. and BAILEY, P. (2023). Results of NAEP Reading Item Automated Scoring Data Challenge (Fall 2021). https://doi.org/10.35542/osf.io/2hevq.

Matthew S. Johnson is Principal Research Director in the Research Division at Educational Testing Service (ETS) in Princeton, New Jersey, USA (e-mail: msjohnson@ets.org).

Statistical Science
2025, Vol. 40, No. 2, 205–210
https://doi.org/10.1214/25-STS974

Rejoinder: Item Response Theory—A Statistical Framework for Educational and Psychological Measurement

Yunxiao Chen, Xiaoou Li, Jingchen Liu and Zhiliang Ying

REFERENCES

ANDERSEN, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc. Ser. B* **32** 283–301. MR0273723

BECHGER, T. M. and MARIS, G. (2015). A statistical test for differential item pair functioning. *Psychometrika* **80** 317–340. MR3353960 https://doi.org/10.1007/s11336-014-9408-y

BOLLEN, K. A. and PEARL, J. (2013). Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research* (S. L. Morgan, ed.) 301–328. Springer, New York, NY.

CHEN, Y. (2020). A continuous-time dynamic choice measurement model for problem-solving process data. *Psychometrika* **85** 1052–1075. MR4204752 https://doi.org/10.1007/s11336-020-09734-1

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023). DIF statistical inference without knowing anchoring items. *Psychometrika* **88** 1097–1122. MR4668563 https://doi.org/10.1007/s11336-023-09930-9

CHEN, Y., LI, X., LIU, J. and YING, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika* **83** 538–562. MR3851946 https://doi.org/10.1007/s11336-018-9610-4

CHEN, Y., LI, X. and ZHANG, S. (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika* **84** 124–146. MR3910911 https://doi.org/10.1007/s11336-018-9646-5

CHEN, Y., LI, X. and ZHANG, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *J. Amer. Statist. Assoc.* **115** 1756–1770. MR4189755 https://doi.org/10.1080/01621459.2019.1635485

COX, D. R. and ISHAM, V. S. (1980). Stochastic Processes, Boca Raton, FL.

DA SILVA, M. A., LIU, R., HUGGINS-MANLEY, A. C. and BAZÁN, J. L. (2019). Incorporating the Q-matrix into multidimensional item response theory models. *Educ. Psychol. Meas.* **79** 665–687. https://doi.org/10.1177/0013164418814898

EMBRETSON, S. E. and YANG, X. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika* **78** 14–36. MR3042817 https://doi.org/10.1007/s11336-012-9296-y

GHOSH, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statist. Probab. Lett.* **23** 165–170. MR1341360 https://doi.org/10.1016/0167-7152(94)00109-L

HOLLAND, P. W. (1988). Differential item performance and the Mantel–Haenszel procedure. In *Test Validity* (H. Wainer and H. I. Braun, eds.) Lawrence Erlbaum Associates, Hillsdale, NJ.

JOHNSON, M. S., LIU, X. and MCCAFFREY, D. F. (2022). Psychometric methods to evaluate measurement and algorithmic bias in automated scoring. *J. Educ. Meas.* **59** 338–361.

MEHRABI, N., MORSTATTER, F., SAXENA, N., LERMAN, K. and GALSTYAN, A. (2021). A survey on bias and fairness in machine learning. *ACM Comput. Surv.* **54** 1–35.

SHEALY, R. and STOUT, W. (1993a). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika* **58** 159–194.

SHEALY, R. T. and STOUT, W. F. (1993b). An item response theory model for test bias and differential test functioning. In *Differential Item Functioning* (P. W. Holland and H. Wainer, eds.) 197–239. Lawrence Erlbaum Associates, Hillsdale, NJ.

WANG, S., DE BOECK, P. and YOTEBIENG, M. (2023). Heywood cases in unidimensional factor models and item response models for binary data. *Appl. Psychol. Meas.* **47** 141–154.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

Yunxiao Chen is Associate Professor, Department of Statistics, London School of Economics and Political Science, London, UK (e-mail: y.chen186@lse.ac.uk). Xiaoou Li is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, USA. Jingchen Liu is Professor, Department of Statistics, Columbia University, New York, New York 10027, USA. Zhiliang Ying is Professor, Department of Statistics, Columbia University, New York, New York 10027, USA.

Statistical Science
2025, Vol. 40, No. 2, 211–218
https://doi.org/10.1214/24-STS924

Demystifying Inferential Models and Confidence Curves: A Fiducial Perspective

Yifan Cui and Jan Hannig

Abstract. Inferential models have recently gained in popularity for valid uncertainty quantification. In this paper, we investigate inferential models by exploring relationships between inferential models, fiducial inference, and confidence curves. In short, we argue that from a certain point of view, inferential models can be viewed as fiducial distribution based confidence curves. We show that all probabilistic uncertainty quantification of inferential models is based on a collection of sets we name principle sets and principle assertions.

Key words and phrases: Confidence curve, confidence distribution, Dempster–Shafer theory, fiducial inference, inferential model.

REFERENCES

[1] AUGUSTIN, T., COOLEN, F. P. A., DE COOMAN, G. and TROFFAES, M. C. M., eds. (2014). *Introduction to Imprecise Probabilities*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3236913 https://doi.org/10.1002/9781118763117

[2] BAYARRI, M. J., BERGER, J. O., FORTE, A. and GARCÍA-DONATO, G. (2012). Criteria for Bayesian model choice with application to variable selection. *Ann. Statist.* **40** 1550–1577. MR3015035 https://doi.org/10.1214/12-AOS1013

[3] BERGER, J. O., BERNARDO, J. M. and SUN, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37** 905–938. MR2502655 https://doi.org/10.1214/07-AOS587

[4] BERGER, J. O., BERNARDO, J. M. and SUN, D. (2012). Objective priors for discrete parameter spaces. *J. Amer. Statist. Assoc.* **107** 636–648. MR2980073 https://doi.org/10.1080/01621459.2012.682538

[5] BIRNBAUM, A. (1961). Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. *J. Amer. Statist. Assoc.* **56** 246–249. MR0121904

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

[7] CASELLA, G. and BERGER, R. L. (2001). *Statistical Inference*, 2nd ed. *The Wadsworth & Brooks/Cole Statistics/Probability Series*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[8] CISEWSKI, J. and HANNIG, J. (2012). Generalized fiducial inference for normal linear mixed models. *Ann. Statist.* **40** 2102–2127. MR3059078 https://doi.org/10.1214/12-AOS1030

[9] COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Stat.* **29** 357–372. MR0094890 https://doi.org/10.1214/aoms/1177706618

[10] CUI, Y. and HANNIG, J. (2019). Nonparametric generalized fiducial inference for survival functions under censoring. *Biometrika* **106** 501–518. MR3992384 https://doi.org/10.1093/biomet/asz016

[11] CUI, Y. and HANNIG, J. (2024). A fiducial approach to the nonparametric deconvolution problem: The discrete case. *Sci. China Math.* **67** 2653–2670. MR4803505 https://doi.org/10.1007/s11425-021-2086-5

[12] CUI, Y., HANNIG, J. and KOSOROK, M. R. (2024). A unified nonparametric fiducial approach to interval-censored data. *J. Amer. Statist. Assoc.* **119** 2230–2241. MR4797936 https://doi.org/10.1080/01621459.2023.2252143

[13] CUI, Y. and XIE, M.-G. (2021). Confidence distribution and distribution estimation for modern statistical inference. arXiv preprint. Available at arXiv:2109.01898.

[14] DEMPSTER, A. P. (2008). The Dempster–Shafer calculus for statisticians. *Internat. J. Approx. Reason.* **48** 365–377. MR2419025 https://doi.org/10.1016/j.ijar.2007.03.004

[15] DUBOIS, D. (2006). Possibility theory and statistical reasoning. *Comput. Statist. Data Anal.* **51** 47–69. MR2297587 https://doi.org/10.1016/j.csda.2006.04.015

[16] DUBOIS, D. and PRADE, H. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, French ed. Plenum Press, New York. MR1104217 https://doi.org/10.1007/978-1-4684-5287-7

[17] EDLEFSEN, P. T., LIU, C. and DEMPSTER, A. P. (2009). Estimating limits from Poisson counting data using Dempster–Shafer analysis. *Ann. Appl. Stat.* **3** 764–790. MR2750681 https://doi.org/10.1214/00-AOAS223

[18] EFRON, B. (2000). R. A. Fisher in the 21st century. *Statist. Sci.* **13** 95–122.

[19] FIELLER, E. C. (1954). Symposium on interval estimation: Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B* **16** 175–185. MR0093076

[20] FISHER, R. A. (1930). Inverse probability. *Proc. Camb. Philos. Soc.* **xxvi** 528–535.

[21] FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen.* 91–98.

[22] FRASER, A. M., FRASER, D. A. S. and STAICU, A.-M. (2010). Second order ancillary: A differential view from continuity.

Yifan Cui is Associate Professor, School of Management and Center for Data Science, Zhejiang University, Hangzhou, Zhejiang 310058, China (e-mail: cuiyf@zju.edu.cn). Jan Hannig is Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA (e-mail: jan.hannig@unc.edu).

Bernoulli **16** 1208–1223. MR2759176 https://doi.org/10.3150/10-BEJ248

[23] FRASER, D. and NADERI, A. (2008). Exponential models: Approximations for probabilities. *Biometrika* **94** 1–9.

[24] FRASER, D., REID, N. and WONG, A. (2005). What a model with data says about theta. *Int. J. Math. Stat. Sci.* **3** 163–178.

[25] FRASER, D. A. S. (2004). Ancillaries and conditional inference. *Statist. Sci.* **19** 333–369. With comments by Ronald W. Butler, Ib M. Skovgaard, Rudolf Beran and a rejoinder by the author. MR2140544 https://doi.org/10.1214/088342304000000323

[26] FRASER, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? *Statist. Sci.* **26** 299–316. MR2918001 https://doi.org/10.1214/11-STS352

[27] FRASER, D. A. S., REID, N., MARRAS, E. and YI, G. Y. (2010). Default priors for Bayesian and frequentist inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 631–654. MR2758239 https://doi.org/10.1111/j.1467-9868.2010.00750.x

[28] HANNIG, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19** 491–544. MR2514173

[29] HANNIG, J., IYER, H., LAI, R. C. S. and LEE, T. C. M. (2016). Generalized fiducial inference: A review and new results. *J. Amer. Statist. Assoc.* **111** 1346–1361. MR3561954 https://doi.org/10.1080/01621459.2016.1165102

[30] LAI, R. C. S., HANNIG, J. and LEE, T. C. M. (2015). Generalized fiducial inference for ultrahigh-dimensional regression. *J. Amer. Statist. Assoc.* **110** 760–772. MR3367262 https://doi.org/10.1080/01621459.2014.931237

[31] LAWLESS, J. F. and FREDETTE, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92** 529–542. MR2202644 https://doi.org/10.1093/biomet/92.3.529

[32] LIU, C. and MARTIN, R. (2015). Frameworks for prior-free posterior probabilistic inference. *Wiley Interdiscip. Rev.: Comput. Stat.* **7** 77–85. MR3348723 https://doi.org/10.1002/wics.1329

[33] LIU, C. and MARTIN, R. (2020). Inferential models and possibility measures. arXiv preprint. Available at arXiv:2008.06874.

[34] LIU, D., LIU, R. Y. and XIE, M. (2014). Exact meta-analysis approach for discrete data and its application to 2×2 tables with rare events. *J. Amer. Statist. Assoc.* **109** 1450–1465. MR3293603 https://doi.org/10.1080/01621459.2014.946318

[35] LIU, D., LIU, R. Y. and XIE, M. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. *J. Amer. Statist. Assoc.* **110** 326–340. MR3338506 https://doi.org/10.1080/01621459.2014.899235

[36] LIU, Y. and HANNIG, J. (2016). Generalized fiducial inference for binary logistic item response models. *Psychometrika* **81** 290–324. MR3505368 https://doi.org/10.1007/s11336-015-9492-7

[37] LIU, Y. and HANNIG, J. (2017). Generalized fiducial inference for logistic graded response models. *Psychometrika* **82** 1097–1125. MR3736343 https://doi.org/10.1007/s11336-017-9554-0

[38] LUO, X., DASGUPTA, T., XIE, M. and LIU, R. Y. (2021). Leveraging the Fisher randomization test using confidence distributions: Inference, combination and fusion learning. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 777–797. MR4320001 https://doi.org/10.1111/rssb.12429

[39] MARTIN, R. (2017). Inferential Models. American Cancer Society, 1–8. Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07997.

[40] MARTIN, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *Internat. J. Approx. Reason.* **113** 39–73. MR3979518 https://doi.org/10.1016/j.ijar.2019.06.005

[41] MARTIN, R. (2021). An imprecise-probabilistic characterization of frequentist statistical inference. arXiv preprint. Available at arXiv:2112.10904.

[42] MARTIN, R. and LIU, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.* **108** 301–313. MR3174621 https://doi.org/10.1080/01621459.2012.747960

[43] MARTIN, R. and LIU, C. (2015). Conditional inferential models: Combining information for prior-free probabilistic inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 195–217. MR3299405 https://doi.org/10.1111/rssb.12070

[44] MARTIN, R. and LIU, C. (2015). Marginal inferential models: Prior-free probabilistic inference on interest parameters. *J. Amer. Statist. Assoc.* **110** 1621–1631. MR3449059 https://doi.org/10.1080/01621459.2014.985827

[45] MARTIN, R. and LIU, C. (2016). *Inferential Models: Reasoning with Uncertainty. Monographs on Statistics and Applied Probability* **147**. CRC Press, Boca Raton, FL. MR3618727

[46] NEUPERT, S. D. and HANNIG, J. (2019). BFF: Bayesian, fiducial, frequentist analysis of age effects in daily diary data. *J. Gerontol., Ser. B* gbz100. https://doi.org/10.1093/geronb/gbz100

[47] NEYMAN, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32** 128–150. MR0005582 https://doi.org/10.1093/biomet/32.2.128

[48] SCHWEDER, T. and HJORT, N. (2003). Frequentist analogues of priors and posteriors. *Econom. Philos. Econ.* 285–317.

[49] SCHWEDER, T. and HJORT, N. L. (2002). Confidence and likelihood. *Scand. J. Stat.* **29** 309–332. MR1909788 https://doi.org/10.1111/1467-9469.00285

[50] SCHWEDER, T. and HJORT, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions. Cambridge Series in Statistical and Probabilistic Mathematics* **41**. Cambridge Univ. Press, New York. MR3558738 https://doi.org/10.1017/CBO9781139046671

[51] SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ. MR0464340

[52] SINGH, K., XIE, M. and STRAWDERMAN, W. E. (2005). Combining information from independent sources through confidence distributions. *Ann. Statist.* **33** 159–183. MR2157800 https://doi.org/10.1214/009053604000001084

[53] SINGH, K., XIE, M. and STRAWDERMAN, W. E. (2007). Confidence distribution (CD)—distribution estimator of a parameter. In *Complex Datasets and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 132–150. IMS, Beachwood, OH. MR2459184 https://doi.org/10.1214/074921707000000102

[54] TIAN, L., WANG, R., CAI, T. and WEI, L.-J. (2011). The highest confidence density region and its usage for joint inferences about constrained parameters. *Biometrics* **67** 604–610. MR2829029 https://doi.org/10.1111/j.1541-0420.2010.01486.x

[55] WANDLER, D. V. and HANNIG, J. (2012). Generalized fiducial confidence intervals for extremes. *Extremes* **15** 67–87. MR2891310 https://doi.org/10.1007/s10687-011-0127-9

[56] WILLIAMS, J. P. and HANNIG, J. (2019). Nonpenalized variable selection in high-dimensional linear model settings via generalized fiducial inference. *Ann. Statist.* **47** 1723–1753. MR3911128 https://doi.org/10.1214/18-AOS1733

[57] WILLIAMS, J. P., STORLIE, C. B., THERNEAU, T. M., JACK, C. R. JR. and HANNIG, J. (2020). A Bayesian approach to multistate hidden Markov models: Application to dementia progression. *J. Amer. Statist. Assoc.* **115** 16–31. MR4078442 https://doi.org/10.1080/01621459.2019.1594831

[58] XIE, M. and SINGH, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int. Stat. Rev.* **81** 3–39. MR3047496 https://doi.org/10.1111/insr.12000

[59] XIE, M., SINGH, K. and STRAWDERMAN, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis.

J. Amer. Statist. Assoc. **106** 320–333. MR2816724 https://doi.org/10.1198/jasa.2011.tm09803
[60] YANG, G., LIU, D., WANG, J. and XIE, M. (2016). Meta-analysis framework for exact inferences with application to the analysis of rare events. *Biometrics* **72** 1378–1386. MR3591622 https://doi.org/10.1111/biom.12497
[61] ZADEH, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* **1** 3–28. MR0480045 https://doi.org/10.1016/0165-0114(78)90029-5

Statistical Science
2025, Vol. 40, No. 2, 219–234
https://doi.org/10.1214/24-STS928

A Geometric Perspective on Bayesian and Generalized Fiducial Inference

Yang Liu, Jan Hannig and Alexander C. Murph

Abstract. Post-data statistical inference concerns making probability statements about model parameters conditional on observed data. When a priori knowledge about parameters is available, post-data inference can be conveniently made from Bayesian posteriors. In the absence of prior information, we may still rely on objective Bayes or generalized fiducial inference (GFI). Inspired by approximate Bayesian computation, we propose a novel characterization of post-data inference with the aid of differential geometry. Under suitable smoothness conditions, we establish that Bayesian posteriors and generalized fiducial distributions (GFDs) can be respectively characterized by absolutely continuous distributions supported on the same differentiable manifold: The manifold is uniquely determined by the observed data and the data generating equation of the fitted model. Our geometric analysis not only sheds light on the connection and distinction between Bayesian inference and GFI, but also allows us to sample from posteriors and GFDs using manifold Markov chain Monte Carlo algorithms. A repeated measures analysis of variance example is presented to illustrate the sampling procedure.

Key words and phrases: Approximate Bayesian computation, Bayesian inference, differentiable manifold, generalized fiducial inference, Markov chain Monte Carlo.

REFERENCES

ALQUIER, P., FRIEL, N., EVERITT, R. and BOLAND, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Stat. Comput.* **26** 29–47. MR3439357 https://doi.org/10.1007/s11222-014-9521-x

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

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

BEAUMONT, M. A., ZHANG, W. and BALDING, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162** 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

BERGER, J. (2006). The case for objective Bayesian analysis. *Bayesian Anal.* **1** 385–402. MR2221271 https://doi.org/10.1214/06-BA115

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

BERGER, J. O., BERNARDO, J. M. and SUN, D. (2015). Overall objective priors. *Bayesian Anal.* **10** 189–221. MR3420902 https://doi.org/10.1214/14-BA915

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

BRUBAKER, M., SALZMANN, M. and URTASUN, R. (2012). A family of MCMC methods on implicitly defined manifolds. In *In Artificial Intelligence and Statistics* 161–172. PMLR.

CHAVEL, I. (2006). *Riemannian Geometry: A Modern Introduction*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **98**. Cambridge Univ. Press, Cambridge. MR2229062 https://doi.org/10.1017/CBO9780511616822

CISEWSKI, J. and HANNIG, J. (2012). Generalized fiducial inference for normal linear mixed models. *Ann. Statist.* **40** 2102–2127. MR3059078 https://doi.org/10.1214/12-AOS1030

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

DEMPSTER, A. P. (1963). Further examples of inconsistencies in the fiducial argument. *Ann. Math. Stat.* **34** 884–891. MR0150865 https://doi.org/10.1214/aoms/1177704011

Yang Liu is an Associate Professor from the Department of Human Development and Quantitative Methodology at University of Maryland, College Park, Maryland 20742, USA (e-mail: yliu87@umd.edu). Jan Hannig is a Professor from the Department of Statistics and Operations Research at the University of North Carolina, Chapel Hill, North Carolina 27599, USA (e-mail: jan.hannig@unc.edu). Alexander C. Murph is a postdoctoral researcher from the Computer, Computational, and Statistical Sciences Division (CCS-6) at the Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA (e-mail: murph@lanl.gov).

DEMPSTER, A. P. (1964). On the difficulities inherent in Fisher's fiducial argument. *J. Amer. Statist. Assoc.* **59** 56–66. MR0158468

DEMPSTER, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Stat.* **37** 355–374. MR0187357 https://doi.org/10.1214/aoms/1177699517

DEMPSTER, A. P. (1968). A generalization of Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **30** 205–232.

DEMPSTER, A. P. (2008). The Dempster–Shafer calculus for statisticians. *Internat. J. Approx. Reason.* **48** 365–377. MR2419025 https://doi.org/10.1016/j.ijar.2007.03.004

DIACONIS, P., HOLMES, S. and SHAHSHAHANI, M. (2013). Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton. Inst. Math. Stat. (IMS) Collect.* **10** 102–125. IMS, Beachwood, OH. MR3586941

FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. CRC Press, London. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

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

FEDERER, H. (1996). *Geometric Measure Theory*. Springer, New York.

FISHER, R. A. (1925). Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* **22** 700–725.

FISHER, R. A. (1930). Inverse probability. *Math. Proc. Cambridge Philos. Soc.* **26** 528–535.

FISHER, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. R. Soc. Lond. Ser. A* **139** 343–348.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen.* **6** 391–398.

FRASER, D. A. S. and NG, K. W. (1980). Multivariate regression analysis with spherical error. *J. Multivariate Anal.* **5** 369–386.

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–534. MR2221284 https://doi.org/10.1214/06-BA117A

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. Taylor & Francis, Boca Raton, FL.

HANNIG, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19** 491–544. MR2514173

HANNIG, J. (2013). Generalized fiducial inference via discretization. *Statist. Sinica* **23** 489–514. MR3086644

HANNIG, J., IYER, H., LAI, R. C. S. and LEE, T. C. M. (2016). Generalized fiducial inference: A review and new results. *J. Amer. Statist. Assoc.* **111** 1346–1361. MR3561954 https://doi.org/10.1080/01621459.2016.1165102

HWANG, C.-R. (1980). Laplace's method revisited: Weak convergence of probability measures. *Ann. Probab.* **8** 1177–1182. MR0602391

KASS, R. E. and WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91** 1343–1370.

LAI, R. C. S., HANNIG, J. and LEE, T. C. M. (2015). Generalized fiducial inference for ultrahigh-dimensional regression. *J. Amer. Statist. Assoc.* **110** 760–772. MR3367262 https://doi.org/10.1080/01621459.2014.931237

LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. MR2954043

LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2010). *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, London. MR2681239 https://doi.org/10.1142/9781848162488

LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2012). Langevin dynamics with constraints and computation of free energy differences. *Math. Comp.* **81** 2071–2125. MR2945148 https://doi.org/10.1090/S0025-5718-2012-02594-4

LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2019). Hybrid Monte Carlo methods for sampling probability measures on submanifolds. *Numer. Math.* **143** 379–421. MR4009691 https://doi.org/10.1007/s00211-019-01056-4

LIU, Y. and HANNIG, J. (2016). Generalized fiducial inference for binary logistic item response models. *Psychometrika* **81** 290–324. MR3505368 https://doi.org/10.1007/s11336-015-9492-7

LIU, Y. and HANNIG, J. (2017). Generalized fiducial inference for logistic graded response models. *Psychometrika* **82** 1097–1125. MR3736343 https://doi.org/10.1007/s11336-017-9554-0

LIU, Y., HANNIG, J. and MURPH, A. C (2025). Supplement to "A Geometric Perspective on Bayesian and Generalized Fiducial Inference." https://doi.org/10.1214/24-STS928SUPP

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

MARTIN, R. and LIU, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.* **108** 301–313. MR3174621 https://doi.org/10.1080/01621459.2012.747960

MARTIN, R. and LIU, C. (2015a). Conditional inferential models: Combining information for prior-free probabilistic inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 195–217. MR3299405 https://doi.org/10.1111/rssb.12070

MARTIN, R. and LIU, C. (2015b). Marginal inferential models: Prior-free probabilistic inference on interest parameters. *J. Amer. Statist. Assoc.* **110** 1621–1631. MR3449059 https://doi.org/10.1080/01621459.2014.985827

MARTIN, R. and LIU, C. (2016). *Inferential Models: Reasoning with Uncertainty. Monographs on Statistics and Applied Probability* **147**. CRC Press, Boca Raton. MR3618727

MATLAB (2021). version 9.11.0 (R2021b). The MathWorks Inc., Natick, Massachusetts.

MURPH, A., HANNIG, J. and WILLIAMS, J. P. (2022a). Introduction to generalized fiducial inference. In *Handbook on Bayesian, Frequentist, and Fiducial Inference* (S. Brooks, A. Gelman, G. Jones and X. L. Meng, eds.). Chapman & Hall, London.

MURPH, A. C., HANNIG, J. and WILLIAMS, J. P. (2022b). Generalized fiducial inference on differentiable manifolds. https://arxiv.org/abs/2209.15473 https://doi.org/10.48550/ARXIV.2209.15473.

POTTHOFF, R. F. and ROY, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51** 313–326. MR0181062 https://doi.org/10.1093/biomet/51.3-4.313

SHI, W. J., HANNIG, J., LAI, R. C. S. and LEE, T. C. M. (2021). Covariance estimation via fiducial inference. *Stat. Theory Relat. Fields* **5** 316–331. MR4335036 https://doi.org/10.1080/24754269.2021.1877950

SISSON, S. A. and FAN, Y. (2011). Likelihood-free MCMC. In *Handbook of Markov Chain Monte Carlo* (S. Brooks,, A. Gelman, G. L. Jones and X. L. Meng, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 313–335. CRC Press, Boca Raton, FL. MR2858454

SISSON, S. A., FAN, Y. and BEAUMONT, M. A. (2018). *Handbook of Approximate Bayesian Computation. Chapman & Hall/CRC Handb. Mod. Stat. Methods* CRC Press, Boca Raton.

STEVENS, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika* **37** 117–129. MR0035955 https://doi.org/10.1093/biomet/37.1-2.117

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–540. MR0898357

ZAPPA, E., HOLMES-CERFON, M. and GOODMAN, J. (2018). Monte Carlo on manifolds: Sampling densities and integrating functions. *Comm. Pure Appl. Math.* **71** 2609–2647. MR3869037 https://doi.org/10.1002/cpa.21783

Statistical Science
2025, Vol. 40, No. 2, 235–258
https://doi.org/10.1214/24-STS929
This research was funded, in whole or in part, by EPSRC, EP/T0184451 and EP/T008650/1.

The Variance-Gamma Distribution: A Review

Adrian Fischer, Robert E. Gaunt and Andrey Sarantsev

Abstract. The variance-gamma (VG) distributions form a four-parameter family which includes as special and limiting cases the normal, gamma and Laplace distributions. Some of the numerous applications include financial modelling and distributional approximation on Wiener space. In this review, we provide an up-to-date account of the basic distributional theory of the VG distribution. Properties covered include probability and cumulative distribution functions, generating functions, moments and cumulants, mode and median, Stein characterisations, representations in terms of other random variables, and a list of related distributions. We also review methods for parameter estimation and some applications of the VG distribution, including the aforementioned applications to financial modelling and distributional approximation on Wiener space.

Key words and phrases: Variance-gamma distribution, distributional theory, estimation, variance-gamma process, financial modelling, approximation on Wiener space.

REFERENCES

[1] ÅBERG, S. and PODGÓRSKI, K. (2011). A class of non-Gaussian second order random fields. *Extremes* **14** 187–222. MR2794324 https://doi.org/10.1007/s10687-010-0119-1

[2] ÅBERG, S., PODGÓRSKI, K. and RYCHLIK, I. (2009). Fatigue damage assessment for a spectral model of non-Gaussian random loads. *Probab. Eng. Mech.* **24** 608–617.

[3] AGUILAR, J.-P. (2020). Some pricing tools for the variance gamma model. *Int. J. Theor. Appl. Finance* **23** 2050025, 35 pp. MR4128081 https://doi.org/10.1142/S0219024920500259

[4] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

[5] ARRAS, B., AZMOODEH, E., POLY, G. and SWAN, Y. (2019). A bound on the Wasserstein-2 distance between linear combinations of independent random variables. *Stochastic Process. Appl.* **129** 2341–2375. MR3958435 https://doi.org/10.1016/j.spa.2018.07.009

[6] AVRAMIDIS, A. N. and L'ECUYER, P. (2006). Efficient Monte Carlo and quasi-Monte Carlo option pricing under the Variance-Gamma model. *Manage. Sci.* **52** 1930–1944.

[7] AZMOODEH, E., EICHELSBACHER, P. and THÄLE, C. (2022). Optimal variance-gamma approximation on the second Wiener chaos. *J. Funct. Anal.* **282** Paper No. 109450, 33 pp. MR4393888 https://doi.org/10.1016/j.jfa.2022.109450

[8] BAI, S. and TAQQU, M. S. (2017). Behavior of the generalized Rosenblatt process at extreme critical exponent values. *Ann. Probab.* **45** 1278–1324. MR3630299 https://doi.org/10.1214/15-AOP1087

[9] BAKSHI, G., CHEN, C. and CAO, Z. (1997). Empirical performance of alternative option pricing models. *J. Finance* **52** 2003–2049.

[10] BARNDORFF-NIELSEN, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Stat.* **5** 151–157. MR0509451

[11] BARNDORFF-NIELSEN, O., KENT, J. and SØRENSEN, M. (1982). Normal variance-mean mixtures and z distributions. *Int. Stat. Rev.* **50** 145–159. MR0678296 https://doi.org/10.2307/1402598

[12] BARNDORFF-NIELSEN, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A* **353** 401–419.

[13] BAXEVANI, A., PODGÓRSKI, K. and WEGENER, J. (2014). Sample path asymmetries in non-Gaussian random processes. *Scand. J. Stat.* **41** 1102–1123. MR3277040 https://doi.org/10.1111/sjos.12086

[14] BEE, M., DICKSON, M. M. and SANTI, F. (2018). Likelihood-based risk estimation for variance-gamma models. *Stat. Methods Appl.* **27** 69–89. MR3769304 https://doi.org/10.1007/s10260-017-0393-z

[15] BIBBY, B. M. and SØRENSEN, M. (2003). Hyperbolic processes in finance. In *Handbook of Heavy Tailed Distributions in Finance* (S. Rachev, ed.) 211–248. Elsevier Science, Amsterdam.

[16] BOGSJÖ, K., PODGÓRSKI, K. and RYCHLIK, I. (2012). Models for road surface roughness. *Veh. Syst. Dyn.* **50** 725–747.

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

Adrian Fischer is Postdoctoral Research Associate, University of Oxford, Oxford, UK (e-mail: adrian.fischer@stats.ox.ac.uk). Robert E. Gaunt is Senior Lecturer, University or Manchester, Manchester, UK (e-mail: robert.gaunt@manchester.ac.uk). Andrey Sarantsev is Assistant Professor, University of Nevada, Reno, Nevada, USA (e-mail: asarantsev@unr.edu).

[18] BREYMANN, W. and LÜTHI, D. (2013). ghyp: A package on generalized hyperbolic distributions. Manual for R Package ghyp.

[19] CARR, P., GEMAN, H., MADAN, D. and YOR, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus*. **75** 305–332.

[20] CARR, P. and MADAN, D. (1999). Option valuation using the fast Fourier transform. *J. Comput. Finance* **2** 61–73.

[21] CERVELLERA, G. P. and TUCCI, M. P. (2017). A note on the estimation of a gamma-variance process: Learning from a failure. *Comput. Econ*. **49** 363–385.

[22] CUI, G., YU, X., IOMMELLI, S. and KONG, L. (2016). Exact distribution for the product of two correlated Gaussian random variables. *IEEE Signal Process. Lett*. **23** 1662–1666.

[23] DAAL, E. A. and MADAN, D. B. (2005). An empirical examination of the variance-gamma model for foreign currency options. *J. Bus*. **78** 2121–2152.

[24] DANG, U. J., BROWNE, R. P. and MCNICHOLAS, P. D. (2015). Mixtures of multivariate power exponential distributions. *Biometrics* **71** 1081–1089. MR3436733 https://doi.org/10.1111/biom.12351

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

[26] EBERLEIN, E. and HAMMERSTEIN, E. A. (2004). Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*. *Progress in Probability* **58** 221–264. Birkhäuser, Basel. MR2096291

[27] EBERLEIN, E. and KELLER, U. (1995). Hyperbolic distributions in finance. *Bernoulli* **1** 281–299.

[28] EBERLEIN, E. and PRAUSE, K. (2002). The generalized hyperbolic model: Financial derivatives and risk measures. In *Mathematical Finance—Bachelier Congress*, 2000 (*Paris*). *Springer Finance* 245–267. Springer, Berlin. MR1960567

[29] EICHELSBACHER, P. and THÄLE, C. (2015). Malliavin–Stein method for variance-gamma approximation on Wiener space. *Electron. J. Probab.* **20** Paper No. 123, 28 pp. MR3425543 https://doi.org/10.1214/EJP.v20-4136

[30] EMBRECHTS, P. (1983). A property of the generalized inverse Gaussian distribution with some applications. *J. Appl. Probab.* **20** 537–544. MR0713503 https://doi.org/10.2307/3213890

[31] FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Stat*. **43** 1634–1643. MR0373022 https://doi.org/10.1214/aoms/1177692395

[32] FINLAY, R. and SENETA, E. (2006). Stationary-increment Student and variance-gamma processes. *J. Appl. Probab.* **43** 441–453. MR2248575 https://doi.org/10.1239/jap/1152413733

[33] FINLAY, R. and SENETA, E. (2008). Stationary-i ncrement variance-gamma and t models: Simulation and parameter estimation. *Int. Stat. Rev.* **76** 167–186.

[34] FIORANI, F. (2004). Option pricing under the variance gamma process. Ph.D. thesis, Univ. Degli Studi Di Trieste.

[35] FU, M. C. (2007). Variance-gamma and Monte Carlo. In *Advances in Mathematical Finance*. *Appl. Numer. Harmon. Anal*. 21–35. Birkhäuser, Boston, MA. MR2359360 https://doi.org/10.1007/978-0-8176-4545-8_2

[36] GAUNT, R. E. (2013). Rates of convergence of variance-gamma approximations via Stein's method. Ph.D. thesis, Univ. Oxford.

[37] GAUNT, R. E. (2014). Variance-gamma approximation via Stein's method. *Electron. J. Probab.* **19** Paper No. 38, 33 pp. MR3194737 https://doi.org/10.1214/EJP.v19-3020

[38] GAUNT, R. E. (2014). Inequalities for modified Bessel functions and their integrals. *J. Math. Anal. Appl*. **420** 373–386. MR3229830 https://doi.org/10.1016/j.jmaa.2014.05.083

[39] GAUNT, R. E. (2017). A Stein characterisation of the generalized hyperbolic distribution. *ESAIM Probab. Stat*. **21** 303–316. MR3743916 https://doi.org/10.1051/ps/2017007

[40] GAUNT, R. E. (2018). Inequalities for integrals of modified Bessel functions and expressions involving them. *J. Math. Anal. Appl*. **462** 172–190. MR3771238 https://doi.org/10.1016/j.jmaa.2018.02.003

[41] GAUNT, R. E. (2019). A note on the distribution of the product of zero-mean correlated normal random variables. *Stat. Neerl*. **73** 176–179. MR3942099 https://doi.org/10.1111/stan.12152

[42] GAUNT, R. E. (2022). Stein factors for variance-gamma approximation in the Wasserstein and Kolmogorov distances. *J. Math. Anal. Appl*. **514** Paper No. 126274, 32 pp. MR4413319 https://doi.org/10.1016/j.jmaa.2022.126274

[43] GAUNT, R. E. (2022). The basic distributional theory for the product of zero mean correlated normal random variables. *Stat. Neerl*. **76** 450–470. MR4495630

[44] GAUNT, R. E. (2023). On the moments of the variance-gamma distribution. *Statist. Probab. Lett*. **201** Paper No. 109884, 4 pp. MR4603853 https://doi.org/10.1016/j.spl.2023.109884

[45] GAUNT, R. E. (2024). On the cumulative distribution function of the variance-gamma distribution. *Bull. Aust. Math. Soc*. To appear.

[46] GAUNT, R. E. and LI, S. (2023). Bounding Kolmogorov distances through Wasserstein and related integral probability metrics. *J. Math. Anal. Appl*. **522** Paper No. 126985, 24 pp. MR4533915 https://doi.org/10.1016/j.jmaa.2022.126985

[47] GAUNT, R. E. and LI, S. (2023). The variance-gamma ratio distribution. *C. R. Math. Acad. Sci. Paris* **361** 1151–1161. MR4659075 https://doi.org/10.5802/crmath.495

[48] GAUNT, R. E. and LI, S. (2024). The distribution of the product of independent variance-gamma random variables. *J. Math. Anal. Appl*. **539** Article 128530.

[49] GAUNT, R. E. and MERKLE, M. (2021). On bounds for the mode and median of the generalized hyperbolic and related distributions. *J. Math. Anal. Appl*. **493** Paper No. 124508, 19 pp. MR4141494 https://doi.org/10.1016/j.jmaa.2020.124508

[50] GAUNT, R. E., MIJOULE, G. and SWAN, Y. (2019). An algebra of Stein operators. *J. Math. Anal. Appl*. **469** 260–279. MR3857522 https://doi.org/10.1016/j.jmaa.2018.09.015

[51] GRADSHETYN, I. S. and RYZHIK, I. M. (2007). *Table of Integrals, Series and Products*, 7th ed. Academic Press, San Diego.

[52] GROENEVELD, R. A. and MEEDEN, G. (1977). The mode, median, and mean inequality. *Amer. Statist*. **31** 120–121. MR0448635 https://doi.org/10.2307/2682959

[53] GU, M., WANG, X. and BERGER, J. O. (2018). Robust Gaussian stochastic process emulation. *Ann. Statist*. **46** 3038–3066. MR3851764 https://doi.org/10.1214/17-AOS1648

[54] GUTTORP, P. and GNEITING, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika* **93** 989–995. MR2285084 https://doi.org/10.1093/biomet/93.4.989

[55] HAMMERSTEIN, E. A. V. (2010). Generalized hyperbolic distributions: Theory and applications to CDO pricing. Ph.D. thesis, Univ. Freiburg.

[56] HEYDE, C. C. (1999). A risky asset model with strong dependence through fractal activity time. *J. Appl. Probab.* **36** 1234–1239. MR1746407 https://doi.org/10.1239/jap/1032374769

[57] HIRSA, A. and MADAN, D. B. (2004). Pricing American options under variance gamma. *J. Comput. Finance* **7** 63–80.

[58] HOLM, H. and ALOUINI, M.-S. (2004). Sum and difference of two squared correlated Nakagami variates with the McKay distribution. *IEEE Trans. Commun*. **52** 1367–1376.

[59] IVANOV, R. V. (2018). Option pricing in the variance-gamma model under the drift jump. *Int. J. Theor. Appl. Finance* **21** 1850018, 19 pp. MR3817272 https://doi.org/10.1142/S0219024918500188

[60] JANKOV MAŠIREVIĆ, D. and POGÁNY, T. K. (2021). On new formulae for cumulative distribution function for McKay Bessel distribution. *Comm. Statist. Theory Methods* **50** 143–160. MR4195014 https://doi.org/10.1080/03610926.2019.1632898

[61] JOHANNESSON, P., PODGÓRSKI, K., RYCHLIK, I. and SHARIATI, N. (2016). AR(1) time series with autoregressive gamma variance for road topography modeling. *Probab. Eng. Mech.* **43** 106–116.

[62] KOLLO, T. (2005). *Advanced Multivariate Statistics with Matrices*. Springer, Berlin.

[63] KOPONEN, I. (1995). Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E* **52** 1197–1199.

[64] KORN, R., KORN, E. and KROISANDT, G. (2010). *Monte Carlo Methods and Models in Finance and Insurance*. *Chapman & Hall/CRC Financial Mathematics Series*. CRC Press, Boca Raton, FL. MR2599032 https://doi.org/10.1201/9781420076196

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

[66] KOZUBOWSKI, T. J., MAZUR, S. and PODGÓRSKI, K. (2023). Matrix variate generalized asymmetric Laplace distributions. *Theory Probab. Math. Statist.* **109** 55–80. MR4652994 https://doi.org/10.1090/tpms/1197

[67] KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). Asymmetric Laplace laws and modeling financial data. *Math. Comput. Modelling* **34** 1003–1021. MR1858834 https://doi.org/10.1016/S0895-7177(01)00114-5

[68] KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2012). Laplace probability distributions and related stochastic processes. In *Probability*: *Interpretation*, *Theory and Applications* (Y. Shmaliy, ed.) 105–145. Nova Science Publishers, Inc., New York.

[69] KOZUBOWSKI, T. J., PODGÓRSKI, K. and RYCHLIK, I. (2013). Multivariate generalized Laplace distribution and related random fields. *J. Multivariate Anal.* **113** 59–72. MR2984356 https://doi.org/10.1016/j.jmva.2012.02.010

[70] KÜCHLER, U. and TAPPE, S. (2013). Tempered stable distributions and processes. *Stochastic Process. Appl.* **123** 4256–4293. MR3096354 https://doi.org/10.1016/j.spa.2013.06.012

[71] KYPRIANOU, A. E. (2014). *Fluctuations of Lévy Processes with Applications*: *Introductory Lectures*, 2nd ed. *Universitext*. Springer, Heidelberg. MR3155252 https://doi.org/10.1007/978-3-642-37632-0

[72] LAM, K., CHANG, E. and LEE, M. C. (2002). Option pricing in the variance-gamma model under the drift jump. *Pac.-Basin Finance J.* **10** 267–285.

[73] LEONENKO, N. N., PETHERICK, S. and SIKORSKII, A. (2011). The Student subordinator model with dependence for risky asset returns. *Comm. Statist. Theory Methods* **40** 3509–3523. MR2860754 https://doi.org/10.1080/03610926.2011.581175

[74] LINDERS, D. and STASSEN, B. (2016). The multivariate variance gamma model: Basket option pricing and calibration. *Quant. Finance* **16** 555–572. MR3473973 https://doi.org/10.1080/14697688.2015.1043934

[75] LOREGIAN, A., MERCURI, L. and RROJI, E. (2012). Approximation of the variance gamma model with a finite mixture of normals. *Statist. Probab. Lett.* **82** 217–224. MR2875204 https://doi.org/10.1016/j.spl.2011.10.004

[76] LUCIANO, E., MARENA, M. and SEMERARO, P. (2016). Dependence calibration and portfolio fit with factor-based subordinators. *Quant. Finance* **16** 1037–1052. MR3516138 https://doi.org/10.1080/14697688.2015.1114661

[77] LUCIANO, E. and SCHOUTENS, W. (2006). A multivariate jump-driven financial asset model. *Quant. Finance* **6** 385–402. MR2261218 https://doi.org/10.1080/14697680600806275

[78] LUCIANO, E. and SEMERARO, P. (2010). Multivariate time changes for Lévy asset models: Characterization and calibration. *J. Comput. Appl. Math.* **233** 1937–1953. MR2564029 https://doi.org/10.1016/j.cam.2009.08.119

[79] MA, P. and BHADRA, A. (2023). Beyond Matérn: On a class of interpretable confluent hypergeometric covariance functions. *J. Amer. Statist. Assoc.* **118** 2045–2058. MR4646625 https://doi.org/10.1080/01621459.2022.2027775

[80] MADAN, D. B., CARR, P. and CHANG, E. C. (1998). The variance gamma process and option pricing. *Eur. Finance Rev.* **2** 74–105.

[81] MADAN, D. B. and MILNE, F. (1991). Option pricing with V.G. martingale components. *Math. Finance* **1** 39–55.

[82] MADAN, D. B. and SENETA, E. (1987). Simulation of estimates using the empirical characteristic function. *Int. Stat. Rev.* **55** 153–161. MR0963336 https://doi.org/10.2307/1403191

[83] MADAN, D. B. and SENETA, E. (1987). Chebyshev polynomial approximations and characteristic function estimation. *J. Roy. Statist. Soc. Ser. B* **49** 163–169. MR0905187

[84] MADAN, D. B. and SENETA, E. (1990). The variance gamma (V.G.) model for share market returns. *J. Bus.* **63** 511–524.

[85] MAEJIMA, M. and TUDOR, C. A. (2012). Selfsimilar processes with stationary increments in the second Wiener chaos. *Probab. Math. Statist.* **32** 167–186. MR2959876

[86] MANGILLI, A., PLASZCZYNSKI, S. and TRISTRAM, M. (2015). Large-scale cosmic microwave background temperature and polarization cross-spectra likelihoods. *Mon. Not. R. Astron. Soc.* **453** 3174–3189.

[87] MATÉRN, B. (1960). *Spatial Variation*: *Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Statens Skogsforskningsinstitut, Stockholm. MR0169346

[88] MCKAY, A. T. (1932). A Bessel function distribution. *Biometrika* **24** 39–44.

[89] MCLEISH, D. L. (1982). A robust alternative to the normal distribution. *Canad. J. Statist.* **10** 89–102. MR0673669 https://doi.org/10.2307/3314901

[90] NADARAJAH, S. and KOTZ, S. (2006). The Bessel ratio distribution. *C. R. Math. Acad. Sci. Paris* **343** 531–534. MR2267589 https://doi.org/10.1016/j.crma.2006.09.031

[91] NADARAJAH, S. and POGÁNY, T. K. (2016). On the distribution of the product of correlated normal random variables. *C. R. Math. Acad. Sci. Paris* **354** 201–204. MR3456899 https://doi.org/10.1016/j.crma.2015.10.019

[92] NADARAJAH, S., SRIVASTAVA, H. M. and GUPTA, A. K. (2007). Skewed Bessel function distributions with application to rainfall data. *Statistics* **41** 333–344. MR2369750 https://doi.org/10.1080/02331880701270606

[93] NITITHUMBUNDIT, T. and CHAN, J. S. (2015). An ECM algorithm for skewed multivariate variance gamma distribution in normal mean-variance representation. Preprint. Available at arXiv:1504.01239.

[94] NITITHUMBUNDIT, T. and CHAN, J. S. K. (2020). ECM algorithm for auto-regressive multivariate skewed variance gamma model with unbounded density. *Methodol. Comput. Appl. Probab.* **22** 1169–1191. MR4129130 https://doi.org/10.1007/s11009-019-09762-0

[95] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

[96] NOURDIN, I. and PECCATI, G. (2015). The optimal fourth moment theorem. *Proc. Amer. Math. Soc.* **143** 3123–3133. MR3336636 https://doi.org/10.1090/S0002-9939-2015-12417-3

[97] NUALART, D. and PECCATI, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. MR2118863 https://doi.org/10.1214/009117904000000621

[98] NZOKEM, A. H. (2021). Gamma variance model: Fractional Fourier transform (FRFT). *J. Phys., Conf. Ser.* **2090** 012094.

[99] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W., eds. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge. MR2723248

[100] PEARSON, K., JEFFEREY, G. B. and ELDERTON, E. M. (1929). On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population. *Biometrika* **21** 164–193.

[101] PEARSON, K., STOUFFER, S. A. and DAVID, F. N. (1932). Further applications in statistics of the $T_m(x)$ Bessel function. *Biometrika* **24** 293–350.

[102] PODGÓRSKI, K. and WEGENER, J. (2011). Estimation for stochastic models driven by Laplace motion. *Comm. Statist. Theory Methods* **40** 3281–3302. MR2860820 https://doi.org/10.1080/03610926.2010.499051

[103] PRESS, S. J. (1967). On the sample covariance from a bivariate normal distribution. *Ann. Inst. Statist. Math.* **19** 355–361. MR0219160 https://doi.org/10.1007/BF02911686

[104] RAMACHANDRAN, B. (1997). On geometric-stable laws, a related property of stable processes, and stable densities of exponent one. *Ann. Inst. Statist. Math.* **49** 299–313. MR1463308 https://doi.org/10.1023/A:1003115013644

[105] RATHGEBER, A. W., JOHANNES, S. and STÖCKL, S. (2016). Modeling share returns—an empirical study on the variance gamma model. *J. Econ. Finance* **40** 653–682.

[106] REED, W. J. (2006). The normal-Laplace distribution and its relatives. In *Advances in Distribution Theory, Order Statistics, and Inference. Stat. Ind. Technol.* 61–74. Birkhäuser, Boston, MA. MR2226215 https://doi.org/10.1007/0-8176-4487-3_4

[107] REED, W. J. (2007). Brownian-Laplace motion and its use in financial modelling. *Comm. Statist. Theory Methods* **36** 473–484. MR2405311 https://doi.org/10.1080/03610920601001766

[108] RUIZ-ANTOLÍN, D. and SEGURA, J. (2016). A new type of sharp bounds for ratios of modified Bessel functions. *J. Math. Anal. Appl.* **443** 1232–1246. MR3514344 https://doi.org/10.1016/j.jmaa.2016.06.011

[109] RYDBERG, T. H. (1999). Generalized hyperbolic diffusion processes with applications in finance. *Math. Finance* **9** 183–201. MR1849003 https://doi.org/10.1111/1467-9965.00067

[110] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. MR1739520

[111] SCOTT, D. and DONG, C. Y. (2018). Package 'VarianceGamma'.

[112] SCOTT, D. J., WÜRTZ, D., DONG, C. and TRAN, T. T. (2011). Moments of the generalized hyperbolic distribution. *Comput. Statist.* **26** 459–476. MR2833142 https://doi.org/10.1007/s00180-010-0219-z

[113] SEGURA, J. (2011). Bounds for ratios of modified Bessel functions and associated Turán-type inequalities. *J. Math. Anal. Appl.* **374** 516–528. MR2729238 https://doi.org/10.1016/j.jmaa.2010.09.030

[114] SEMERARO, P. (2008). A multivariate variance gamma model for financial applications. *Int. J. Theor. Appl. Finance* **11** 1–18. MR2398464 https://doi.org/10.1142/S0219024908004701

[115] SENETA, E. (2004). Fitting the variance-gamma model to financial data. *J. Appl. Probab.* **41** 177–187. MR2057573 https://doi.org/10.1239/jap/1082552198

[116] SENETA, E. (2007). The early years of the variance-gamma process. In *Advances in Mathematical Finance. Appl. Numer. Harmon. Anal.* 3–19. Birkhäuser, Boston, MA. MR2359359 https://doi.org/10.1007/978-0-8176-4545-8_1

[117] SICHEL, H. S. (1973). Statistical valuation of diamondiferous deposits. *J. S. Afr. Inst. Min. Metall.* **73** 235–243.

[118] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

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

[120] TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329 https://doi.org/10.1007/BF00532868

[121] TEICHROEW, D. (1957). The mixture of normal distributions with different variances. *Ann. Math. Stat.* **28** 510–512. MR0089551 https://doi.org/10.1214/aoms/1177706981

[122] TJETJEP, A. and SENETA, E. (2008). Skewed normal variance-mean models for asset pricing and the method of moments. *Int. Stat. Rev.* **74** 109–126.

[123] TOMY, L. and JOSE, K. K. (2009). Generalized normal-Laplace AR process. *Statist. Probab. Lett.* **79** 1615–1620. MR2547420 https://doi.org/10.1016/j.spl.2009.04.003

[124] TRINDADE, A. A., ZHU, Y. and ANDREWS, B. (2010). Time series models with asymmetric Laplace innovations. *J. Stat. Comput. Simul.* **80** 1317–1333. MR2745603 https://doi.org/10.1080/00949650903071088

[125] VAN ZWET, W. R. (1979). Mean, median, mode. II. *Stat. Neerl.* **33** 1–5. MR0533245 https://doi.org/10.1111/j.1467-9574.1979.tb00657.x

[126] WEISSTEIN, E. W. Chi-squared distribution. From MathWorld—A Wolfram web resource. Available at https://mathworld.wolfram.com/Chi-SquaredDistribution.html. Last visited on 09/03/2023.

[127] WU, F. (2008). Applications of the normal Laplace and generalized normal Laplace distributions. MSc thesis, Univ. Victoria.

[128] YAMAZATO, M. (1978). Unimodality of infinitely divisible distribution functions of class L. *Ann. Probab.* **6** 523–531. MR0482941

Statistical Science
2025, Vol. 40, No. 2, 259–271
https://doi.org/10.1214/24-STS930

Catalytic Priors: Using Synthetic Data to Specify Prior Distributions in Bayesian Analysis

Dongming Huang, Feicheng Wang, Donald B. Rubin and S. C. Kou

Abstract. Catalytic prior distributions provide general, easy-to-use, and interpretable specifications of prior distributions for Bayesian analysis. They are particularly beneficial when the observed data are inadequate to stably estimate a complex target model. A catalytic prior distribution is constructed by augmenting the observed data with synthetic data that are sampled from the predictive distribution of a simpler model estimated from the observed data. We illustrate the usefulness of the catalytic prior approach using an example from labor economics. In the example, the resulting Bayesian inference reflects many important aspects of the observed data, and the estimation accuracy and predictive performance of the inference based on the catalytic prior are superior to, or comparable to, that of other commonly used prior distributions. We further explore the connection between the catalytic prior approach and a few popular regularization methods. We expect the catalytic prior approach to be useful in many applications.

Key words and phrases: Synthetic data, insufficient data, stable estimation, regularization methods.

REFERENCES

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

BOX, G. E. P. (1979). Robustness in the strategy of scientific model building. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds.) 201–236. Academic Press, San Diego. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B* **26** 211–243. MR0192611

CLOGG, C. C., RUBIN, D. B., SCHENKER, N., SCHULTZ, B. and WEIDMAN, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *J. Amer. Statist. Assoc.* **86** 68–78.

DEMPSTER, A. P. (1968). A generalization of Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **30** 205–232. MR0238428

FRANK, L. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–135.

FRIEDLANDER, D. and GUERON, J. M. (1992). Are high-cost services more effective than low-cost services? In *Evaluating Welfare and Training Programs* 143–98. Harvard Univ. Press, Cambridge.

FRIEDLANDER, D. and HAMILTON, G. (1993). *The Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study*. Manpower Demonstration Research Corporation, New York.

FRIEDLANDER, D. and ROBINS, P. K. (1995). Evaluating program evaluations: New evidence on commonly used nonexperimental methods. *Amer. Econ. Rev.* **85** 923–937.

GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2** 1360–1383. MR2655663 https://doi.org/10.1214/08-AOAS191

HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–960. MR0867618

HOTZ, V. J., IMBENS, G. W. and MORTIMER, J. H. (2005). Predicting the efficacy of future training programs using past experiences at other locations. *J. Econometrics* **125** 241–270. MR2143377 https://doi.org/10.1016/j.jeconom.2004.04.009

HUANG, D., STEIN, N., RUBIN, D. B. and KOU, S. C. (2020). Catalytic prior distributions with application to generalized linear models. *Proc. Natl. Acad. Sci. USA* **117** 12004–12010. MR4240265 https://doi.org/10.1073/pnas.1920913117

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—For Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

Dongming Huang is Assistant Professor, Department of Statistics and Data Science, National University of Singapore, Singapore 117546, Singapore (e-mail: stahd@nus.edu.sg). Feicheng Wang is Ph.D. graduate, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA (e-mail: fwangfeicheng@gmail.com). Donald B. Rubin is Emeritus Professor of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA (e-mail: dbrubin@fas.harvard.edu). S. C. Kou is Professor of Statistics and Professor of Biostatistics, Harvard University, Cambridge, Massachusetts 02138, USA (e-mail: kou@stat.harvard.edu).

JACKSON, D. D. (1979). The use of a priori data to resolve non-uniqueness in linear inversion. *Geophys. J. Int.* **57** 137–157.

LEAMER, E. E. and CHAMBERLAIN, G. (1976). A Bayesian interpretation of pretesting. *J. Roy. Statist. Soc. Ser. B* **38** 85–94. MR0400521

MARQUARDT, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **12** 591–612.

PRATT, J. W. (1965). Bayesian interpretation of standard inference statements. *J. Roy. Statist. Soc. Ser. B* **27** 169–203. MR0196830

PRATT, J. W., RAIFFA, H. and SCHLAIFER, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *J. Amer. Statist. Assoc.* **59** 353–375. MR0161444

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

RUBIN, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *J. Amer. Statist. Assoc.* **75** 591–593.

SAVAGE, L. J., BARNARD, G., CORNFIELD, J., BROSS, I., BOX, G. E. P., GOOD, I. J., LINDLEY, D. V., CLUNIES-ROSS, C. W., PRATT, J. W. et al. (1962). On the foundations of statistical inference: Discussion. *J. Amer. Statist. Assoc.* **57** 307–326.

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

VON NEUMANN, J. (1947). The mathematician. In *The Works of the Mind* (R. B. Heywood, ed.) **I** 180–196. Univ. Chicago Press, Chicago, IL. MR0021929

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

Statistical Science
2025, Vol. 40, No. 2, 272–288
https://doi.org/10.1214/24-STS932

On the Use of Auxiliary Variables in Multilevel Regression and Poststratification

Yajuan Si

Abstract. Multilevel regression and poststratification (MRP) is a popular method for addressing selection bias in subgroup estimation, with broad applications across fields from social sciences to public health. In this paper, we examine the inferential validity of MRP in finite populations, evaluating the impact of poststratification and model specification. The success of MRP relies on the availability of auxiliary information that is strongly related to the outcome. To enhance the fitting performance of the outcome model, we recommend modeling the inclusion probabilities conditionally on auxiliary variables and incorporating flexible functions of estimated inclusion probabilities as predictors in the mean structure. We present a statistical data integration framework that yields robust inferences for probability and nonprobability surveys, addressing various challenges in practical applications. Our simulation studies indicate the statistical validity of MRP, which involves a tradeoff between bias and variance, with greater benefits for subgroup estimates with small sample sizes, compared to alternative methods. We have applied our methods to the Adolescent Brain Cognitive Development (ABCD) Study, which collected information on U.S. children across multiple geographic locations to provide national representation, but is subject to selection bias as a nonprobability sample. We focus on the cognition assessment measure of diverse groups of children in the ABCD study and show that the use of auxiliary variables affects the findings on cognitive performance.

Key words and phrases: Data integration, nonprobability sample, robust inference, model-based, selection/nonresponse bias.

REFERENCES

[1] BAKER, R., BRICK, J. M., BATES, N. A., BATTAGLIA, M., COUPER, M. P., DEVER, J. A., GILE, K. J. and TOURANGEAU, R. (2013). Summary report of the AAPOR task force on nonprobability sampling. *J. Surv. Stat. Methodol.* **1** 90–143.

[2] BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

[3] BETHLEHEM, J. G. (2001). Weighting nonresponse adjustments based on auxiliary information. In *Survey Nonresponse* R. M. Groves, (D. A. Dillman, J. L. Eltinge and R. J. A. Little, eds.). Wiley, New York.

[4] BRADLEY, R. H. and CORWYN, R. F. (2002). Socioeconomic status and child development. *Annu. Rev. Psychol.* **53** 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

[5] BREIDT, F. J., CLAESKENS, G. and OPSOMER, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika* **92** 831–846. MR2234189 https://doi.org/10.1093/biomet/92.4.831

[6] BREIDT, F. J. and OPSOMER, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statist. Sci.* **32** 190–205. MR3648955 https://doi.org/10.1214/16-STS589

[7] BRONIECKI, P., LEEMANN, L. and WÜEST, R. (2022). Improved multilevel regression with poststratification through machine learning (autoMrP). *J. Polit.* **84** 597–601.

[8] BUTTICE, M. K. and HIGHTON, B. (2013). How does multilevel regression and poststratification perform with conventional national surveys? *Polit. Anal.* **21** 449–467.

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

[10] COMPTON, W. M., DOWLING, G. J. and GARAVAN, H. (2019). Ensuring the best use of data: The adolescent brain cognitive development study. *JAMA, J. Amer. Med. Assoc. Southeast Asia, Suppl., Pediatr.* **173** 809–810. https://doi.org/10.1001/jamapediatrics.2019.2081

Yajuan Si is Research Associate Professor, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48104, USA (e-mail: yajuan@umich.edu).

[11] COVELLO, L., GELMAN, A., SI, Y. and WANG, S. (2021). Routine hospital-based SARS-CoV-2 testing outperforms state-based data in predicting clinical burden. *Epidemiology* **32** 792–799.
[12] DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804 https://doi.org/10.1080/01621459.1992.10475217
[13] DONG, Q., ELLIOTT, M. R. and RAGHUNATHAN, T. E. (2014). A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Surv. Methodol.* **40** 29–46.
[14] DOWNES, M. and CARLIN, J. B. (2020). Multilevel regression and poststratification versus survey sample weighting for estimating population quantities in large population health studies. *Amer. J. Epidemiol.* **189** 717–725. https://doi.org/10.1093/aje/kwaa053
[15] DOWNES, M., GURRIN, L. C., ENGLISH, D. R., PIRKIS, J., CURRIER, D., SPITTAL, M. J. and CARLIN, J. B. (2018). Multilevel regression and poststratification: A modeling approach to estimating population quantities from highly selected survey samples. *Amer. J. Epidemiol.* **187** 1780–1790. https://doi.org/10.1093/aje/kwy070
[16] ELLIOTT, M. R. and VALLIANT, R. (2017). Inference for nonprobability samples. *Statist. Sci.* **32** 249–264. MR3648958 https://doi.org/10.1214/16-STS598
[17] FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019
[18] GARAVAN, H., BARTSCH, H., CONWAY, K., DECASTRO, A., GOLDSTEIN, R. Z., HEERINGA, S., JERNIGAN, T., POTTER, A., THOMPSON, W. et al. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Dev. Cogn. Neurosci.* **32** 16–22. https://doi.org/10.1016/j.dcn.2018.04.004
[19] GELMAN, A. and CARLIN, J. B. (2001). Poststratification and weighting adjustments. In *Survey Nonresponse* (R. Groves, D. Dillman, J. Eltinge and R. J. Little, eds.) Wiley, New York.
[20] GELMAN, A. and CARPENTER, B. (2020). Bayesian analysis of tests with unknown specificity and sensitivity. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 1269–1283. MR4166866 https://doi.org/10.1111/rssc.12435
[21] GELMAN, A., GOEL, S., RIVERS, D. and ROTHSCHILD, D. (2016). The mythical swing voter. *Q. J. Polit. Sci.* **11** 103–130.
[22] GELMAN, A. and HILL, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, Cambridge.
[23] GELMAN, A. and LITTLE, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Surv. Methodol.* **23** 127–135.
[24] GHITZA, Y. and GELMAN, A. (2013). Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *Amer. J. Polit. Sci.* **57** 762–776.
[25] GHOSH, M. (2020). Small area estimation: Its evolution in five decades. *Stat. Transit.* **21** 1–22.
[26] GHOSH, M. and MEEDEN, G. (1997). *Bayesian Methods for Finite Population Sampling*. *Monographs on Statistics and Applied Probability* **79**. CRC Press, London. MR1469494 https://doi.org/10.1007/978-1-4899-3416-1
[27] GOODRICH, B. (2022). rstanarm: Bayesian applied regression modeling via Stan version 2.21.3. Available at https://cran.r-project.org/web/packages/rstanarm.
[28] HEERINGA, S. G. and BERGLUND, P. A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) study baseline data. https://doi.org/10.1101/2020.02.10.942011
[29] HOLT, D. and SMITH, T. M. F. (1979). Post stratification. *J. R. Stat. Soc., A* **142** 33–46. MR0626773
[30] KALTON, G. and KASPRZYK, D. (1988). The treatment of missing survey data. *Surv. Methodol.* **12** 1–16.
[31] KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227
[32] KASTELLEC, J. P., LAX, J. R., MALECKI, M. and PHILLIPS, J. H. (2015). Polarizing the electoral connection: Partisan representation in Supreme Court confirmation politics. *J. Polit.* **77** 787–804.
[33] KEIDING, N. and LOUIS, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. Roy. Statist. Soc. Ser. A* **179** 319–376. MR3461587 https://doi.org/10.1111/rssa.12136
[34] KIM, J. K., PARK, S., CHEN, Y. and WU, C. (2021). Combining non-probability and probability survey samples through mass imputation. *J. Roy. Statist. Soc. Ser. A* **184** 941–963. MR4305566 https://doi.org/10.1111/rssa.12696
[35] KIM, J.-K. and TAM, S.-M. (2021). Data integration by combining big data and survey sample data for finite population inference. *Int. Stat. Rev.* **89** 382–401. MR4411911 https://doi.org/10.1111/insr.12434
[36] LAHIRI, P. and MUKHERJEE, K. (2007). On the design-consistency property of hierarchical Bayes estimators in finite population sampling. *Ann. Statist.* **35** 724–737. MR2336865 https://doi.org/10.1214/009053606000001262
[37] LAU, A. (2018). Comparing MRP to raking for online opt-in polls. Available at https://medium.com/pew-research-center-decoded/comparing-mrp-to-raking-for-online-opt-in-polls-b05444d9931d.
[38] LAUDERDALE, B. E., BAILEY, D., BLUMENAU, J. and RIVERS, D. (2020). Model-based pre-election polling for national and sub-national outcomes in the US and UK. *Int. J. Forecast.* **36** 399–413.
[39] LAX, J. R. and PHILLIPS, J. H. (2009). Gay rights in the states: Public opinion and policy responsiveness. *Amer. Polit. Sci. Rev.* **103** 367–386.
[40] LAX, J. R. and PHILLIPS, J. H. (2009). How should we estimate public opinion in the states? *Amer. J. Polit. Sci.* **53** 107–121.
[41] LEEMANN, L. and WASSERFALLEN, F. (2017). Extending the use and prediction precision of subnational public opinion estimation. *Amer. J. Polit. Sci.* **61** 1003–1022.
[42] LI, K. and SI, Y. (2024). Embedded multilevel regression and poststratification: Model-based inference with incomplete auxiliary information. *Stat. Med.* **43** 256–278. MR4684815
[43] LITTLE, R. J. A. (1983). Comment on "An evaluation of model-dependent and probability-sampling inferences in sample surveys", by M. H. Hansen, W. G. Madow and B. J. Tepping. *J. Amer. Statist. Assoc.* **78** 797–799. MR0664675
[44] LITTLE, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *Int. Stat. Rev.* **54** 139–157.
[45] LITTLE, R. J. A. (1993). Post-stratification: A modeler's perspective. *J. Amer. Statist. Assoc.* **88** 1001–1012.
[46] LITTLE, R. and AN, H. (2004). Robust likelihood-based analysis of multivariate data with missing values. *Statist. Sinica* **14** 949–968. MR2089342
[47] LITTLE, R. J. A. and VARTIVARIAN, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Surv. Methodol.* **31** 161–168.
[48] LITTLE, R. J. A., WEST, B. T., BOONSTRA, P. S. and HU, J. (2020). Measures of the degree of departure from ignorable sample selection. *J. Surv. Stat. Methodol.* **8** 932–964.

[49] LITTLE, R. J. A. and WU, M.-M. (1991). Models for contingency tables with known margins when target and sampled populations differ. *J. Amer. Statist. Assoc.* **86** 87–95.

[50] MAKELA, S., SI, Y. and GELMAN, A. (2018). Bayesian inference under cluster sampling with probability proportional to size. *Stat. Med.* **37** 3849–3868. MR3869158 https://doi.org/10.1002/sim.7892

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

[52] ORNSTEIN, J. T. (2020). Stacked regression and poststratification. *Polit. Anal.* **28** 293–301.

[53] RAFEI, A., FLANNAGAN, C. A. C. and ELLIOTT, M. R. (2018). Big data for finite population inference: Applying quasi-random approaches to naturalistic driving data using Bayesian additive regression trees. *J. Surv. Stat. Methodol.* **8** 148–180. MR3858513 https://doi.org/10.4310/SII.2018.v11.n4.a1

[54] REILLY, C., GELMAN, A. and KATZ, J. (2001). Poststratification without population level information on the poststratifying variable with application to political polling. *J. Amer. Statist. Assoc.* **96** 1–11. MR1973779 https://doi.org/10.1198/016214501750332640

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

[56] RUBIN, D. B. (1983). Comment on "An evaluation of model-dependent and probability-sampling inferences in sample surveys," by M. H. Hansen, W. G. Madow and B. J. Tepping. *J. Amer. Statist. Assoc.* **78** 803–805.

[57] SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. *Springer Series in Statistics*. Springer, New York. MR1140409 https://doi.org/10.1007/978-1-4612-4378-6

[58] SI, Y. (2025). Supplement to "On the use of auxiliary variables in multilevel regression and poststratification." https://doi.org/10.1214/24-STS932SUPP

[59] SI, Y., COVELLO, L., WANG, S., COVELLO, T. and GELMAN, A. (2022). Beyond vaccination rates: A synthetic random proxy metric of total SARS-CoV-2 immunity seroprevalence in the community. *Epidemiology* **33** 457–464.

[60] SI, Y., PILLAI, N. S. and GELMAN, A. (2015). Bayesian nonparametric weighted sampling inference. *Bayesian Anal.* **10** 605–625. MR3420817 https://doi.org/10.1214/14-BA924

[61] SI, Y., TRANGUCCI, R., GABRY, J. S. and GELMAN, A. (2020). Bayesian hierarchical weighting adjustment and survey inference. *Surv. Methodol.* **46** 181–214.

[62] SI, Y. and ZHOU, P. (2021). Bayes-raking: Bayesian finite population inference with known margins. *J. Surv. Stat. Methodol.* **9** 833–855.

[63] SMITH, T. M. F. (1983). On the validity of inferences from nonrandom samples. *J. Roy. Statist. Soc. Ser. A* **146** 394–403. MR0769995 https://doi.org/10.2307/2981454

[64] STAN DEVELOPMENT TEAM (2020). RStan: The R interface to Stan. Available at https://CRAN.R-project.org/package=rstan.

[65] TAN, Z. (2007). Comment: Understanding OR, PS and DR. *Statist. Sci.* **22** 560–568. MR2420461 https://doi.org/10.1214/07-STS227A

[66] TANG, X., GHOSH, M., HA, N. S. and SEDRANSK, J. (2018). Modeling random effects using global-local shrinkage priors in small area estimation. *J. Amer. Statist. Assoc.* **113** 1476–1489. MR3902223 https://doi.org/10.1080/01621459.2017.1419135

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

[68] VALLIANT, R., DEVER, J. A. and KREUTER, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. *Statistics for Social and Behavioral Sciences*. Springer, New York. MR3088726 https://doi.org/10.1007/978-1-4614-6449-5

[69] VOLKOW, N. D., KOOB, G. F., CROYLE, R. T., BIANCHI, D. W., GORDON, J. A., KOROSHETZ, W. J., PÉREZ-STABLE, E. J., RILEY, W. T., BLOCH, M. H. et al. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **2018** 4–7.

[70] WANG, L., GRAUBARD, B. I., KATKI, H. A. and LI, Y. (2020). Improving external validity of epidemiologic cohort analyses: A kernel weighting approach. *J. Roy. Statist. Soc. Ser. A* **183** 1293–1311. MR4114487 https://doi.org/10.1111/rssa.12564

[71] WANG, W., ROTHSCHILD, D., GOEL, S. and GELMAN, A. (2015). Forecasting elections with non-representative polls. *Int. J. Forecast.* **31** 980–991.

[72] YANG, S. and KIM, J. K. (2020). Statistical data integration in survey sampling: A review. *Jpn. J. Stat. Data Sci.* **3** 625–650. MR4181993 https://doi.org/10.1007/s42081-020-00093-w

[73] YANG, S., KIM, J. K. and SONG, R. (2020). Doubly robust inference when combining probability and non-probability samples with high dimensional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 445–465. MR4084171

[74] YOUGOV (2020). 2020 presidential election model. Available at https://today.yougov.com/2020-presidential-election.

[75] ZAHORSKI, A. (2020). Multilevel regression with poststratification for the national level Viber/Street poll on the 2020 presidential election in Belarus. Available at https://arxiv.org/abs/2009.06615.

[76] ZANGENEH, S. Z. and LITTLE, R. J. A. (2015). Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *J. Surv. Stat. Methodol.* **3** 162–192.

[77] ZHANG, G. and LITTLE, R. (2009). Extensions of the penalized spline of propensity prediction method of imputation. *Biometrics* **65** 911–918. MR2649864 https://doi.org/10.1111/j.1541-0420.2008.01155.x

[78] ZHANG, X., HOLT, J. B., YUN, S., LU, H., GREENLUND, K. J. and CROFT, J. B. (2015). Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. *Amer. J. Epidemiol.* **182** 127–137. https://doi.org/10.1093/aje/kwv002

[79] ZHOU, H., ELLIOTT, M. R. and RAGHUNATHAN, T. E. (2016). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics* **72** 242–252. MR3500593 https://doi.org/10.1111/biom.12413

Statistical Science
2025, Vol. 40, No. 2, 289–312
https://doi.org/10.1214/24-STS938

Orthogonal and Linear Regressions and Pencils of Confocal Quadrics

Vladimir Dragović and Borislav Gajić

Abstract. This paper enhances and develops bridges between statistics, mechanics, and geometry. For a given system of points in $\mathbb{R}^k$ representing a sample of full rank, we construct an explicit pencil of confocal quadrics with the following properties: (i) All the hyperplanes for which the hyperplanar moments of inertia for the given system of points are equal, are tangent to the same quadrics from the pencil of quadrics. As an application, we develop regularization procedures for the orthogonal least square method, analogues of lasso and ridge methods from linear regression. (ii) For any given point P among all the hyperplanes that contain it, the best fit is the tangent hyperplane to the quadric from the confocal pencil corresponding to the maximal Jacobi coordinate of the point P; the worst fit among the hyperplanes containing P is the tangent hyperplane to the ellipsoid from the confocal pencil that contains P. The confocal pencil of quadrics provides a universal tool to solve the restricted principal component analysis restricted at any given point. Both results (i) and (ii) can be seen as generalizations of the classical result of Pearson on orthogonal regression. They have natural and important applications in the statistics of the errors-in-variables models (EIV). For the classical linear regressions we provide a geometric characterization of hyperplanes of least squares in a given direction among all hyperplanes which contain a given point. The obtained results have applications in restricted regressions, both ordinary and orthogonal ones. For the latter, a new formula for test statistic is derived. The developed methods and results are illustrated in natural statistics examples.

Key words and phrases: Data ellipsoid, confocal pencil of quadrics, planar moments of inertia, restricted regression, regularization and shrinkage, restricted PCA.

REFERENCES

[1] ADCOCK, R. J. (1877). Note on the method of least squares. *Analyst* **4** 183–184.
[2] ADCOCK, R. J. (1878). A problem in least squares. *Analyst* **5** 53–54.
[3] AKRITAS, M. and BERSHADY, M. (1996). Linear regression for astronomical data with measurement errors and intrinsic scatter. *Astrophys. J.* **470** 706–714.
[4] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. *Wiley Publications in Statistics*. Wiley, New York. MR0091588
[5] ARNOLD, V. I. (1989). *Mathematical Methods of Classical Mechanics*, 2nd ed. *Graduate Texts in Mathematics* **60**. Springer, New York. Translated from the Russian by K. Vogtmann and A. Weinstein. MR0997295 https://doi.org/10.1007/978-1-4757-2063-1
[6] ARVIDSSON, D., FRIDOLFSSON, J. and BÖRJESSON, M. (2019). Measurement of physical activity in clinical practice using accelerometers. *J. Intern. Med.* **286** 137–153. https://doi.org/10.1111/joim.12908
[7] BEN GILLEN, B., SNOWBERG, E. and YARIV, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *J. Polit. Econ.* **127**.
[8] CARROLL, R. J. and RUPPERT, D. (1996). The use and misuse of orthogonal regression estimation in linear errors-in-variables models. *Amer. Statist.* **50** 1–6.
[9] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models*: *A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138
[10] CASELLA, G. and BERGER, R. L. (2001). *Statistical Inference*, 2nd ed. *Duxbury Advanced Series*. Duxbury, N. Scituate.

Vladimir Dragović is Full Professor, Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas, USA (e-mail: Vladimir.Dragovic@utdallas.edu). Borislav Gajić is Research Professor, Mathematical Institute SANU, Belgrade, Serbia (e-mail: gajab@mi.sanu.ac.rs).

[11] COHEN, J. E. and D'EUSTACHIO, P. (1978). An affine linear model for the relation between two sets of frequency counts. Response to a query. *Biometrics* **34** 514–516.

[12] COHEN, J. E., D'EUSTACHIO, P. and EDELMAN, G. M. (1977). The specific antigen-binding cell populations of individual fetal mouse spleens: Repertoire composition, size, and genetic control. *J. Exp. Med.* **146** 394–411. https://doi.org/10.1084/jem.146.2.394

[13] COUREL-IBANEZ, J., MARTINEZ-CAVA, A., MORAN-NAVARRO, R. et al. (2019). Reproducibility and repeatability of five different technologies for bar velocity measurement in resistance training. *Ann. Biomed. Eng.* **47** 1523–1538.

[14] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. *Princeton Mathematical Series*, *Vol.* 9. Princeton Univ. Press, Princeton, NJ. MR0016588

[15] DEGROOT, M. H. and SCHERVISH, M. (2012). *Probability and Statistics*. 4th ed. Addison-Wesley, Reading.

[16] DEMPSTER, A. (1969). *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading.

[17] DRAGOVIĆ, V. and GAJIĆ, B. (2023). Points with rotational ellipsoids of inertia, envelopes of hyperplanes which equally fit the system of points in $\mathbb{R}^k$, and ellipsoidal billiards. *Phys. D* **451** Paper No. 133776, 15. MR4591364 https://doi.org/10.1016/j.physd.2023.133776

[18] DRAGOVIĆ, V. and GAJIĆ, B. (2024). Geometric Approach and Closed Exact Formulae for the Lasso. arxiv. Available at arXiv:2407.08058 [math:OC].

[19] DRAGOVIĆ, V. and GAJIĆ, B. (2024). Analogues of Lasso in orthogonal regression. Unpublished manuscript.

[20] DRAGOVIĆ, V. and GAJIĆ, B. (2025). Supplement to "Orthogonal and linear regressions and pencils of confocal quadrics." https://doi.org/10.1214/24-STS938SUPP

[21] DRAGOVIĆ, V. and RADNOVIĆ, M. (2008). Hyperelliptic Jacobians as billiard algebra of pencils of quadrics: Beyond Poncelet porisms. *Adv. Math.* **219** 1577–1607. MR2458147 https://doi.org/10.1016/j.aim.2008.06.021

[22] DRAGOVIĆ, V. and RADNOVIĆ, M. (2011). *Poncelet Porisms and Beyond*: *Integrable Billiards*, *Hyperelliptic Jacobians and Pencils of Quadrics*. *Frontiers in Mathematics*. Birkhäuser, Basel. MR2798784 https://doi.org/10.1007/978-3-0348-0015-0

[23] FOMBY, T. B., HILL, R. C. and JOHNSON, S. R. (1984). *Advanced Econometric Methods*. Springer, New York. MR0753284 https://doi.org/10.1007/978-1-4419-8746-4

[24] FORBES, J. D. (1857). Further experiments and remarks on the measurement of heights by the boiling point of water. *Trans. R. Soc. Edinb.* **21** 135–143.

[25] FRIENDLY, M., MONETTE, G. and FOX, J. (2013). Elliptic insights: Understanding statistical methods through elliptical geometry. *Statist. Sci.* **28** 1–39. MR3075337 https://doi.org/10.1214/12-STS402

[26] FULLER, W. A. (1978). An affine linear model for the relation between two sets of frequency counts: Response to query. *Biometrics* **34** 517–521.

[27] FULLER, W. A. (1987). *Measurement Error Models*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0898653 https://doi.org/10.1002/9780470316665

[28] GALTON, F. (1886). Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. G. B. Irel.* **15** 246–263.

[29] GARDINER, M. (1893). On "confocal quadrics of moments of inertia" pertaining to all planes in space, and loci and envelopes of straight lines whose "moments of inertia" are constants. *Proc. R. Soc. Vic.* **5** 200–208.

[30] GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

[31] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[32] JACOBI, C. G. J. (1865). *Vorlesungen über dynamik*, **51**. Springer, Berlin. English Translation: (2009) *Jacobi's Lectures on Dynamics*.

[33] JURDJEVIC, V. (2023). Integrable systems: In the footprints of the greats. *Mathematics* **11**. 1063.

[34] KELLY, B. (2007). Some aspects of measurement error in linear regression of astronomical data. *Astrophys. J.* **665** 1489–1506.

[35] MEIJER, E., OCZKOWSKI, E. and WANSBEEK, T. (2021). How measurement error affects inference in linear regression. *Empir. Econ.* **60** 131–155.

[36] PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2** 559–572.

[37] ROGERS, A. J. (2013). Concentration ellipsoids, their planes of support, and the linear regression model. *Econometric Rev.* **32** 220–243. MR2988926 https://doi.org/10.1080/07474938.2011.608055

[38] SEBER, G. A. F. and LEE, A. J. (2003). *Linear Regression Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, New York. MR1958247 https://doi.org/10.1002/9780471722199

[39] SUSLOV, G. K. (1900). *Fundamentals of Analytical Mechanics*, *Vol.* 1. Tipografiia Imperatatorskago universiteta Sv. Vladimiira N.T. Korchak-Novitskago, Kiev. [in Russian].

[40] WEISBERG, S. (1985). *Applied Linear Regression*, 2nd ed. Wiley, New York.

Statistical Science
2025, Vol. 40, No. 2, 313–328
https://doi.org/10.1214/24-STS934

Cox Reduction and Confidence Sets of Models: A Theoretical Elucidation

Rebecca M. Lewis and Heather S. Battey

Abstract. For sparse high-dimensional regression problems, Battey and Cox (*Proc. R. Soc. Lond. A* **474** (2018) 20170631) and Cox and Battey (*Proc. Natl. Acad. Sci. USA* **114** (2017) 8592–8595) emphasised the need for confidence sets of models: an enumeration of those small sets of variables that fit the data equivalently well in a suitable statistical sense. This is to be contrasted with the single model returned by penalised regression procedures, effective for prediction but potentially misleading for subject-matter understanding. The proposed construction of such sets relied on preliminary reduction of the full set of variables. While various possibilities could be considered for this, Cox and Battey (*Proc. Natl. Acad. Sci. USA* **114** (2017) 8592–8595) proposed a succession of regression fits based on incomplete block designs. The purpose of the present paper is to provide some new insight into that work. For an unspecified reduction strategy, we begin by characterising models that are likely to be retained in the model confidence set, emphasising geometric aspects. We then evaluate possible reduction schemes based on penalised regression or marginal screening, before theoretically elucidating the reduction of Cox and Battey (*Proc. Natl. Acad. Sci. USA* **114** (2017) 8592–8595). We identify features of the covariate matrix that may reduce its efficacy, and indicate improvements to the original proposal. An advantage of the approach is its ability to reveal its own stability or fragility for the data at hand.

Key words and phrases: Genomics, high-dimensional regression, model uncertainty, partially balanced incomplete block arrangements, sparsity.

REFERENCES

[1] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337
[2] BATTEY, H. S. and COX, D. R. (2018). Large numbers of explanatory variables: A probabilistic assessment. *Proc. R. Soc. Lond. A* **474** Paper No. 20170631, 9. MR3842843
[3] BIND, M. and RUBIN, D. B. (2023). High-dimensional randomization-based inference capitalizing on classical design and modern computing. *Behaviormetrika* **50** 9–29.
[4] BRAGG, F., KARTSONAKI, C. and GUO, Y. E. A. (2022). The role of NMR-based circulating metabolic biomarkers in development and risk prediction of new onset type 2 diabetes. *Sci. Rep.* **12** 15071.
[5] BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388
[6] BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. H. (2010). Variable selection in high-dimensional linear models: Partially faithful distributions and the PC-simple algorithm. *Biometrika* **97** 261–278. MR2650737 https://doi.org/10.1093/biomet/asq008
[7] BÜHLMANN, P., KALISCH, M. and MEIER, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annu. Rev. Stat. Appl.* **1** 255–278.
[8] CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265
[9] COCHRAN, W. G. (1938). The omission or addition of an independent variable in multiple linear regression. *Suppl. J. R. Stat. Soc.* **5** 171–176.
[10] COX, D. R. (1960). Regression analysis when there is prior information about supplementary variables. *J. Roy. Statist. Soc. Ser. B* **22** 172–176. MR0114270
[11] COX, D. R. (1968). Notes on some aspects of regression analysis (with discussion). *J. R. Stat. Soc., A* **131** 265–279.
[12] COX, D. R. (1992). Causality: Some statistical aspects. *J. Roy. Statist. Soc. Ser. A* **155** 291–301. MR1157712 https://doi.org/10.2307/2982962

Rebecca M. Lewis is a Ph.D. student, Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7, UK (e-mail: rebecca.lewis15@imperial.ac.uk). Heather S. Battey is a Reader in Statistics, Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7, UK (e-mail: h.battey@imperial.ac.uk).

[13] COX, D. R. and BATTEY, H. S. (2017). Large numbers of explanatory variables, a semi-descriptive analysis. *Proc. Natl. Acad. Sci. USA* **114** 8592–8595. MR3688013 https://doi.org/10.1073/pnas.1703764114

[14] COX, D. R. and DONNELLY, C. A. (2011). *Principles of Applied Statistics*. Cambridge Univ. Press, Cambridge. MR2817147 https://doi.org/10.1017/CBO9781139005036

[15] COX, D. R. and SNELL, E. J. (1974). The choice of variables in observational studies. *J. R. Stat. Soc., Ser. C* **23** 51–59.

[16] COX, D. R. and SNELL, E. J. (1989). *Analysis of Binary Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **32**. CRC Press, London. MR1014891

[17] COX, D. R. and WERMUTH, N. (1990). An approximation to maximum likelihood estimates in reduced models. *Biometrika* **77** 747–761. MR1086687 https://doi.org/10.1093/biomet/77.4.747

[18] COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. *Monographs on Statistics and Applied Probability* **67**. CRC Press, London. MR1456990

[19] COX, D. R. and WERMUTH, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 937–941. MR2017879 https://doi.org/10.1111/1467-9868.00424

[20] DICKER, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269–284. MR3215347 https://doi.org/10.1093/biomet/ast065

[21] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

[22] EVANS, R. J. (2020). Model selection and local geometry. *Ann. Statist.* **48** 3513–3544. MR4185818 https://doi.org/10.1214/19-AOS1940

[23] FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839 https://doi.org/10.1111/j.1467-9868.2011.01005.x

[24] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

[25] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

[26] FAN, J., SHAO, Q.-M. and ZHOU, W.-X. (2018). Are discoveries spurious? Distributions of maximum spurious correlations and their applications. *Ann. Statist.* **46** 989–1017. MR3797994 https://doi.org/10.1214/17-AOS1575

[27] FRIEDMAN, J., HASTIE, T. J. and TIBSHIRANI, R. J. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **43**. 1–22.

[28] GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.

[29] HANSEN, P. R., LUNDE, A. and NASON, J. M. (2011). The model confidence set. *Econometrica* **79** 453–497. MR2809377 https://doi.org/10.3982/ECTA5771

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

[31] HOELTGEBAUM, H. H. (2023). HCmodelSets: Regression with a large number of potential explanatory variables. R package version 1.1.3.

[32] HOELTGEBAUM, H. H. and BATTEY, H. S. (2019). HCmodelSets: An R package for specifying sets of well-fitting models in high dimensions. *R J.* **11** 370–379.

[33] KARTSONAKI, C., PANG, Y., MILLWOOD, I., YANG, L., GUO, Y., WALTERS, R., LV, J., HILL, M., YU, C. et al. (2022). Circulating proteins and risk of pancreatic cancer: A case-subcohort study among Chinese adults. *Int. J. Epidemiol.* **51** 817–829.

[34] LE CAM, L. (1960). *Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses. Univ. California Publ. Statist.* **3**. 37–98. MR0126903

[35] LEWIS, R. and BATTEY, H. S. (2023). On inference in high-dimensional logistic regression models with separated data. *Biometrika*.

[36] LEWIS, R. M. and BATTEY, H. S. (2025). Supplement to "Cox reduction and confidence sets of models: A theoretical elucidation." https://doi.org/10.1214/24-STS934SUPP

[37] LI, S., SESIA, M., ROMANO, Y., CANDÈS, E. and SABATTI, C. (2022). Searching for robust associations with a multi-environment knockoff filter. *Biometrika* **109** 611–629. MR4472838 https://doi.org/10.1093/biomet/asab055

[38] MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1032. MR0997578

[39] P., Y., KARTSONAKI, C., LV, J., MILLWOOD, I. Y., FAIRHURST-HUNTER, Z., TURNBULL, I., BRAGG, F., HILL, M. R., YU, C., GUO, Y. et al. (2022). Adiposity, metabolomic biomarkers, and risk of nonalcoholic fatty liver disease: A case-cohort study. *Am. J. Clin. Nutr.* **115** 799–810.

[40] REID, S., TIBSHIRANI, R. and FRIEDMAN, J. (2016). A study of error variance estimation in Lasso regression. *Statist. Sinica* **26** 35–67. MR3468344

[41] SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York. MR0116429

[42] STÄDLER, N., BÜHLMANN, P. and VAN DE GEER, S. (2010). ℓ_1-penalization for mixture regression models. *TEST* **19** 209–256. MR2677722 https://doi.org/10.1007/s11749-010-0197-z

[43] SU, W., BOGDAN, M. and CANDÈS, E. (2017). False discoveries occur early on the Lasso path. *Ann. Statist.* **45** 2133–2150. MR3718164 https://doi.org/10.1214/16-AOS1521

[44] SU, W. J. (2018). When is the first spurious variable selected by sequential regression procedures? *Biometrika* **105** 517–527. MR3842882 https://doi.org/10.1093/biomet/asy032

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

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

[47] YATES, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.* **26** 424–455.

[48] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Statistical Science
2025, Vol. 40, No. 2, 329–345
https://doi.org/10.1214/25-STS992

A Conversation with Vijay Nair

Elizaveta Levina and Naveen N. Narisetty

Abstract. Vijayan N. Nair (Vijay) was born in Malaysia. He received his undergraduate degree in Economics from the University of Malaya and his Ph.D. in Statistics from the University of California, Berkeley. He spent 15 years as a Research Scientist at Bell Labs in New Jersey before turning to academic life at the University of Michigan in 1993. He was Donald A. Darling Collegiate Professor of Statistics and Professor of Industrial and Operations Engineering at Michigan, and he also served as department chair of Statistics for over 10 years and cofounded the Michigan Institute for Data Science. After retiring from Michigan in 2016, he moved to Wells Fargo, where he is currently Head of a Research and Development Group.

Vijay has broad research interests and has contributed to industrial and engineering statistics, reliability, design and analysis of experiments, network tomography and behavioral interventions. His current research interests include risk modeling, machine learning and generative AI. He is a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, the American Society for Quality and the American Association for the Advancement of Science. He delivered the Deming Lecture at the Joint Statistical Meetings in 2013 and the Gosset Lecture at the World Statistics Congress in 2017. His extensive record of service to the profession includes serving as Coeditor-in-Chief of the International Statistical Review for two terms, Editor of Technometrics, President of the International Statistical Institute and President of the International Society for Business and Industrial Statistics.

Key words and phrases: Reliability, industrial statistics, risk modeling, machine learning, generative AI.

REFERENCES

BICKEL, P. J. and DOKSUM, K. A. (1976). *Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day Series in Probability and Statistics*. Holden-Day, Inc., San Francisco, CA–Düsseldorf–Johannesburg. MR0443141

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York–Chichester–Brisbane. MR0483116

COLLINS, L. M., MURPHY, S. A., NAIR, V. N. and STRECHER, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Ann. Behav. Med.* **30** 65–73. https://doi.org/10.1207/s15324796abm3001_8

FRASER, D. A. S. (1958). *Statistics: An Introduction. A Wiley Publication in Mathematical Statistics*. Wiley, New York; CRC Press, London. MR0101576

FRIEDMAN, D. J., HANSEN, M. H., HOYER, J. R. and NAIR, V. N. (1993). U.S. Patent: “Method for characterizing failed circuits on semiconductor wafers”.

FRIEDMAN, D. J., HANSEN, M. H., NAIR, V. N. and JAMES, D. (1997). Model-free estimation of defect clustering in integrated circuit fabrication. *IEEE Trans. Semicond. Manuf.* **10** 344–359.

GEORGESCU, J. (2022). Bring back the golden days of Bell Labs. *Nat. Rev. Phys.* **4** 76–78.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations. A Wiley Publication in Applied Statistics*. Wiley, New York. MR0440802

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York; CRC Press, London. MR0045368

LAWRENCE, E., MICHAILIDIS, G. and NAIR, V. N. (2006). Network delay tomography using flexicast experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 785–813. MR2301295 https://doi.org/10.1111/j.1467-9868.2006.00567.x

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York; CRC Press, London. MR0107933

MALLOWS, C. L., NAIR, V. N., SHEPP, L. A. and VARDI, Y. (1985). Optimal sequential selection of secretaries. *Math. Oper. Res.* **10** 709–715. MR0812828 https://doi.org/10.1287/moor.10.4.709

Elizaveta (Liza) Levina is the Vijay Nair Collegiate Professor of Statistics and Department Chair, Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: elevina@umich.edu). Naveen N. Narisetty obtained his Ph.D. in Statistics (under the cosupervision of Vijay Nair) from University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: naveennn@umich.edu).

NAIR, V. N. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika* **68** 99–103. MR0614946 https://doi.org/10.1093/biomet/68.1.99

NAIR, V. N. (1986). Testing in industrial experiments with ordered categorical data (with discussion) (Wilcoxon prize for best applications paper). *Technometrics* **28** 283–311.

NAIR, V. N. (1992). Taguchi's parameter design: A panel discussion. *Technometrics* 127–161.

NAIR, V. N., HANSEN, M. and SHI, J. (2000). Statistics in advanced manufacturing. *J. Amer. Statist. Assoc.* 1002–1005.

NAIR, V. N. and PREGIBON, D. (1993). Discussion of 'Is industrial statistics out of control' by D. Banks. *Statist. Sci.* 356–409.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York; CRC Press, London. MR0116429

XI, B., MICHAILIDIS, G. and NAIR, V. N. (2006). Estimating network loss rates using active tomography. *J. Amer. Statist. Assoc.* **101** 1430–1448. MR2279470 https://doi.org/10.1198/016214506000000366

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

IMS EDITORS