STATISTICAL SCIENCE

Volume 33, Number 3 August 2018

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 33, Number 3, August 2018. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

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

Statistical Science **Volume 33, Number 3 (299–467) August 2018**

Volume 33 Number 3 August 2018

Statistical Science
2018, Vol. 33, No. 3, 299–318
https://doi.org/10.1214/17-STS629

A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications

Alex Reinhart

Abstract. Self-exciting spatio-temporal point process models predict the rate of events as a function of space, time, and the previous history of events. These models naturally capture triggering and clustering behavior, and have been widely used in fields where spatio-temporal clustering of events is observed, such as earthquake modeling, infectious disease, and crime. In the past several decades, advances have been made in estimation, inference, simulation, and diagnostic tools for self-exciting point process models. In this review, I describe the basic theory, survey related estimation and inference techniques from each field, highlight several key applications, and suggest directions for future research.

Key words and phrases: Epidemic-Type Aftershock Sequence, conditional intensity, Hawkes process, stochastic declustering.

REFERENCES

ADELFIO, G. and CHIODI, M. (2015a). Alternated estimation in semi-parametric space–time branching-type point processes with application to seismic catalogs. *Stoch. Environ. Res. Risk Assess.* **29** 443–450.

ADELFIO, G. and CHIODI, M. (2015b). FLP estimation of semi-parametric models for space–time point processes and diagnostic tools. *Spat. Stat.* **14** 119–132. MR3429716

BACRY, E., MASTROMATTEO, I. and MUZY, J.-F. (2015). Hawkes processes in finance. *Mark. Microstruct. Liq.* **1** 1550005.

BADDELEY, A., MØLLER, J. and PAKES, A. G. (2007). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* **60** 627–649.

BADDELEY, A., RUBAK, E. and MØLLER, J. (2011). Score, pseudo-score and residual diagnostics for spatial point process models. *Statist. Sci.* **26** 613–646.

BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 617–666. MR2210685

BAUWENS, L. and HAUTSCH, N. (2009). Modelling financial high frequency data using point processes. In *Handbook of Financial Time Series* (T. Mikosch, J.-P. Kreiß, R. A. Davis and T. G. Andersen, eds.) 953–979. Springer, Berlin.

BERNASCO, W., JOHNSON, S. D. and RUITER, S. (2015). Learning where to offend: Effects of past on future burglary locations. *Appl. Geogr.* **60** 120–129.

BRAGA, A. A., PAPACHRISTOS, A. V. and HUREAU, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Q.* **31** 633–663.

BRAY, A. and SCHOENBERG, F. P. (2013). Assessment of point process models for earthquake forecasting. *Statist. Sci.* **28** 510–520. MR3161585

BRAY, A., WONG, K., BARR, C. D. and SCHOENBERG, F. P. (2014). Voronoi residual analysis of spatial point process models with applications to California earthquake forecasts. *Ann. Appl. Stat.* **8** 2247–2267. MR3292496

BROCKMANN, D., HUFNAGEL, L. and GEISEL, T. (2006). The scaling laws of human travel. *Nature* **439** 462–465.

CHEN, J. M., HAWKES, A. G., SCALAS, E. and TRINH, M. (2017). Performance of information criteria used for model selection of Hawkes process models of financial data. Available at arXiv:1702.06055.

CHIODI, M. and ADELFIO, G. (2011). Forward likelihood-based predictive approach for space–time point processes. *Environmetrics* **22** 749–757. MR2843141

CLARKE, R. V. and CORNISH, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime Justice* **6** 147–185.

CLEMENTS, R. A., SCHOENBERG, F. P. and VEEN, A. (2012). Evaluation of space–time point process models using super-thinning. *Environmetrics* **23** 606–616.

COHEN, L. E. and FELSON, M. (1979). Social change and crime rate trends: A routine activity approach. *Am. Sociol. Rev.* **44** 588–608.

COHEN, J., GORR, W. L. and OLLIGSCHLAEGER, A. M. (2007). Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geogr. Anal.* **39** 105–127.

Alex Reinhart is a Ph.D. Student, Department of Statistics & Data Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, Pennsylvania 15213, USA (e-mail: areinhar@stat.cmu.edu).

COWLING, A. and HALL, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **58** 551–563. MR1394366

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, New York.

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

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

DIGGLE, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. *Monographs on Statistics and Applied Probability* **128**. CRC Press, Boca Raton, FL. MR3113855

DIGGLE, P. J., MORAGA, P., ROWLINGSON, B. and TAYLOR, B. M. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statist. Sci.* **28** 542–563.

FOTHERINGHAM, A. S. and WONG, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plann. A.* **23** 1025–1044.

FOX, E. W., SCHOENBERG, F. P. and GORDON, J. S. (2016). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *Ann. Appl. Stat.* **10** 1725–1756. MR3553242

FOX, E. W., SHORT, M. B., SCHOENBERG, F. P., CORONGES, K. D. and BERTOZZI, A. L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *J. Amer. Statist. Assoc.* **111** 564–584.

FREED, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annu. Rev. Earth Planet. Sci.* **33** 335–367.

GONZÁLEZ, J. A., RODRÍGUEZ-CORTÉS, F. J., CRONIE, O. and MATEU, J. (2016). Spatio-temporal point process statistics: A review. *Spat. Stat.* **18** 505–544.

GORR, W. L. and LEE, Y. (2015). Early warning system for temporary crime hot spots. *J. Quant. Criminol.* **31** 25–47.

GRAY, A. G. and MOORE, A. W. (2003). Nonparametric density estimation: Toward computational tractability. In *SIAM International Conference on Data Mining* 203–211. SIAM, Philadelphia, PA.

GREEN, B., HOREL, T. and PAPACHRISTOS, A. V. (2017). Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Intern. Med.* **177** 326–333.

HARTE, D. S. (2012). Bias in fitting the ETAS model: A case study based on New Zealand seismicity. *Geophys. J. Int.* **192** 390–412.

HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **51** 83–90.

HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503.

JOHNSON, D. H. (1996). Point process models of single-neuron discharges. *J. Comput. Neurosci.* **3** 275–299.

KUMAZAWA, T. and OGATA, Y. (2014). Nonstationary ETAS models for nonstandard earthquakes. *Ann. Appl. Stat.* **8** 1825–1852. MR3271355

LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413.

LEWIS, E., MOHLER, G., BRANTINGHAM, P. J. and BERTOZZI, A. L. (2011). Self-exciting point process models of civilian deaths in Iraq. *Secur. J.* **25** 244–264.

LIPPIELLO, E., GIACCO, F., DE ARCANGELIS, L., MARZOCCHI, W. and GODANO, C. (2014). Parameter estimation in the ETAS model: Approximations and novel methods. *Bull. Seismol. Soc. Am.* **104** 985–994.

LOEFFLER, C. and FLAXMAN, S. (2016). Is gun violence contagious? Available at arXiv:1611.06713.

MARSAN, D. and LENGLINÉ, O. (2008). Extending earthquakes' reach through cascading. *Science* **319** 1076–1079.

MARSAN, D. and LENGLINÉ, O. (2010). A new estimation of the decay of aftershock density with distance to the mainshock. *J. Geophys. Res.* **115** B09302.

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. Wiley, New York.

MEYER, S. (2010). Spatio-temporal infectious disease epidemiology based on point processes. Master's thesis, Ludwig-Maximilians-Univ. München.

MEYER, S., ELIAS, J. and HÖHLE, M. (2012). A space–time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68** 607–616.

MEYER, S. and HELD, L. (2014). Power-law models for infectious disease spread. *Ann. Appl. Stat.* **8** 1612–1639. MR3271346

MEYER, S., WARNKE, I., RÖSSLER, W. and HELD, L. (2016). Model-based testing for space–time interaction using point processes: An application to psychiatric hospital admissions in an urban area. *Spat. Spatio-temporal Epidemiol.* **17** 15–25.

MOHLER, G. O. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497.

MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108.

MØLLER, J. and RASMUSSEN, J. G. (2005). Perfect simulation of Hawkes processes. *Adv. in Appl. Probab.* **37** 629–646. MR2156552

MUSMECI, F. and VERE-JONES, D. (1992). A space–time clustering model for historical earthquakes. *Ann. Inst. Statist. Math.* **44** 1–11.

NANDAN, S., OUILLON, G., WIEMER, S. and SORNETTE, D. (2017). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to California. *J. Geophys. Res. Solid Earth* **122** 5118–5143.

NSOESIE, E. O., BROWNSTEIN, J. S., RAMAKRISHNAN, N. and MARATHE, M. V. (2013). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza Other Respir. Viruses* **8** 309–316.

OGATA, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30** 243–261. MR0514494

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

OGATA, Y. (1999). Seismicity analysis through point-process modeling: A review. *Pure Appl. Geophys.* **155** 471–507.

OGATA, Y. and KATSURA, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. *Ann. Inst. Statist. Math.* **40** 29–39.

OGATA, Y., KATSURA, K. and TANEMURA, M. (2003). Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **52** 499–509.

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

PENG, R. D., SCHOENBERG, F. P. and WOODS, J. A. (2005). A space–time conditional intensity model for evaluating a wildfire hazard index. *J. Amer. Statist. Assoc.* **100** 26–35.

PORTER, M. D. and WHITE, G. (2012). Self-exciting hurdle models for terrorist activity. *Ann. Appl. Stat.* **6** 106–124. MR2951531

RASMUSSEN, J. G. (2013). Bayesian inference for Hawkes processes. *Methodol. Comput. Appl. Probab.* **15** 623–642. MR3085883

RATCLIFFE, J. H. and RENGERT, G. F. (2008). Near-repeat patterns in Philadelphia shootings. *Secur. J.* **21** 58–76.

RATHBUN, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *J. Statist. Plann. Inference* **51** 55–74.

RIPLEY, B. D. (1977). Modelling spatial patterns. *J. Roy. Statist. Soc. Ser. B* **39** 172–212.

ROSS, G. J. (2016). Bayesian estimation of the ETAS model for earthquake occurrences. Preprint.

SARMA, S. V., NGUYEN, D. P., CZANNER, G., WIRTH, S., WILSON, M. A., SUZUKI, W. and BROWN, E. N. (2011). Computing confidence intervals for point process models. *Neural Comput.* **23** 2731–2745.

SCHOENBERG, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *J. Amer. Statist. Assoc.* **98** 789–795. MR2055487

SCHOENBERG, F. P. (2013). Facilitated estimation of ETAS. *Bull. Seismol. Soc. Am.* **103** 601–605.

SCHOENBERG, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statist. Sinica* **26** 861–879. MR3497774

SCHOENBERG, F. P., HOFFMAN, M. and HARRIGAN, R. (2017). A recursive point process model for infectious diseases. Available at arXiv:1703.08202.

SHORT, M. B., D'ORSOGNA, M. R., BRANTINGHAM, P. J. and TITA, G. E. (2009). Measuring and modeling repeat and near-repeat burglary effects. *J. Quant. Criminol.* **25** 325–339.

SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, Boca Raton, FL.

STAN DEVELOPMENT TEAM (2016). Stan modeling language users guide and reference manual. Available at http://mc-stan.org.

TOWNSLEY, M., HOMEL, R. and CHASELING, J. (2003). Infectious burglaries: A test of the near repeat hypothesis. *Br. J. Criminol.* **43** 615–633.

VEEN, A. and SCHOENBERG, F. P. (2008). Estimation of space–time branching process models in seismology using an EM-type algorithm. *J. Amer. Statist. Assoc.* **103** 614–624.

VERE-JONES, D. (2009). Some models and procedures for space–time point processes. *Environ. Ecol. Stat.* **16** 173–195.

WANG, Q., SCHOENBERG, F. P. and JACKSON, D. D. (2010). Standard errors of parameter estimates in the ETAS model. *Bull. Seismol. Soc. Am.* **100** 1989–2001.

WEISBURD, D. (2015). The law of crime concentration and the criminology of place. *Criminology* **53** 133–157.

ZHUANG, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. Roy. Statist. Soc. Ser. B* **68** 635–653.

ZHUANG, J. (2011). Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth Planets Space* **63** 207–216.

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc.* **97** 369–380.

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.* **109** B05301.

ZIPKIN, J. R., SCHOENBERG, F. P., CORONGES, K. and BERTOZZI, A. L. (2016). Point-process models of social network interactions: Parameter estimation and missing data recovery. *European J. Appl. Math.* **27** 502–529. MR3491509

Statistical Science
2018, Vol. 33, No. 3, 319–322
https://doi.org/10.1214/18-STS650

Comment on "A Review of Self-Exciting Spatiotemporal Point Process and Their Applications" by Alex Reinhart

Yosihiko Ogata

Abstract. In my discussion, I would like to comment on our early reactions to Hawkes' enlightening paper on the self-exciting model; further, I would like to comment on developments of the extended models with some applications.

Key words and phrases: Akaike Bayesian Information Criterion (ABIC), Akaike information criterion (AIC), causality analysis, conditional intensity function, empirical Bayesian method, epidemic-type aftershock sequence (ETAS) model, hierarchical space-time ETAS (HIST-ETAS) model, maximum-likelihood method, penalized log-likelihood, statistical seismology, study of earthquake predictability, thinning simulation method.

REFERENCES

ADAMOPOULOS, L. (1976). Cluster models for earthquakes: Regional comparisons. *Math. Geol.* **8** 463–475.

AKAIKE, H. (1980). Likelihood and the Bayes procedure. In *Bayesian Statistics* (*Valencia*, 1979) 143–166. Univ. Press, Valencia. MR0638876

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen & Co., Ltd., London. MR0199942

FLETCHER, R. and POWELL, M. J. D. (1963/1964). A rapidly convergent descent method for minimization. *Comput. J.* **6** 163–168. MR0152116

GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 255–277. MR0319314

HAINZL, S. and OGATA, Y. (2005). Detecting fluid signals in seismicity data through statistical earthquake modeling. *J. Geophys. Res.* **110** B05S07. DOI:10.1029/2004JB003247.

HAN, P., ZHUANG, J., HATTORI, K. and OGATA, Y. (2016). An interdisciplinary approach for earthquake modelling and forecasting. In 2016 *Fall Meeting of the American Geophysical Union* (*AGU*), Moscone Center, San Francisco, CA, 13 December 2016 (Oral).

HAWKES, A. G. and ADAMOPOULOS, L. (1973). Cluster models for earthquakes—Regional comparisons. *Bull. Int. Stat. Inst.* **45** 454–461.

OGATA, Y. (1999). Seismicity analyses through point-process modelling: A review. In *Seismicity Patterns*, *Their Statistical Significance and Physical Meaning* (M. Wyss, K. Shimazaki and A. Ito, eds.). *Pure and Applied Geophysics* **155** 471–507. Birkhauser, Basel.

JORDAN, T. H., CHEN, Y. T., GASPARINI, P., MADARIAGA, R. and MARZOCCHI, W. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Ann. Geophys.* **54** 315–391.

KUMAZAWA, T. and OGATA, Y. (2014). Nonstationary ETAS models for nonstandard earthquakes. *Ann. Appl. Stat.* **8** 1825–1852. MR3271355

KUMAZAWA, T., OGATA, Y., KIMURA, K., MAEDA, K. and KOBAYASHI, A. (2016). Background rates of swarm earthquakes that are synchronized with volumetric strain changes. *Earth Planet. Sci. Lett.* **442** 51–60. DOI:10.1016/j.epsl.2016.02.049.

LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413. MR0546120

NISHIZAWA, O., LEI, X. and NAGAO, T. (1994). Hazard function analysis of seismo-electric signals in Greece. In *Electromagnetic Phenomena Related to Earthquake Prediction* (M. Hayakawa and Y. Fujinawa, eds.) 459–474. Terra Publishing Company, Tokyo.

OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* **27** 23–31.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

Yosihiko Ogata is Emeritus Professor, The Institute of Statistical Mathematics, Midori-Cho 10-3, Tachikawa city, Tokyo, Japan 190-8562 (e-mail: ogata@ism.ac.jp).

OGATA, Y. (2004). Space-time model for regional seismicity and detection of crustal stress changes. *J. Geophys. Res.* **109** B03308.

OGATA, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth Planets Space* **63** 217–229.

OGATA, Y. (2013). A prospect of earthquake prediction research. *Statist. Sci.* **28** 521–541. MR3161586

OGATA, Y. (2017). Statistics of earthquake activity: Models and methods for earthquake predictability studies. *Annu. Rev. Earth Planet. Sci.* **45** 497–527. DOI:10.1146/annurev-earth-063016-015918.

OGATA, Y. and AKAIKE, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. Roy. Statist. Soc. Ser. B* **44** 102–107. MR0655379

OGATA, Y., AKAIKE, H. and KATSURA, K. (1982). The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process. *Ann. Inst. Statist. Math.* **34** 373–387.

OGATA, Y. and KATSURA, K. (1986). Point-process models with linearly parametrized intensity for application to earthquake data. *J. Appl. Probab.* **23A** 291–310. MR0803179

OGATA, Y., MATSU'URA, R. S. and KATSURA, K. (1993). Fast likelihood computation of epidemic type aftershock-sequence model. *Geophys. Res. Lett.* **20** 2143–2146.

OGATA, Y., KATSURA, K., FALCONE, G., NANJO, K. Z. and ZHUANG, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bull. Seismol. Soc. Am.* **103** 1692–1708.

OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. MR0541960

PARZEN, E., TANABE, K. and KITAGAWA, G. (eds.) (1998). *Selected Papers of Hirotugu Akaike*. Springer, New York. MR1486823

RUBIN, I. (1972). Regular point processes and their detection. *IEEE Trans. Inform. Theory* **18** 547–557. MR0378096

UTSU, T., OGATA, Y. and MATSU'URA, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth* **43** 1–33.

VERE-JONES, D. (1970). Stochastic models for earthquake occurrence. *J. Roy. Statist. Soc. Ser. B* **32** 1–62. MR0272087

VERE-JONES, D. and OZAKI, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34** 189–207.

WANG, T., ZHUANG, J., KATO, T. and BEBBINGTON, M. (2013). Assessing the potential improvement in short-term earthquake forecasts from incorporation of GPS data. *Geophys. Res. Lett.* **40** 2631–2635.

WHITTLE, P. (1951). *Hypothesis Testing in Times Series Analysis*. Almqvist & Wiksells Boktryckeri AB, Uppsala.

ZHUANG, J., VERE-JONES, D., GUAN, H., OGATA, Y. and MA, L. (2005). Preliminary analysis of observations on the ultra-low frequency electric field in a region around Beijing. *Pure Appl. Geophys.* **162** 1367–1396. DOI:10.1007/s00024-004-2674-3.

ZHUANG, J., OGATA, Y., VERE-JONES, D., MA, L. and GUAN, H. (2014). Statistical modeling of earthquake occurrences based on external geophysical observations: With an illustrative application to the ultra-low frequency ground electric signals observed in the Beijing region. In *Seismic Imaging, Fault Rock Damage and Healing* (Y. Li, ed.) 351–376. De Gruyter, Berlin together with Higher Education Press, Beijing.

Statistical Science
2018, Vol. 33, No. 3, 323–324
https://doi.org/10.1214/18-STS651

Comment on "A Review of Self-Exciting Spatio-Temporal Point Process and Their Applications" by Alex Reinhart

Jiancang Zhuang

REFERENCES

GUO, Y., ZHUANG, J. and ZHOU, S. (2015). An improved space–time ETAS model for inverting the rupture geometry from seismicity triggering. *J. Geophys. Res.* **120** 3309–3323.

GUO, Y., ZHUANG, J., HIRATA, N. and ZHOU, S. (2017). Heterogeneity of direct aftershock productivity of the main shock rupture. *J. Geophys. Res.* **122** 5288–5305.

ZHUANG, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 635–653. MR2301012

ZHUANG, J. and MATEU, J. (2018). A semi-parametric spatiotemporal Hawkes-type point process model with periodic background for crime data. Submitted.

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.* **109** B05301.

Jiancang Zhuang is Associate Professor, Institute of Statistical Mathematics, Research Organization of Information and Systems, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan, (e-mail: zhuangjc@ism.ac.jp).

Statistical Science
2018, Vol. 33, No. 3, 325–326
https://doi.org/10.1214/18-STS652

Comment on "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications" by Alex Reinhart

Frederic Paik Schoenberg

REFERENCES

ADELFIO, G. and SCHOENBERG, F. P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Ann. Inst. Statist. Math.* **61** 929–948. MR2556772

BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 617–666. MR2210685

CRONIE, O. and VAN LIESHOUT, M. N. M. (2016). Bandwidth selection for kernel estimators of the spatial intensity function. Available at arXiv:1611.10221.

DIGGLE, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. *Monographs on Statistics and Applied Probability* **128**. CRC Press, Boca Raton, FL. MR3113855

HARTE, D. S. (2013). Bias in fitting the ETAS model: A case study based on New Zealand seismicity. *Geophys. J. Int.* **192** 390–412.

SCHOENBERG, F. P. (2013). Facilitated estimation of ETAS. *Bull. Seismol. Soc. Am.* **103** 601–605.

VEEN, A. and SCHOENBERG, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *J. Amer. Statist. Assoc.* **103** 614–624. MR2523998

Frederic Paik Schoenberg is Professor of Statistics, University of California, Los Angeles, 8152 MS building, Los Angeles, California 90095, USA. (e-mail: frederic@stat.ucla.edu).

Statistical Science
2018, Vol. 33, No. 3, 327–329
https://doi.org/10.1214/18-STS653

Self-Exciting Point Processes: Infections and Implementations

Sebastian Meyer

Abstract. This is a contribution to the discussion of Reinhart's "Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications" [*Statist. Sci.* **33** (2018)], which synthesizes developments from various research fields. Here, I discuss some experiences from modeling the spread of infectious diseases. Furthermore, I try to complement the review with regard to the availability of software for the described models, which I think is essential in "paving the way for new uses."

Key words and phrases: Spatio-temporal modeling, infectious disease epidemiology, statistical software.

REFERENCES

ADELFIO, G. and CHIODI, M. (2015). FLP estimation of semi-parametric models for space–time point processes and diagnostic tools. *Spat. Stat.* **14** 119–132. MR3429716

ALDRIN, M., HUSEBY, R. B. and JANSEN, P. A. (2015). Space–time modelling of the spread of pancreas disease (PD) within and between Norwegian marine salmonid farms. *Preventive Veterinary Medicine* **121** 132–141.

BADDELEY, A. and TURNER, R. (2005). `spatstat`: An R package for analyzing spatial point patterns. *J. Stat. Softw.* **12** 1–42.

BROCKMANN, D., HUFNAGEL, L. and GEISEL, T. (2006). The scaling laws of human travel. *Nature* **439** 462–465.

DIGGLE, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat. Methods Med. Res.* **15** 325–336. MR2242245

GIBBONS, C. L., MANGEN, M.-J. J., PLASS, D., HAVELAAR, A. H., BROOKE, R. J., KRAMARZ, P., PETERSON, K. L., STUURMAN, A. L., CASSINI, A., FÈVRE, E. M. and KRETZSCHMAR, M. E. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: A comparison of methods. *BMC Public Health* **14** 1–17.

HARTE, D. (2010). `PtProcess`: An R package for modelling marked point processes indexed by time. *J. Stat. Softw.* **35** 1–32.

HELD, L. and MEYER, S. (2018). Forecasting based on surveillance data. In *Handbook of Infectious Disease Data Analysis* (L. Held, N. Hens, P. D. O'Neill and J. Wallinga, eds.) Chapman & Hall/CRC, Boca Raton, FL. To appear.

HÖHLE, M. (2009). Additive-multiplicative regression models for spatio-temporal epidemics. *Biom. J.* **51** 961–978. MR2744450

HÖHLE, M. (2016). Infectious disease modelling. In *Handbook of Spatial Epidemiology* (A. B. Lawson, S. Banerjee, R. P. Haining and M. D. Ugarte, eds.) 477–500. CRC Press, Boca Raton, FL. MR3586838

JALILIAN, A. (2018). `ETAS`: Modeling earthquake data using ETAS model. R package version 0.4.4, Comprehensive R Archive Network.

KASAHARA, A., YAGI, Y. and ENESCU, B. (2016). `etas_solve`: A robust program to estimate the ETAS model parameters. *Seismological Research Letters* **87** 1143.

LOMBARDI, A. M. (2017). SEDA: A software package for the statistical earthquake data analysis. *Sci. Rep.* **7** 44171.

MEYER, S., ELIAS, J. and HÖHLE, M. (2012). A space–time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68** 607–616. MR2959628

MEYER, S. and HELD, L. (2014). Power-law models for infectious disease spread. *Ann. Appl. Stat.* **8** 1612–1639. MR3271346

MEYER, S., HELD, L. and HÖHLE, M. (2017). Spatio-temporal analysis of epidemic phenomena using the R package `surveillance`. *J. Stat. Softw.* **77** 1–55.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

READ, J. M., LESSLER, J., RILEY, S., WANG, S., TAN, L. J., KWOK, K. O., GUAN, Y., JIANG, C. Q. and CUMMINGS, D. A. T. (2014). Social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. Lond., B Biol. Sci.* **281**.

ROSS, G. J. (2017). `bayesianETAS`: Bayesian estimation of the ETAS model for earthquake occurrences. R package version 1.0.3, Comprehensive R Archive Network.

Sebastian Meyer is a Research Fellow at the Institute of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany (e-mail: seb.meyer@fau.de).

SCHEEL, I., ALDRIN, M., FRIGESSI, A. and JANSEN, P. A. (2007). A stochastic model for infectious salmon anemia (ISA) in Atlantic salmon farming. *J. R. Soc. Interface* **4** 699–706.

SCHRÖDLE, B., HELD, L. and RUE, H. (2012). Assessing the impact of a movement network on the spatiotemporal spread of infectious diseases. *Biometrics* **68** 736–744. MR3055178

TAYLOR, B., DAVIES, T., ROWLINGSON, B. and DIGGLE, P. (2013). Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate log-Gaussian Cox processes in R. *J. Stat. Softw.* **63** 1–48.

THE INSTITUTE OF STATISTICAL MATHEMATICS (2016). SAPP: Statistical analysis of point processes. R package version 1.0.7, Comprehensive R Archive Network.

Statistical Science
2018, Vol. 33, No. 3, 330–333
https://doi.org/10.1214/18-STS654

Rejoinder: A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications

Alex Reinhart

REFERENCES

ACHAB, M., BACRY, E., GAÏFFAS, S., MASTROMATTEO, I. and MUZY, J.-F. (2017). Uncovering causality from multivariate Hawkes integrated cumulants. In *Proceedings of the* 34*th International Conference on Machine Learning* **70**.

CHEN, S., WITTEN, D. and SHOJAIE, A. (2017). Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process. *Electron. J. Stat.* **11** 1207–1234. MR3634334

EICHLER, M., DAHLHAUS, R. and DUECK, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *J. Time Series Anal.* **38** 225–242. MR3611742

MOHLER, G. O. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497. DOI:10.1016/j.ijforecast.2014.01.004.

OATES, C. J. (2015). Accelerated non-parametrics for cascades of Poisson processes. *Stat* **4** 183–195. MR3405400

OGATA, Y., MATSU'URA, R. S. and KATSURA, K. (1993). Fast likelihood computation of epidemic type aftershock-sequence model. *Geophys. Res. Lett.* **20** 2143–2146. DOI:10.1029/93gl02142.

XU, H., FARAJTABAR, M. and ZHA, H. (2016). Learning granger causality for Hawkes processes. In *Proceedings of the* 33*rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 1717–1726. PMLR, New York.

ZHUANG, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 635–653. MR2301012

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.* **109** B05301. DOI:10.1029/2003JB002879.

Alex Reinhart is PhD candidate, Department of Statistics & Data Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, Pennsylvania 15213, USA (e-mail: areinhar@stat.cmu.edu).

Statistical Science
2018, Vol. 33, No. 3, 334–357
https://doi.org/10.1214/18-STS659

On the Relationship between the Theory of Cointegration and the Theory of Phase Synchronization

Rainer Dahlhaus, István Z. Kiss and Jan C. Neddermeyer

Abstract. The theory of cointegration has been a leading theory in econometrics with powerful applications to macroeconomics during the last decades. On the other hand, the theory of phase synchronization for weakly coupled complex oscillators has been one of the leading theories in physics for many years with many applications to different areas of science. For example, in neuroscience phase synchronization is regarded as essential for functional coupling of different brain regions. In an abstract sense, both theories describe the dynamic fluctuation around some equilibrium. In this paper, we point out that there exists a very close connection between both theories. Apart from phase jumps, a stochastic version of the Kuramoto equations can be approximated by a cointegrated system of difference equations. As one consequence, the rich theory on statistical inference for cointegrated systems can immediately be applied for statistical inference on phase synchronization based on empirical data. This includes tests for phase synchronization, tests for unidirectional coupling and the identification of the equilibrium from data including phase shifts. We study two examples on a unidirectionally coupled Rössler–Lorenz system and on electrochemical oscillators. The methods from cointegration may also be used to investigate phase synchronization in complex networks. Conversely, there are many interesting results on phase synchronization which may inspire new research on cointegration.

Key words and phrases: Cointegration, phase synchronization, weakly coupled oscillators, driver response relationship, Rössler–Lorenz system.

REFERENCES

ALLEFELD, C. and KURTHS, J. (2004a). An approach to multivariate phase synchronization analysis and its application to event-related potentials. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **14** 417–426. MR2069592

ALLEFELD, C. and KURTHS, J. (2004b). Testing for phase synchronization. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **14** 405–416. MR2069591

BANERJEE, A., GALBRAITH, J., DOLADO, J. and HENDRY, D. (1993). *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford Univ. Press, Oxford.

BAPTISTA, M., SILVA, T., SARTORELLI, J., CALDAS, I. and ROSA, E. JR. (2003). Phase synchronization in the perturbed Chua circuit. *Phys. Rev. E* **67** 056212.

BLASIUS, B., HUPPERT, A. and STONE, L. (1999). Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399** 354–359.

BOCCALETTI, S., PECORA, L. and PELAEZ, A. (2001). Unifying framework for synchronization of coupled dynamical systems. *Phys. Rev. E* **63** 066219.

BOCCALETTI, S., KURTHS, J., OSIPOV, G., VALLADARES, D. L. and ZHOU, C. S. (2002). The synchronization of chaotic systems. *Phys. Rep.* **366** 1–101. MR1913567

BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. Reprint of the 1981 edition. MR1853554

Rainer Dahlhaus is Professor, Institute of Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany (e-mail: dahlhaus@statlab.uni-heidelberg.de). István Z. Kiss is Arts and Sciences Associate Professor of Chemistry, Department of Chemistry, Saint Louis University, 3501 Laclede Ave. St. Louis, Missouri 63103, USA (e-mail: izkiss@slu.edu). Jan C. Neddermeyer is Quantitative Analyst, DZ BANK AG, 60325 Frankfurt am Main, Germany (e-mail: jc@neddermeyer.net).

COLGIN, L. and MOSER, E. (2010). Gamma oscillations in the hippocampus. *Physiology* **25** 319–329.

DAHLHAUS, R., KURTHS, R., MAAS, P. and TIMMER, J., eds. (2008). *Mathematical Methods in Time Series Analysis and Digital Image Processing*. Springer, Berlin and Heidelberg.

DAHLHAUS, R., DUMONT, T., LE CORFF, S. and NEDDERMEYER, J. C. (2017). Statistical inference for oscillation processes. *Statistics* **51** 61–83. MR3600462

DAVID, O., COSMELLI, D., LACHAUX, J.-P., BAILLET, S., GARNERO, L. and MARTINERIE, J. (2003). A theoretical and experimental introduction to the non-invasive study of large-scale neural phase synchronization in human beings. *Internat. J. Comput. Cog.* **1** 53–77.

DESHAZER, D. J., BREBAN, R., OTT, E. and ROY, R. (2001). Detecting phase synchronization in a chaotic laser array. *Phys. Rev. Lett.* **87** 044101.

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431. MR0548036

DICKEY, D. A. and FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057–1072. MR0625773

ELSON, R., SELVERSTON, A., HUERTA, R., RULKOV, N., RABINOVICH, M. and ABARBANEL, H. (1998). Synchronous behavior of two coupled biological neurons. *Phys. Rev. Lett.* **81** 5692–5695.

ENGEL, A., FRIES, P. and SINGER, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat. Rev., Neurosci.* **2** 704–716.

ENGLE, R. F. and GRANGER, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55** 251–276. MR0882095

ENGLE, R. and WHITE, H. (1999). *Cointegration, Causality and Forecasting*. Oxford Univ. Press, Oxford.

FELL, J. and AXMACHER, N. (2011). The role of phase synchronization in memory processes. *Nat. Rev., Neurosci.* **12** 105–118.

FULLER, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed. Wiley, New York. MR1365746

GRANGER, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16** 121–130.

GREENE, W. (2008). *Econometric Analysis*, 6th ed. Pearson Prentice Hall, New Jersey.

GROSSMANN, A., KRONLAND-MARTINET, R. and MORLET, J. (1989). Reading and understanding continuous wavelet transforms. In *Wavelets, Time-Frequency Methods and Phase Space. Inverse Probl. Theoret. Imaging* (J. Combes, ed.) 2–20. Springer, Berlin. MR1010896

GUAN, S., LAI, C.-H. and WEI, G. W. (2005). Phase synchronization between two essentially different chaotic systems. *Phys. Rev. E* (3) **72** 016205. MR2178376

HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. MR1278033

HANNAN, E. J. (1973). The estimation of frequency. *J. Appl. Probab.* **10** 510–519. MR0370977

HENNINGSEN, A. and HAMANN, J. D. (2007). systemfit: A package for estimating systems of simultaneous equations in r. *J. Stat. Softw.* **23** 1–40.

HORVATH, M. T. K. and WATSON, M. W. (1995). Testing for cointegration when some of the cointegrating vectors are prespecified. *Econometric Theory* **11** 984–1014. Trending multiple time series (New Haven, CT, 1993). MR1458947

JOHANSEN, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59** 1551–1580. MR1135641

JOHANSEN, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford Univ. Press, New York. MR1487375

JUSELIUS, K. (2006). *The Cointegrated VAR Model: Methodology and Applications. Advanced Texts in Econometrics*. Oxford Univ. Press, Oxford. MR2358946

KAMMERDINER, A. R. and PARDALOS, P. M. (2010). Analysis of multichannel EEG recordings based on generalized phase synchronization and cointegrated VAR. In *Computational Neuroscience. Springer Optim. Appl.* **38** 317–339. Springer, New York. MR2656819

KAMMERDINER, A., BOYKO, N., YE, N., HE, J. and PARDALOS, P. (2010). Integration of signals in complex biophysical system. In *Dynamics of Information Systems* (M. Hirsch, P. Pardalos and R. Murphey, eds.) 197–211. Springer, New York. MR2572384

KESSLER, M. and RAHBEK, A. (2001). Asymptotic likelihood based inference for co-integrated homogenous Gaussian diffusions. *Scand. J. Stat.* **28** 455–470. MR1858411

KISS, I. Z. and HUDSON, J. L. (2001). Phase synchronization and suppression of chaos through intermittency in forcing of an electrochemical oscillator. *Phys. Rev. E* **64** 046215. DOI:10.1103/PhysRevE.64.046215.

KISS, I. and HUDSON, J. (2002). Phase synchronization of nonidentical chaotic electrochemical oscillators. *Phys. Chem. Chem. Phys.* **4** 2638–2647.

KISS, I., LV, Q. and HUDSON, J. (2005). Synchronization of non-phase-coherent chaotic electrochemical oscillations. *Phys. Rev. E* **71** 035201.

KOCAREV, L. and PARLITZ, U. (1996). Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems. *Phys. Rev. Lett.* **76** 1816–1819.

KREMERS, J. J. M., ERICSSON, N. R. and DOLADO, J. J. (1992). The power of cointegration tests. *Oxf. Bull. Econ. Stat.* **54** 325–48.

KURAMOTO, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In *Lecture Notes in Phys.* **39** 420–422. Springer, Berlin. MR0676492

KURAMOTO, Y. (1984). *Chemical Oscillations, Waves, and Turbulence. Springer Series in Synergetics* **19**. Springer, Berlin. MR0762432

KWIATKOWSKI, D., PHILLIPS, P., SCHMIDT, P. and SHIN, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econometrics* **54** 159–178.

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368

MARAUN, D. and KURTHS, J. (2005). Epochs of phase coherence between El Niño/Southern Oscillation and Indian monsoon. *Geophys. Res. Lett.* **32**. DOI:10.1029/2005GL023225.

MORMANN, F., LEHNERTZ, K., DAVID, P. and ELGER C, E. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Phys. D* **144** 358–369.

MOSCONI, R. and OLIVETTI, F. (2005). Bivariate generalizations of the ACD models. Presented at the *Journal of Applied Econometrics Annual Conference*, *Venezia*.

OSIPOV, G. V., KURTHS, J. and ZHOU, C. (2007). *Synchronization in Oscillatory Networks*. *Springer Series in Synergetics*. Springer, Berlin. MR2350638

PALUŠ, M. and VEJMELKA, M. (2007). Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E* (3) **75** 056211. MR2361831

PALUŠ, M., KOMÁREK, V., HRNČÍŘ, Z. and ŠTĚRBOVÁ, K. (2001). Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E* **63** 046211. DOI:10.1103/PhysRevE.63.046211.

PALUT, Y. and ZANONE, P.-G. (2005). A dynamical analysis of tennis: Concepts and data. *J. Sports Sci*. **23** 1021–1032.

PARASCHAKIS, K. and DAHLHAUS, R. (2012). Frequency and phase estimation in time series with quasi periodic components. *J. Time Series Anal*. **33** 13–31. MR2877605

PECORA, L. M. and CARROLL, T. L. (1990). Synchronization in chaotic systems. *Phys. Rev. Lett*. **64** 821–824. MR1038263

PFAFF, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*, 2nd ed. *Use R!* Springer, New York. MR2450313

PHILLIPS, P. C. B. (1991). Error correction and long-run equilibrium in continuous time. *Econometrica* **59** 967–980. MR1113542

PHILLIPS, P. C. B. and OULIARIS, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* **58** 165–193. MR1046923

PIKOVSKY, A., ROSENBLUM, M. and KURTHS, J. (2001a). *Synchronization*: *A Universal Concept in Nonlinear Sciences*. *Cambridge Nonlinear Science Series* **12**. Cambridge Univ. Press, Cambridge, MA. MR1869044

PIKOVSKY, A. S., ROSENBLUM, M. G., OSIPOV, G. V. and KURTHS, J. (1997). Phase synchronization of chaotic oscillators by external driving. *Phys. D* **104** 219–238. MR1453087

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes in C*, 2nd ed. Cambridge Univ. Press, Cambridge.

PUJOL-PERE, A., CALVO, O., MATIAS, M. and KURTHS, J. (2003). Experimental study of imperfect phase synchronization in the forced Lorenz system. *Chaos* **13** 319–326.

QUIAN QUIROGA, R., KREUZ, T. and GRASSBERGER, P. (2000). Learning driver-response relationships from synchronization patterns. *Phys. Rev. E* **61** 5142–5148.

QUIAN QUIROGA, R., KREUZ, T. and GRASSBERGER, P. (2002). Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E* **65** 041903. DOI:10.1103/PhysRevE.65.041903.

ROSENBLUM, M., PIKOVSKY, A. and KURTHS, J. (1996). Phase synchronization of chaotic oscillators. *Phys. Rev. Lett*. **76**. 1804–1807.

SAIKKONEN, P. and LÜTKEPOHL, H. (2000). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory* **16** 373–406. MR1769280

SCHELTER, B., WINTERHALDER, M., TIMMER, J. and PEIFER, M. (2007). Testing for phase synchronization. *Phys. Lett. A* **366** 382–390.

STEFANOVSKA, A. (2002). Cardiorespiratory interactions. *Nonlinear Phenom. Complex Syst*. **5** 462–469. MR2013768

STEFANOVSKA, A., HAKEN, H., MCCLINTOCK, P. V. E., HOŽIČ, M., BAJROVIĆ, F. and RIBARIČ, S. (2000). Reversible transitions between synchronization states of the cardiorespiratory system. *Phys. Rev. Lett*. **85** 4831–4834. DOI:10.1103/PhysRevLett.85.4831.

STROGATZ, S. H. (2000). From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Phys. D* **143** 1–20. Bifurcations, patterns and symmetry. MR1783382

TASS, P., ROSENBLUM, M. G., WEULE, J., KURTHS, J., PIKOVSKY, A., VOLKMANN, J., SCHNITZLER, A. and FREUND, H. J. (1998). Detection of n:m phase locking from noisy data: Application to magnetoencephalography. *Phys. Rev. Lett*. **81** 3291–3294.

VAN LEEUWEN, P., GEUE, D., THIEL, D., CYSARZ, D., LANGE, S., ROMANO, M., WESSEL, N., KURTHS, J. and GRÖNEMEYER, D. (2009). Influence of paced maternal breathing on fetal-maternal heart rate coordination. In *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS*) **106** 13661–13666.

VARELA, F., LACHAUX, J.-P., RODRIGUEZ, E. and MARTINERIE, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nat. Rev. Neurosci*. **2** 229–239.

WINFREE, A. T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. *J. Theor. Biol*. **16** 15–42.

WOMELSDORF, T., SCHOFFELEN, J.-M., OOSTENVELD, R., SINGER, W., DESIMONE, R., ENGEL, A. K. and FRIES, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* **316** 1609–1612.

Statistical Science
2018, Vol. 33, No. 3, 358–385
https://doi.org/10.1214/17-STS641

Confidentiality and Differential Privacy in the Dissemination of Frequency Tables

Yosef Rinott, Christine M. O'Keefe, Natalie Shlomo and Chris Skinner

Abstract. For decades, national statistical agencies and other data custodians have been publishing frequency tables based on census, survey and administrative data. In order to protect the confidentiality of individuals represented in the data, tables based on original data are modified before release. Recently, in response to user demand for more flexible and responsive table publication services, frequency table publication schemes have been augmented with on-line table generating servers such as the US Census Bureau FactFinder and the Australian Bureau of Statistics (ABS) TableBuilder. These systems allow users to build their own custom tables, and make use of automated perturbation routines to protect confidentiality. Motivated by the growing popularity of table generating servers, in this paper we study confidentiality protection for perturbed frequency tables, including the trade-off with analytical utility, focusing on a version of the ABS TableBuilder as a concrete example of a data release mechanism, and examining its properties. Confidentiality protection is assessed in terms of the differential privacy standard, and this paper can be used as a practical introduction to differential privacy, to calculations related to its application, to the relationship between confidentiality protection and utility and to confidentiality in general.

Key words and phrases: Differential privacy, statistical disclosure control, contingency tables, utility.

REFERENCES

ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, H. B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security* 308–318. ACM, New York.

ANDERSSON, K., JANSSON, I. and KRAFT, K. (2015). Protection of frequency tables—current work at statistics Sweden. In *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality* (*Helsinki*, *Finland*, 5–7 *October*). 20 pp.

AUGUSTE, K. (1883). La cryptographie militaire. *J. Sci. Mil.* **9** 538.

BARAK, B., CHAUDHURI, K., DWORK, C., KALE, S., MCSHERRY, F. and TALWAR, K. (2007). Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In *Proceedings of the* 26*th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (*PODS*) 273–282.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis*: *Theory and Practice*. MIT Press, Cambridge, MA. MR0381130

BRENNER, H. and NISSIM, K. (2010). Impossibility of differentially private universally optimal mechanisms. In *Foundations of Computer Science* (*FOCS*), 2010 51*st Annual IEEE Symposium on* 71–80. IEEE, New York.

CHAREST, A.-S. (2010). How can we analyse differentially-private synthetic datasets? *J. Priv. Confid.* **2** 21–33.

CHAUDHURI, K. and MISHRA, N. (2006). When random sampling preserves privacy. In *Proceedings of the* 26*th Annual International Conference on Advances in Cryptology*: *CRYPTO* 2006 (C. Dwork, ed.). *LNCS* **4117** 198–213. Springer, Berlin.

Yosef Rinott is Professor, The Federmann Center for the Study of Rationality, Hebrew University, Jerusalem 91904, Israel (e-mail: yosef.rinott@mail.huji.ac.il). Christine O'Keefe is Senior Principal Research Scientist, CSIRO, GPO Box 1700, Canberra ACT 2601, Australia (e-mail: Christine.O'Keefe@csiro.au). Natalie Shlomo is Professor of Social Statistics, University of Manchester, Manchester M13 9PL, United Kingdom (e-mail: natalie.shlomo@manchester.ac.uk). Chris Skinner is Professor of Statistics, London School of Economics and Political Science, London WC2A 2AE, United Kingdom (e-mail: c.j.skinner@lse.ac.uk).

CHIPPERFIELD, J., GOW, D. and LOONG, B. (2016). The Australian Bureau of Statistics and releasing frequency tables via a remote server. *Stat. J. IAOS* **32** 53–64.

COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley, New York. MR2239987

DRECHSLER, J. (2012). New data dissemination approaches in old Europe—synthetic datasets for a German establishment survey. *J. Appl. Stat.* **39** 243–265.

DRECHSLER, J. and REITER, J. P. (2011). An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets. *Comput. Statist. Data Anal.* **55** 3232–3243.

DUNCAN, G. T., ELLIOT, M. and SALAZAR-GONZÀLEZ, J. J. (2011). *Statistical Confidentiality*. Springer, New York.

DUNCAN, G. T., FIENBERG, S. E., KRISHNAN, R., PADMAN, R. and ROEHRIG, S. F. (2001). Disclosure limitation methods and information loss for tabular data. In *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies* 135–166.

DWORK, C. (2006). Differential privacy. In *ICALP* 2006 (M. Bugliesi, B. Preneel, V. Sassone and I. Wegener, eds.). *Lecture Notes in Computer Science* **4052** 1–12. Springer, Heidelberg.

DWORK, C. and ROTH, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–407.

DWORK, C. and ROTHBLUM, G. N. (2016). Concentrated differential privacy. Preprint. Available at arXiv:1603.01887.

DWORK, C., ROTHBLUM, G. N. and VADHAN, S. (2010). Boosting and differential privacy. In 2010 *IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS* 2010 51–60. IEEE Computer Soc., Los Alamitos, CA. MR3024775

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *3rd IACR Theory of Cryptography Conference* 265–284.

EVETT, I. W., JACKSON, G., LAMBERT, J. A. and MCCROSSAN, S. (2000). The impact of the principles of evidence interpretation on the structure and content of statements. *Sci. Justice* **40** 233–239.

FELLEGI, I. P. (1972). On the question of statistical confidentiality. *J. Amer. Statist. Assoc.* **67** 7–18.

FIENBERG, S. E., RINALDO, A. and YANG, X. (2010). Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In *PSD'*2010 *Privacy in Statistical Databases* (J. Domingo-Ferrer and E. Magkos, eds.). *LNCS* **6344** 187–199. Springer, Berlin.

FIENBERG, S. E. and SLAVKOVIĆ, A. B. (2008). A survey of statistical approaches to preserving confidentiality of contingency table entries. In *Privacy-Preserving Data Mining* 291–312. Springer, Berlin.

FRASER, B. and WOOTON, J. (2005). A proposed method for confidentialising tabular output to protect against differencing. In *Joint UNECE/Eurostat Conference on Statistical Disclosure Control, Geneva, Switzerland*, 9–11 *November*. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.35.e.pdf.

FULLER, W. A. (1993). Masking procedures for microdata disclosure limitation. *J. Off. Stat.* **9** 383–383.

GABOARDI, M., ARIAS, E. J. G., HSU, J., ROTH, A. and WU, Z. S. (2016). Dual query: Practical private query release for high dimensional data. *J. Priv. Confid.* **7** 53–77.

GENG, Q. and VISWANATH, P. (2016). The optimal noise-adding mechanism in differential privacy. *IEEE Trans. Inform. Theory* **62** 925–951.

GHOSH, A., ROUGHGARDEN, T. and SUNDARARAJAN, M. (2012). Universally utility-maximizing privacy mechanisms. *SIAM J. Comput.* **41** 1673–1693.

GOMATAM, S. and KARR, A. (2003). Distortion measures for categorical data swapping. Technical report, National Institute of Statistical Sciences. Available at www.niss.org/downloadabletechreports.html.

GOTZ, M., MACHANAVAJJHALA, A., WANG, G., XIAO, X. and GEHRKE, J. (2012). Publishing search logs—a comparative study of privacy guarantees. *IEEE Trans. Knowl. Data Eng.* **24** 520–532.

GYMREK, M., MCGUIRE, A. L., GOLAN, D., HALPERIN, E. and ERLICH, Y. (2013). Identifying personal genomes by surname inference. *Science* **339** 321–324.

HARDT, M., LIGETT, K. and MCSHERRY, F. (2012). A simple and practical algorithm for differentially private data release. In *Advances in Neural Information Processing Systems* 2339–2347.

HAY, M., RASTOGI, V., MIKLAU, G. and SUCIU, D. (2010). Boosting the accuracy of differentially private histograms through consistency. *Proc. VLDB Endow.* **3** 1021–1032.

HAY, M., MACHANAVAJJHALA, A., MIKLAU, G., CHEN, Y. and ZHANG, D. (2016). Principled evaluation of differentially private algorithms using DPBench. In *Proceedings of the* 2016 *International Conference on Management of Data* 139–154 ACM, New York.

HOMER, N., SZELINGER, S., REDMAN, M., DUGGAN, D., TEMBE, W., MUEHLING, J., PEARSON, J. V., STEPHAN, D. A., NELSON, S. F. and CRAIG, D. W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* **4** e1000167.

HUNDEPOOL, A., DOMINGO-FERRER, J., FRANCONI, L., GIESSING, S., NORDHOLT, E. S., SPICER, K. and DE WOLF, P. P. (2012). *Statistical Disclosure Control*. Wiley, Chichester. MR3026260

JANSSON, I. (2012). Issues and plans for the disclosure control of the Swedish Census 2011. Technical Report No. 2012-04-02, Statistika centralbyrån.

KAIROUZ, P., OH, S. and VISWANATH, P. (2017). The composition theorem for differential privacy. *IEEE Trans. Inform. Theory* **63** 4037–4049. MR3677761

KARR, A. F., KOHNEN, C. N., OGANIAN, A., REITER, J. P. and SANIL, A. P. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *Amer. Statist.* **60** 224–232.

KARWA, V., KIFER, D. and SLAVKOVIĆ, A. B. (2015). Private posterior distributions from variational approximations. Preprint. Available at arXiv:1511.07896.

KARWA, V., SLAVKOVIĆ, A. et al. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.* **44** 87–112. MR3449763

LI, C., MIKLAU, G., HAY, M., MCGREGOR, A. and RASTOGI, V. (2015). The matrix mechanism: Optimizing linear counting queries under differential privacy. *VLDB J.* **24** 757–781.

LITTLE, R. (1993). Statistical analysis of masked data. *J. Off. Stat.* **9** 407–426.

LIU, F. (2017). Generalized gaussian mechanism for differential privacy. Preprint. Available at arXiv:1602.06028v5.

LONGHURST, J., TROMANS, N., YOUNG, C. and MILLER, C. (2007). Statistical disclosure control for the 2011 UK census. In *Joint UNECE/Eurostat conference on Statistical Disclosure Control*, *Manchester*, 17–19 *December*. Available at http://ec.europa.eu/eurostat/documents/1001617/4569122/TOPIC-3-WP-28-IP-LONGHURST-ET-ALREV.pdf.

MACHANAVAJJHALA, A., KIFER, D., ABOWD, J., GEHRKE, J. and VILHUBER, L. (2008). Privacy: Theory meets practice on the map. In *Proceedings of the IEEE* 24*th International Conference on Data Engineering ICDE* 277–286.

MARLEY, J. K. and LEAVER, V. L. (2011). A method for confidentialising user-defined tables: Statistical properties and a risk-utility analysis. In *Proc*. 58*th Congress of the International Statistical Institute*, *ISI* 2011 21–26.

MCSHERRY, F. and MIRONOV, I. (2009). Differentially private recommender systems: Building privacy into the net. In *Proceedings of the* 15*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 627–636. ACM, New York.

MCSHERRY, F. and TALWAR, K. (2007). Mechanism design via differential privacy. In *Foundations of Computer Science*, 2007. *FOCS'*07. 48*th Annual IEEE Symposium on* 94–103. IEEE, New York.

NARAYANAN, A. and SHMATIKOV, V. (2008). Robust de-anonymization of large datasets. In *Proc IEEE Security & Privacy Conference* 111–125.

O'KEEFE, C. M. and CHIPPERFIELD, J. O. (2013). A summary of attack methods and protective measures for fully automated remote analysis systems. *Int*. *Stat*. *Rev.* **81** 426–455.

RUBIN, D. B. (1993). Discussion: Statistical disclosure limitation. *J. Off*. *Stat*. **9** 462–468.

SHANNON, C. E. (1949). Communication theory of secrecy systems. *Bell Syst*. *Tech*. *J*. **28** 656–715.

SHLOMO, N. (2007). Statistical disclosure control methods for census frequency tables. *Int*. *Stat*. *Rev.* **75** 199–217.

SHLOMO, N., ANTAL, L. and ELLIOT, M. (2015). Measuring disclosure risk and data utility for flexible table generators. *J. Off*. *Stat*. **31** 305–324.

SHLOMO, N. and YOUNG, C. (2008). Invariant post-tabular protection of census frequency counts. In *PSD'*2008 *Privacy in Statistical Databases* (J. Domingo-Ferrer and Y. Saygin, eds.). *LNCS* **5261** 77–89. Springer, Berlin.

STEINKE, T. and ULLMAN, J. (2016). Between pure and approximate differential privacy. *J. Priv. Confid.* **7** 3–22.

SWEENEY, L. (1997). Weaving technology and policy together to maintain confidentiality. *J. Law Med*. *Ethics* **25** 98–110.

THOMPSON, G., BROADFOOD, S. and ELAZAR, D. (2013). Methodology for automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics. In *Joint UNECE/Eurostat conference on Statistical Disclosure Control*, *Ottawa*, 28–30 *October*. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_1_ABS.pdf.

UHLER, C., SLAVKOVIĆ, A. and FIENBERG, S. E. (2013). Privacy-preserving data sharing for genome-wide association studies. *J. Priv. Confid*. **5** 137–166.

VAN DEN HOUT, A. and VAN DER HEIJDEN, P. G. M. (2002). Randomized response, statistical disclosure control and misclassification: A review. *Int*. *Stat*. *Rev.* **70** 269–288.

WANG, Y., LEE, J. and KIFER, D. (2017). Revisiting differentially private hypothesis tests for categorical data. Preprint. Available at arXiv:1511.03376v4.

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist*. *Assoc*. **60** 63–69.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist*. *Assoc*. **105** 375–389.

WILLENBORG, L. and DE WAAL, T. (2001). *Elements of Statistical Disclosure Control*. *Lecture Notes in Statistics* **155**. Springer, Berlin.

Statistical Science
2018, Vol. 33, No. 3, 386–412
https://doi.org/10.1214/18-STS648

Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo

Paul Fearnhead, Joris Bierkens, Murray Pollock and Gareth O. Roberts

Abstract. Recently, there have been conceptually new developments in Monte Carlo methods through the introduction of new MCMC and sequential Monte Carlo (SMC) algorithms which are based on continuous-time, rather than discrete-time, Markov processes. This has led to some fundamentally new Monte Carlo algorithms which can be used to sample from, say, a posterior distribution. Interestingly, continuous-time algorithms seem particularly well suited to Bayesian analysis in big-data settings as they need only access a small sub-set of data points at each iteration, and yet are still guaranteed to target the true posterior distribution. Whilst continuous-time MCMC and SMC methods have been developed independently we show here that they are related by the fact that both involve simulating a piecewise deterministic Markov process. Furthermore, we show that the methods developed to date are just specific cases of a potentially much wider class of continuous-time Monte Carlo algorithms. We give an informal introduction to piecewise deterministic Markov processes, covering the aspects relevant to these new Monte Carlo algorithms, with a view to making the development of new continuous-time Monte Carlo more accessible. We focus on how and why sub-sampling ideas can be used with these algorithms, and aim to give insight into how these new algorithms can be implemented, and what are some of the issues that affect their efficiency.

Key words and phrases: Bayesian statistics, big data, Bouncy Particle Sampler, continuous-time importance sampling, control variates, SCALE, Zig-Zag Sampler.

REFERENCES

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

BAKER, J., FEARNHEAD, P., FOX, E. B. and NEMETH, C. (2017). Control Variates for Stochastic Gradient MCMC. ArXiv e-prints, 1706.05439.

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

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab.* **10** 85–104. MR2394037

BESKOS, A. and ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15** 2422–2444. MR2187299

BIERKENS, J. (2016). Non-reversible Metropolis–Hastings. *Stat. Comput.* **26** 1213–1228. MR3538633

BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2016). The Zig–Zag Process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* To appear. arXiv:1607.03188.

Paul Fearnhead is Professor of Statistics, Department of Mathematics and Statistics, Lancaster University (e-mail: p.fearnhead@lancaster.ac.uk). Joris Bierkens is Assistant Professor, DIAM, TU Delft (e-mail: joris.bierkens@tudelft.nl). Murray Pollock is Assistant Professor, Department of Statistics, University of Warwick (e-mail: M.Pollock@warwick.ac.uk). Gareth O. Roberts is Professor of Statistics, Department of Statistics, University of Warwick (e-mail: Gareth.O.Roberts@warwick.ac.uk).

BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. MR3655855

BIERKENS, J., BOUCHARD-COTE, A., DUNCAN, A., DOUCET, A., FEARNHEAD, P., ROBERTS, G. and VOLLMER, S. (2017). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Probab. Lett*. To appear. arXiv:1701.04244.

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2017). The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc*. To appear.

BURQ, Z. A. and JONES, O. D. (2008). Simulation of Brownian motion at first-passage times. *Math. Comput. Simulation* **77** 64–71. MR2388251

CARPENTER, J., CLIFFORD, P. and FEARNHEAD, P. (1999). An improved particle filter for non-linear problems. *IEE Proc. Radar Sonar Navig*. **146** 2–7.

ÇINLAR, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ. MR0380912

DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388. MR0790622

DAVIS, M. H. A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré B, Probab. Stat*. **37** 155–194. MR1819122

DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978

DOUC, R., MOULINES, E. and OLSSON, J. (2014). Long-term stability of sequential Monte Carlo methods under verifiable conditions. *Ann. Appl. Probab.* **24** 1767–1802. MR3226163

DOUCET, A., GODSILL, S. J. and ANDRIEU, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput*. **10** 197–208.

DUBEY, K. A., REDDI, S. J., WILLIAMSON, S. A., POCZOS, B., SMOLA, A. J. and XING, E. P. (2016). Variance reduction in stochastic gradient Langevin dynamics. In *Advances in Neural Information Processing Systems* 1154–1162.

ETHIER, S. N. and KURTZ, T. G. (2005). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Statistics*. Wiley, New York.

FEARNHEAD, P., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Particle filters for partially observed diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **70** 755–777. MR2523903

FEARNHEAD, P., LATUSZYNSKI, K., ROBERTS, G. O. and SERMAIDIS, G. (2016). Continuous-time importance sampling: Monte Carlo methods which avoid time-discretisation error. Available at https://arxiv.org/abs/1712.06201.

FOULKES, W., MITAS, L., NEEDS, R. and RAJAGOPAL, G. (2001). Quantum Monte Carlo simulations of solids. *Rev. Modern Phys*. **73** 33.

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **73** 123–214. MR2814492

GUSTAFSON, P. (1998). A guided walk Metropolis algorithm. *Stat. Comput*. **8** 357–364.

KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist*. **5** 1–25. MR1380850

LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q*. **26** 403–413. MR0546120

LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. MR3694589

LIU, J. S. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc*. **90** 567–576. MR3363399

LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc*. **93** 1032–1044. MR1649198

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

MA, Y.-A., CHEN, T. and FOX, E. (2015). A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems* 2917–2925.

MCGRAYNE, S. B. (2011). *The Theory That Would Not die*: *How Bayes' Rule Cracked the Enigma Code*, *Hunted down Russian Submarines*, *& Emerged Triumphant from Two Centuries of Controversy*. Yale Univ. Press, New Haven, CT. MR3235684

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

NEAL, R. M. (2003). Slice sampling. *Ann. Statist*. **31** 705–767. MR1994729

NEAL, R. M. (2004). Improving asymptotic variance of MCMC estimators: non-reversible chains are better Technical report, No. 0406, Department of Statistics, University of Toronto.

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

NEISWANGER, W., WANG, C. and XING, E. P. (2014). Asymptotically exact, embarrassingly parallel MCMC. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence* 623–632. AUAI Press, Arlington.

ØKSENDAL, B. (1985). *Stochastic Differential Equations*. *Universitext*. Springer, Berlin. MR0804391

PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2017). Stochastic bouncy particle sampler. In *Proceedings of ICML*.

PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* (3) **85** 026703.

POLLOCK, M., JOHANSEN, A. M. and ROBERTS, G. O. (2016). On the exact and ε-strong simulation of (jump) diffusions. *Bernoulli* **22** 794–856. MR3449801

POLLOCK, M., FEARNHEAD, P., JOHANSEN, A. and ROBERTS, G. O. (2016). An unbiased and scalable Monte Carlo method for Bayesian inference for big data. arXiv:1609.03436.

QUIROZ, M., VILLANI, M. and KOHN, R. (2015). Speeding up MCMC by efficient data subsampling. *J. Amer. Statist. Assoc.* To appear. Available at https://www.tandfonline.com/doi/abs/10.1080/01621459.2018.1448827.

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

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691

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

SHERLOCK, C. and THIERY, A. H. (2017). A discrete bouncy particle sampler. arXiv:1707.05200.

SRIVASTAVA, S., CEVHER, V., TRAN-DINH, Q. and DUNSON, D. B. (2015). WASP: Scalable Bayes via barycenters of subset posteriors. In *AISTATS*.

TIERNEY, L. and MIRA, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Stat. Med.* **18** 2507–2515.

VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise deterministic Markov chain Monte Carlo. arXiv:1707.05296.

WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the* 28*th International Conference on Machine Learning* (*ICML*-11) 681–688.

Statistical Science
2018, Vol. 33, No. 3, 413–426
https://doi.org/10.1214/18-STS649

Fractionally Differenced Gegenbauer Processes with Long Memory: A Review

G. S. Dissanayake, M. S. Peiris and T. Proietti

Abstract. The main objective of this paper is to review and promote the usefulness of generalized fractionally differenced Gegenbauer processes in time series and econometric research endeavours. In particular, theoretical and computational aspects centered around fractionally differenced Gegenbauer processes with long memory together with a number of interesting and elegant extensions will be discussed. In-depth conceptual developments and large scale simulation study results are presented for clarity and completeness. This survey highlights a number of gaps in the existing literature of this subject area and becomes a valuable reference source for time series practitioners.

Key words and phrases: Gegenbauer process, long memory, heteroskedasticity, fractional difference, volatility, spectral density, stationarity, invertibility.

REFERENCES

ABRAHAMS, M. and DEMPSTER, A. (1979). Research on seasonal analysis. Progress report on the asa/census project on seasonal adjustment. Technical report. Dept. Statistics, Harvard Univ., Boston, MA.

ANDĚL, J. (1986). Long memory time series models. *Kybernetika* (*Prague*) **22** 105–123. MR0849684

ANDERSON, B. D. O. and MOORE, J. B. (1979). *Optimal Filtering*. Prentice-Hall, New York.

ANH, V. V., ANGULO, J. M. and RUIZ-MEDINA, M. D. (1999). Possible long-range dependence in fractional random fields. *J. Statist. Plann. Inference* **80** 95–110. MR1713795

ANH, V. V., LUNNEY, K. and PEIRIS, S. (1997). Stochastic models for characterisation and prediction of time series with long-range dependence and fractality. *Environ. Model. Softw.* **12** 67–73.

AOKI, M. (1990). *State Space Modeling of Time Series*, 2nd ed. Springer, Berlin. MR1073171

ARTECHE, J. (2007). The analysis of seasonal long memory: The case of Spanish inflation. *Oxf. Bull. Econ. Stat.* **69** 749–772.

ARTECHE, J. (2012). Standard and seasonal long memory in volatility: An application to Spanish inflation. *Empir. Econ.* **42** 693–712.

ARTECHE, J. and ROBINSON, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *J. Time Series Anal.* **21** 1–25. MR1766171

BAILLIE, R. T. (1996). Long memory processes and fractional integration in econometrics. *J. Econometrics* **73** 5–59. MR1410000

BAILLIE, R. T., BOLLERSLEV, T. and MIKKELSEN, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **74** 3–30. MR1409033

BEAUMONT, P. and RAMACHANDRAN, R. (2001). Robust estimation of GARMA model parameters with an application to cointegration among interest rates of industrialized countries. *Comput. Econ.* **17** 179–201.

BERAN, J. (1992). Statistical methods for data with long-range dependence. *Statist. Sci.* **7** 404–416.

BERAN, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika* **80** 817–822. MR1282790

BERAN, J. (1994). *Statistics for Long-Memory Processes*. *Monographs on Statistics and Applied Probability* **61**. Chapman & Hall, New York. MR1304490

BERAN, J., FENG, Y., GHOSH, S. and KULIK, R. (2013). *Long-Memory Processes*: *Probabilistic Properties and Statistical Methods*. Springer, Heidelberg. MR3075595

BISOGNIN, C. and LOPES, S. R. C. (2009). Properties of seasonal long memory processes. *Math. Comput. Modelling* **49** 1837–1851. MR2532092

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051

G. S. Dissanayake is Senior Lecturer, Statistics, NSBM Green University, Homagama, Sri Lanka and Honorary Associate, School of Mathematics and Statistics, University of Sydney, Sydney NSW, Australia (e-mail: gnanadarshad@gmail.com). M. S. Peiris is Professor of Statistics, School of Mathematics and Statistics, University of Sydney, Sydney NSW, Australia. T. Proietti is Professor, Economics and Statistics, University of Rome Tor Vergata, Rome, Italy and Fellow, Center for Research in Econometric Analysis of Time Series, Aarhus University, Aarhus, Denmark.

BOX, G. E. P. and JENKINS, G. M. (1970). *Times Series Analysis. Forecasting and Control*. Holden-Day, San Francisco, CA. MR0272138

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. Springer, New York. MR1093459

BROCKWELL, P. J. and DAVIS, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York. MR1416563

CHAN, N. H. and PALMA, W. (1998). State space modeling of long-memory processes. *Ann. Statist.* **26** 719–740. MR1626083

CHAN, N. H. and PALMA, W. (2006). Estimation of long-memory time series models: A survey of different likelihood-based methods. *Adv. Econom.* **20** 89–121. MR2498114

CHUNG, C.-F. (1996). A generalized fractionally integrated autoregressive moving-average process. *J. Time Series Anal.* **17** 111–140. MR1381168

CORSI, F. (2009). A simple approximate long-memory model of realized volatility. *J. Financ. Econom.* **7** 174–196.

DISSANAYAKE, G. S. and PEIRIS, M. S. (2011). Generalized fractional processes with conditional heteroskedasticity. *Sri Lankan J. Appl. Stat.* **12** 1–12.

DISSANAYAKE, G. S., PEIRIS, M. S. and PROIETTI, T. (2014). Estimation of generalized fractionally differenced processes with conditionally heteroskedastic errors. In *International Work Conference on Time Series*, *Proceedings ITISE* 2014 (I. R. Ruiz and G. R. Garcia, eds.). Copicentro Granada S L 871–890.

DISSANAYAKE, G. S., PEIRIS, M. S. and PROIETTI, T. (2015). State space modeling of seasonal Gegenbauer processes with long memory. Working Paper, School of Mathematics and Statistics, Univ. Sydney, Australia.

DISSANAYAKE, G. S., PEIRIS, M. S. and PROIETTI, T. (2016). State space modeling of Gegenbauer processes with long memory. *Comput. Statist. Data Anal.* **100** 115–130. MR3505794

DISSANAYAKE, G. S., PEIRIS, M. S., PROIETTI, T. and WANG, Q. (2015). Nearly efficient testing and asymptotics of a long memory GARMA$(0, \delta, 0)$ process, Working Paper, School of Mathematics and Statistics, Univ. Sydney, Australia.

DOLADO, J. J., GONZALO, J. and MAYORAL, L. (2002). A fractional Dickey–Fuller test for unit roots. *Econometrica* **70** 1963–2006. MR1925162

DURBIN, J. and KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Methods*. *Oxford Statistical Science Series* **24**. Oxford Univ. Press, Oxford. MR1856951

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

ERDELYI, A., MAGNUS, W., OBERHETTINGER, F. and TRICOMI, F. G. (1953). *Higher Transcendental Functions*, *Vol. II*, *Bateman Manuscript Project*. McGraw-Hill. New York.

FERRARA, L. and GUEGAN, D. (2001). Forecasting with k-factor Gegenbauer processes: Theory and applications. *J. Forecast.* **20** 581–601.

FERRARA, L., GUEGAN, D. and LU, Z. (2010). Testing fractional order of long memory processes: A Monte Carlo study. *Comm. Statist. Simulation Comput.* **39** 795–806. MR2785667

GIRAITIS, L., HIDALGO, J. and ROBINSON, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *Ann. Statist.* **29** 987–1023. MR1869236

GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London. MR2977317

GIRAITIS, L. and LEIPUS, R. (1995). A generalized fractionally differencing approach in long-memory modeling. *Liet. Mat. Rink.* **35** 65–81. MR1357812

GONÇALVES, E. (1987). Une généralisation des processus ARMA. *Ann. Écon. Stat.* **5** 109–145. MR0907560

GOULD, H. W. (1974). Coefficient identities for powers of Taylor and Dirichlet series. *Amer. Math. Monthly* **81** 3–14. MR0340038

GRADSHTEYN, I. S. and RYZHIK, I. M. (1980). *Tables of Integrals Series and Products*. Academic Press, New York.

GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1** 15–29. MR0605572

GRASSI, S. and SANTUCCI DE MAGISTRIS, P. (2014). When long memory meets the Kalman filter: A comparative study. *Comput. Statist. Data Anal.* **76** 301–319. MR3209443

GRAY, H. L., ZHANG, N.-F. and WOODWARD, W. A. (1989). On generalized fractional processes. *J. Time Series Anal.* **10** 233–257. MR1028940

GRAY, H. L., ZHANG, N.-F. and WOODWARD, W. A. (1994). A correction: "On generalized fractional processes" [J. Time Ser. Anal. **10** (1989), no. 3, 233–257; MR1028940 (90m:62208)]. *J. Time Series Anal.* **15** 561–562. MR1292167

GUEGAN, D. (2000). A new model: The k-factor GIGARCH process. *J. Signal Process.* **4** 265–271.

GUÉGAN, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Rev.* **24** 113–149. MR2190313

HARVEY, A. C. (1989). *Forecasting*, *Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press, Cambridge.

HARVEY, A. C. and PROIETTI, T., eds. (2005). *Readings in Unobserved Components Models*. Oxford Univ. Press, Oxford. MR2381922

HASSLER, U. (1994). (Mis)specification of long memory in seasonal time series. *J. Time Series Anal.* **15** 19–30. MR1256854

HASSLER, U. and WOLTERS, J. (1994). On the power of unit root tests against fractional alternatives. *Econom. Lett.* **45** 1–5.

HASSLER, U. and WOLTERS, J. (1995). Long memory in inflation rates: International evidence. *J. Bus. Econom. Statist.* **13** 37–45.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. MR0614953

HSU, N.-J. and TSAI, H. (2009). Semiparametric estimation for seasonal long-memory time series using generalized exponential models. *J. Statist. Plann. Inference* **139** 1992–2009. MR2497555

JANSSON, M. and NIELSEN, M. Ø. (2012). Nearly efficient likelihood ratio tests of the unit root hypothesis. *Econometrica* **80** 2321–2332. MR3013726

KALMAN, R. E. (1961). A new approach to linear filtering and prediction problems. *Trans. Am. Soc. Mech. Eng.* **83D** 35–45.

KALMAN, R. E. and BUCY, R. S. (1961). New results in linear filtering and prediction theory. *Trans. Am. Soc. Mech. Eng.* **83** 95–108. MR0234760

KOOPMAN, S. J., OOMS, M. and CARNERO, M. A. (2007). Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *J. Amer. Statist. Assoc.* **102** 16–27. MR2345529

LIEBERMAN, O. and PHILLIPS, P. C. B. (2008). Refined inference on long memory in realized volatility. *Econometric Rev.* **27** 254–267. MR2424814

LING, S. and LI, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.* **92** 1184–1194. MR1482150

LOBATO, I. N. and SAVIN, N. E. (1998). Real and spurious long-memory properties of stock-market data. *J. Bus. Econom. Statist.* **16** 261–283. MR1648533

LU, Z. and GUEGAN, D. (2011). Estimation of time-varying long memory parameter using wavelet method. *Comm. Statist. Simulation Comput.* **40** 596–613. MR2783896

MCALEER, M. and MEDEIROS, M. C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *J. Econometrics* **147** 104–119. MR2472985

MONTANARI, A., ROSSO, R. and TAQQU, M. S. (2000). A seasonal fractional ARIMA modelapplied to Nile river monthly flows at Aswan. *Water Resour. Res.* **36** 1249–1259.

OHANISSIAN, A., RUSSELL, J. R. and TSAY, R. S. (2008). True or spurious long memory? A new test. *J. Bus. Econom. Statist.* **26** 161–175. MR2420145

OOMS, M. (1995). Flexible seasonal long memory and economic time series. Technical Report EI-9515/A. Econometric Institute, Erasmus Univ., Rotterdam.

OPPENHEIM, G. and VIANO, M.-C. (2004). Aggregation of random parameters Ornstein–Uhlenbeck or AR processes: Some convergence results. *J. Time Series Anal.* **25** 335–350. MR2062677

PALMA, W. (2007). *Long-Memory Time Series*: *Theory and Methods*. Wiley-Interscience, Hoboken, NJ. MR2297359

PALMA, W. and CHAN, N. H. (2005). Efficient estimation of seasonal long-range-dependent processes. *J. Time Series Anal.* **26** 863–892. MR2203515

PEARLMAN, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive–moving average process. *Biometrika* **67** 232–233. MR0570525

PEIRIS, M. S. (2003). Improving the quality of forecasting using generalized AR models: An application to statistical quality control. *Stat. Methods* **5** 156–171. MR2198741

PEIRIS, S., ALLEN, D. and PEIRIS, U. (2005). Generalized autoregressive models with conditional heteroscedasticity: An application to financial time series modeling. In *Proceedings of the Workshop on Research Methods*: *Statistics and Finance* 75–83.

PEIRIS, S. and ASAI, M. (2016). Generalized fractional processes with long memory and time dependent volatility revisited. *Econometrics* **4** 37.

PEIRIS, S. and THAVANESWARAN, A. (2007). An introduction to volatility models with indices. *Appl. Math. Lett.* **20** 177–182. MR2283907

PHILLIPS, P. C. B. and XIAO, Z. (1998). A primer on unit root testing. *J. Econ. Surv.* **12** 423–470.

PORTER-HUDAK, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Amer. Statist. Assoc.* **85** 338–344.

RAINVILLE, E. D. (1960). *Special Functions*. The Macmillan Co., New York. MR0107725

RAY, B. K. (1993). Modeling long-memory processes for optimal long-range prediction. *J. Time Series Anal.* **14** 511–525. MR1243579

REISEN, V. A., RODRIGUES, A. L. and PALMA, W. (2006). Estimation of seasonal fractionally integrated processes. *Comput. Statist. Data Anal.* **50** 568–582. MR2201879

ROBINSON, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics* **47** 67–84. MR1087207

SHEPHARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models*: *In Econometrics*, *Finance and Other Fields* (D. R. Cox, D. B. Hinkley, and O. E. Barndorff-Nielsen, eds.), Chapman & Hall, London.

SHITAN, M. and PEIRIS, S. (2008). Generalized autoregressive (GAR) model: A comparison of maximum likelihood and Whittle estimation procedures using a simulation study. *Comm. Statist. Simulation Comput.* **37** 560–570. MR2749969

SHITAN, M. and PEIRIS, S. (2009). On properties of the second order generalized autoregressive GAR(2) model with index. *Math. Comput. Simulation* **80** 367–377. MR2582119

SHITAN, M. and PEIRIS, S. (2013). Approximate asymptotic variance–covariance matrix for the Whittle estimators of GAR(1) parameters. *Comm. Statist. Theory Methods* **42** 756–770. MR3028974

SLUTSKY, E. (1927). The summation of random causes as the source of cyclic processes. *Econometrica* **5** 105–146.

SPOLIA, S. K., CHANDLER, S. and O'CONNOR, K. M. (1980). An autocorrelation approach for parameter estimation of fractional order equal-root autoregressive models using hypergeometric functions. *J. Hydrol.* **47** 1–18.

TAYLOR, A. M. R. (2005). Fluctuation tests for a change in persistence. *Oxf. Bull. Econ. Stat.* **67** 207–230.

VELASCO, C. and ROBINSON, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Amer. Statist. Assoc.* **95** 1229–1243. MR1804246

WANG, Q., LIN, Y.-X. and GULATI, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* **19** 143–164. MR1965845

WOODWARD, W. A., CHENG, Q. C. and GRAY, H. L. (1998). A k-factor GARMA long-memory model. *J. Time Series Anal.* **19** 485–504. MR1650550

YULE, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series?—A study in sampling and the nature of time-series. *J. Roy. Statist. Soc.* **89** 1–63.

Statistical Science
2018, Vol. 33, No. 3, 427–443
https://doi.org/10.1214/18-STS661

A Unified Theory of Confidence Regions and Testing for High-Dimensional Estimating Equations

Matey Neykov, Yang Ning, Jun S. Liu and Han Liu

Abstract. We propose a new inferential framework for constructing confidence regions and testing hypotheses in statistical models specified by a system of high-dimensional estimating equations. We construct an influence function by projecting the fitted estimating equations to a sparse direction obtained by solving a large-scale linear program. Our main theoretical contribution is to establish a unified Z-estimation theory of confidence regions for high-dimensional problems. Different from existing methods, all of which require the specification of the likelihood or pseudo-likelihood, our framework is likelihood-free. As a result, our approach provides valid inference for a broad class of high-dimensional constrained estimating equation problems, which are not covered by existing methods. Such examples include, noisy compressed sensing, instrumental variable regression, undirected graphical models, discriminant analysis and vector autoregressive models. We present detailed theoretical results for all these examples. Finally, we conduct thorough numerical simulations, and a real dataset analysis to back up the developed theoretical results.

Key words and phrases: Post-regularization inference, estimating equations, confidence regions, hypothesis tests, Dantzig selector, instrumental variables, graphical models, discriminant analysis, vector autoregressive models.

REFERENCES

BARBER, R. F. and KOLAR, M. (2015). Rocket: Robust confidence intervals via Kendall's tau for transelliptical graphical models. Preprint. Available at arXiv:1502.07641.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983

BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097

BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2013). Honest confidence regions for logistic regression with a large number of controls. Preprint. Available at arXiv:1304.3969.

CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395

CAI, T. T., LIANG, T. and RAKHLIN, A. (2014). Geometrizing local rates of convergence for linear inverse problems. Preprint. Available at arXiv:1404.4408.

CAI, T. and LIU, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *J. Amer. Statist. Assoc.* **106** 1566–1577. MR2896857

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

Matey Neykov is Assistant Professor, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: mneykov@stat.cmu.edu). Yang Ning is Assistant Professor, Department of Statistical Science, Cornell University, Ithaca, New York 14853, USA (e-mail: yn265@cornell.edu). Jun S. Liu is Associate Professor, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA (e-mail: jliu@stat.harvard.edu). Han Liu is Professor, Department of Electrical Engineering and Computer Science and Department of Statistics, Northwestern University, Evanston, Illinois 60208, USA (e-mail: hanliu@northwestern.edu).

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

CHEN, M., REN, Z., ZHAO, H. and ZHOU, H. (2016). Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *J. Amer. Statist. Assoc.* **111** 394–406. MR3494667

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory* **57** 5467–5484. MR2849368

GAUTIER, E. and TSYBAKOV, A. (2011). High-dimensional instrumental variables regression and confidence sets. Preprint. Available at arXiv:1105.2454.

GODAMBE, V. P. (1991). *Estimating functions*. Clarendon Press, Oxford.

GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional gaussian copula graphical models. Preprint. Available at arXiv:1502.02347.

HAN, F., LU, H. and LIU, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* **16** 3115–3150. MR3450535

HOLMES, K., ROBERTS, O. L., THOMAS, A. M. and CROSS, M. J. (2007). Vascular endothelial growth factor receptor-2: Structure, function, intracellular signalling and therapeutic inhibition. *Cellular Signalling* **19** 2003–2012.

JANKOVÁ, J. and VAN DE GEER, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* **9** 1205–1229. MR3354336

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2013). Exact inference after model selection via the lasso. Preprint. Available at arXiv:1311.6238.

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453

LIU, H., HAN, F. and ZHANG, C.-H. (2012). Transelliptical graphical models. In *Advances in Neural Information Processing Systems*.

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403

LU, S., LIU, Y., YIN, L. and ZHANG, K. (2015). Confidence intervals and regions for the lasso using stochastic variational inequality techniques in optimization. Technical report.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*: *Probability and Mathematical Statistics*. Academic Press, London. MR0560319

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics, Vol. IV. Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

NEYKOV, M., NING, Y., LIU, J. S and LIU, H. (2018). Supplement to "A Unified Theory of Confidence Regions and Testing for High Dimensional Estimating Equations." DOI:10.1214/18-STS661SUPP.

NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450

NING, Y. and LIU, H. (2013). High-dimensional semiparametric bigraphical models. *Biometrika* **100** 655–670. MR3094443

NING, Y. and LIU, H. (2014). Sparc: Optimal estimation and asymptotic inference under semiparametric sparsity. Preprint. Available at arXiv:1412.2295.

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695

SHAH, R. D. and SAMWORTH, R. J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. MR3008271

SMALL, C. G. and YANG, Z. (1999). Multiple roots of estimating functions. *Canad. J. Statist.* **27** 585–598. MR1745824

TAYLOR, J., LOCKHART, R., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). Post-selection adaptive inference for least angle regression and the lasso. Preprint. Available at arXiv:1401.3889.

TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

VOORMAN, A., SHOJAIE, A. and WITTEN, D. (2014). Inference in high dimensions with the penalized score test. Preprint. Available at arXiv:1401.2678.

WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. MR2543689

ZAHN, J. M., POOSALA, S., OWEN, A. B., INGRAM, D. K., LUSTIG, A., CARTER, A., WEERARATNA, A. T., TAUB, D. D., GOROSPE, M., MAZAN-MAMCZARZ, K. et al. (2007). AGEMAP: A gene expression database for aging in mice. *PLoS Genet.* **3** e201.

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940

ZHAO, S. D. (2012). Survival analysis with high-dimensional covariates, with applications to cancer genomics. Ph.D. thesis, Harvard Univ.

ZHU, Y. and BRADIC, J. (2016). Linear hypothesis testing in dense high-dimensional linear models. Preprint. Available at arXiv:1610.02987.

Statistical Science
2018, Vol. 33, No. 3, 444–457
https://doi.org/10.1214/18-STS655

A Conversation with Tom Louis

Lance A. Waller

Abstract. Thomas A. Louis received his BA in Mathematics from Dartmouth College in 1966 and his Ph.D. in Mathematical Statistics from Columbia University in 1972. He served as a NIH Postdoctoral Fellow at Imperial College, London, from 1972–1973 and has held faculty positions at Boston University, Harvard School of Public Health, the University of Minnesota and the Bloomberg School of Public Health at Johns Hopkins University. In addition, he served as a Senior Statistical Scientist at the RAND Corporation, and as Associate Director for Research and Methodology and Chief Scientist at the U.S. Census Bureau. Tom has served as President of both the Eastern North American Region of the International Biometric Society and as President of the International Biometric Society. He is a Fellow of the American Statistical Association, the American Association for the Advancement of Science and the Institute of Mathematical Statistics. As of January 2018, Tom is *Emeritus* Professor, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University. In addition to his many statistical accomplishments, Tom is a strong advocate for professional development and a life-long lover of time on the water.

Key words and phrases: Biography, Applied Statistics, Bayesian Statistics.

REFERENCES

APPEL, L. J., CLARK, J. M., YEH, H.-C., WANG, N.-Y., COUGHLIN, J. W., DAUMIT, G., MILLER, E., DALCIN, A., JEROME, G., GELLER, S., NORONHA, G., POZEFSKY, T., CHARLESTON, J., REYNOLDS, J., RUBIN, R., LOUIS, T. A. and BRANCATI, F. L. (2011). Accomplishing weight loss in clinical practice—A comparative effectiveness trial. *N. Engl. J. Med.* **365** 1959–1968.

BARNARD, G. A. (1963). Discussion of "The spectral analysis of point patterns". *J. Roy. Statist. Soc. Ser. B* **25** 294.

BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264–296. MR0171334

BESAG, J. (1975). Statistical analysis of non-lattice data. *Statistician* **24** 179–195.

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152

CARLIN, B. P. and LOUIS, T. A. (2009). *Bayesian Methods for Data Analysis*, 3rd ed. Chapman & Hall/CRC Press, Boca Raton, FL.

COMMITTEE TO REVIEW THE HEALTH CONSEQUENCES OF SERVICE DURING THE PERSIAN GULF WAR (1996). In Health Consequences of Service During the Persian Gulf War: Recommendations for Research and Information Systems. National Academies Press, Washington DC.

HILTS, V. L. (1978). Aliis exterendum, or, the origins of the Statistical Society of London. *Isis* **69** 21–43. MR0495096

INGALL, T., O'FALLON, M., ASPLUND, K., GOLDFRANK, L., HERTZBERG, V., LOUIS, T. A. and CHRISTIANSON, T. J. H. (2004). Findings from the reanalysis of the NINDS tissue plasminogen activator for acute ischemic stroke treatment trial. *Stroke* **35** 2418–2424.

KEMPEN, J. H., ALTAWEEL, M. M., HOLBROOK, J. T., JABS, D. A., LOUIS, T. A., SUGAR, E. A. and THORNE, J. E. (2011). Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant therapy for intermediate, posterior and panuveitis: The multicenter uveitis steroid treatment trial. *Ophthalmology* **118** 1916–1926.

LOUIS, T. A. (2007). Our future as history. *Biometrics* **63** 1–9, 309. MR2345569

LOUIS, T. A., ALBERT, A. and HEGHINIAN, S. (1978). Screening for the early detection of cancer III: Estimation of disease natural history. *Math. Biosci.* **40** 111–144.

LOUIS, T. A., FINEBERG, H. V. and MOSTELLER, F. (1985). Findings for public health from meta-analyses. *Ann. Rev. Public Health* **6** 1–20.

MCCAFFREY, D. F., LOCKWOOD, J. R., HAMILTON, L., KORETZ, D. and LOUIS, T. A. (2004). Models for value-added modelling of teacher effects. *J. Educ. Behav. Stat.* **29** 67–102.

MOSES, L. and LOUIS, T. A. (1992). Statistical consulting in clinical research: The two-way street. In *Medical Uses of Statis-*

Lance A. Waller is Rollins Professor and Chair, Department of Biostatistics and Bioinformatics Rollins School of Public Health Emory University 1518 Clifton Road, NE Atlanta, Georgia 30322, USA (e-mail: lwaller@emory.edu).

tics, 2nd ed. (J. C. Bailar and F. Mosteller, eds.). NEJM Books, Waltham, MA.

QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360. MR0081040

SHEN, W. and LOUIS, T. A. (1998). Triple-goal estimates in two-stage hierarchical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 455–471. MR1616061

SUTCLIFFE, C. G., KOBAYASHI, T., HAMAPUMBU, H., SHIELDS, T., MHARAKURWA, S., THUMA, P. E., LOUIS, T. A., GLASS, G. and MOSS, W. J. (2012). Reduced risk of malaria parasitemia following household screening and treatment: A cross-sectional and longitudinal cohort study. *PLoS ONE* **7** e31396.

Statistical Science
2018, Vol. 33, No. 3, 458–467
https://doi.org/10.1214/18-STS656

A Conversation with Jim Pitman

David Aldous

Abstract. Jim Pitman was born in June 1949, received a Ph.D. in 1974 from the University of Sheffield with advisor Terry Speed, and since 1979 has been in the U.C. Berkeley Statistics department. He is known for research on many topics within probability, in particular for a long-running collaboration with Marc Yor on distributional properties of Brownian motion, and for his influential lecture notes *Combinatorial Stochastic Processes*. The following conversation took place at his home in December 2017 and February 2018.

Key words and phrases: Mathematical probability, Markov chain, Brownian motion, combinatorial stochastic processes.

REFERENCES

[1] ALDOUS, D. and PITMAN, J. (1979). On the zero-one law for exchangeable events. *Ann. Probab.* **7** 704–723. MR0537216

[2] ALDOUS, D. and PITMAN, J. (1983). The asymptotic speed and shape of a particle system. In *Probability, Statistics and Analysis*. *London Mathematical Society Lecture Note Series* **79** 1–23. Cambridge Univ. Press, Cambridge. MR0696018

[3] ALDOUS, D. and PITMAN, J. (1998). The standard additive coalescent. *Ann. Probab.* **26** 1703–1726. MR1675063

[4] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983*. *Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. MR0883646

[5] ALDOUS, D. J. and PITMAN, J. (1994). Brownian bridge asymptotics for random mappings. *Random Structures Algorithms* **5** 487–512. MR1293075

[6] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading. MR0229267

[7] BREIMAN, L. (1969). *Probability and Stochastic Processes with a View Toward Applications*. Houghton Miffin Co, Boston, MA. MR0254942

[8] ÇINLAR, E., CHUNG, K. L. and GETOOR, R. K., eds. (1983) In *Seminar on Stochastic Processes*, 1982. *Progress in Probability and Statistics* **5**. Birkhäuser, Boston, MA. MR0733662

[9] DIACONIS, P. and FREEDMAN, D. (1980). de Finetti's theorem for Markov chains. *Ann. Probab.* **8** 115–130. MR0556418

[10] DIACONIS, P. and FREEDMAN, D. (1980). de Finetti's generalizations of exchangeability. In *Studies in Inductive Logic and Probability, Vol. II* 233–249. Univ. California Press, Berkeley, CA. MR0587993

[11] DYNKIN, E. B. (1961). *Theory of Markov Processes*. Prentice-Hall, Englewood Cliffs, NJ. MR0131900

[12] EDWARDS, R. E. (1967). *Fourier Series: A Modern Introduction. Vol. I*. Holt, Rinehard & Winston, New York. MR0216227

[13] EDWARDS, R. E. (1967). *Fourier Series: A Modern Introduction. Vol. II*. Holt, Rinehart & Winston, New York. MR0222538

[14] EVANS, S. N. and PITMAN, J. (1998). Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 339–383. MR1625867

[15] EWENS, W. J. (1969). *Population Genetics*. Methuen, London. MR0270768

[16] FREEDMAN, D. (1971). *Markov Chains*. Holden-Day, San Francisco, CA. MR0292176

[17] GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **325** 83–102, 244–245. MR2160320

[18] GNEDIN, A. and PITMAN, J. (2005). Regenerative composition structures. *Ann. Probab.* **33** 445–479. MR2122798

[19] GREENWOOD, P. and PITMAN, J. (1980). Construction of local time and Poisson point processes from nested arrays. *J. Lond. Math. Soc.* (2) **22** 182–192. MR0579823

[20] GREENWOOD, P. and PITMAN, J. (1980). Fluctuation identities for Lévy processes and splitting at the maximum. *Adv. in Appl. Probab.* **12** 893–902. MR0588409

[21] ITÔ, K. and MCKEAN, H. P. JR. (1965). *Diffusion Processes and Their Sample Paths*. *Die Grundlehren der Mathematischen Wissenschaften*, *Band* 125. Academic Press, New York. MR0199891

[22] JACOBSEN, M. (1974). Splitting times for Markov processes and a generalised Markov property for diffusions. *Z. Wahrsch. Verw. Gebiete* **30** 27–43. MR0375477

[23] LE GALL, J.-F. (1989). Marches aléatoires, mouvement brownien et processus de branchement. In *Séminaire de Probabilités, XXIII*. *Lecture Notes in Math.* **1372** 258–274. Springer, Berlin. MR1022916

[24] LE GALL, J.-F. (1993). The uniform random tree in a Brownian excursion. *Probab. Theory Related Fields* **96** 369–383. MR1231930

David Aldous is the Professor in the Statistics Department of the University of California, Berkeley, 367 Evans Hall, Berkeley, California 94720, USA (e-mail: aldous@stat.berkeley.edu; URL: www.stat.berkeley.edu/~aldous).

[25] LOVÁSZ, L. and WINKLER, P. (1998). Mixing times. In *Microsurveys in Discrete Probability* (*Princeton*, *NJ*, 1997). *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **41** 85–133. Amer. Math. Soc., Providence, RI. MR1630411

[26] MEYER, P. A. (ed.) (1976). *Séminaire de Probabilités*, *X*. *Lecture Notes in Mathematics*, *Vol.* 511. Springer, Berlin. MR0431319

[27] NEVEU, J. and PITMAN, J. (1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion. In *Séminaire de Probabilités*, *XXIII*. *Lecture Notes in Math.* **1372** 239–247. Springer, Berlin. MR1022914

[28] NEVEU, J. and PITMAN, J. W. (1989). The branching process in a Brownian excursion. In *Séminaire de Probabilités*, *XXIII*. *Lecture Notes in Math.* **1372** 248–257. Springer, Berlin. MR1022915

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

[30] PITMAN, J. (1999). Coalescent random forests. *J. Combin. Theory Ser. A* **85** 165–193. MR1673928

[31] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

[32] PITMAN, J. and YOR, M. (1981). Bessel processes and infinitely divisible laws. In *Stochastic Integrals* (*Proc. Sympos., Univ. Durham*, *Durham*, 1980). *Lecture Notes in Math.* **851** 285–370. Springer, Berlin. MR0620995

[33] PITMAN, J. and YOR, M. (1982). A decomposition of Bessel bridges. *Z. Wahrsch. Verw. Gebiete* **59** 425–457. MR0656509

[34] PITMAN, J. and YOR, M. (1986). Asymptotic laws of planar Brownian motion. *Ann. Probab.* **14** 733–779. MR0841582

[35] PITMAN, J. W. (1975). One-dimensional Brownian motion and the three-dimensional Bessel process. *Adv. in Appl. Probab.* **7** 511–526. MR0375485

[36] PITMAN, J. W. (1977). Occupation measures for Markov chains. *Adv. in Appl. Probab.* **9** 69–86. MR0433600

[37] SHETH, R. K. and PITMAN, J. (1997). Coagulation and branching process models of gravitational clustering. *Mon. Not. R. Astron. Soc.* **289** 66–82.

[38] STEELE, J. M. (2001). *Stochastic Calculus and Financial Applications*. *Applications of Mathematics* (*New York*) **45**. Springer, New York. MR1783083

[39] TAKÁCS, L. (1967). *Combinatorial Methods in the Theory of Stochastic Processes*. Wiley, New York. MR0217858

[40] WILLIAMS, D. (1974). Path decomposition and continuity of local time for one-dimensional diffusions. I. *Proc. Lond. Math. Soc.* (3) **28** 738–768. MR0350881

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

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

President-Elect: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

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

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

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

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

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

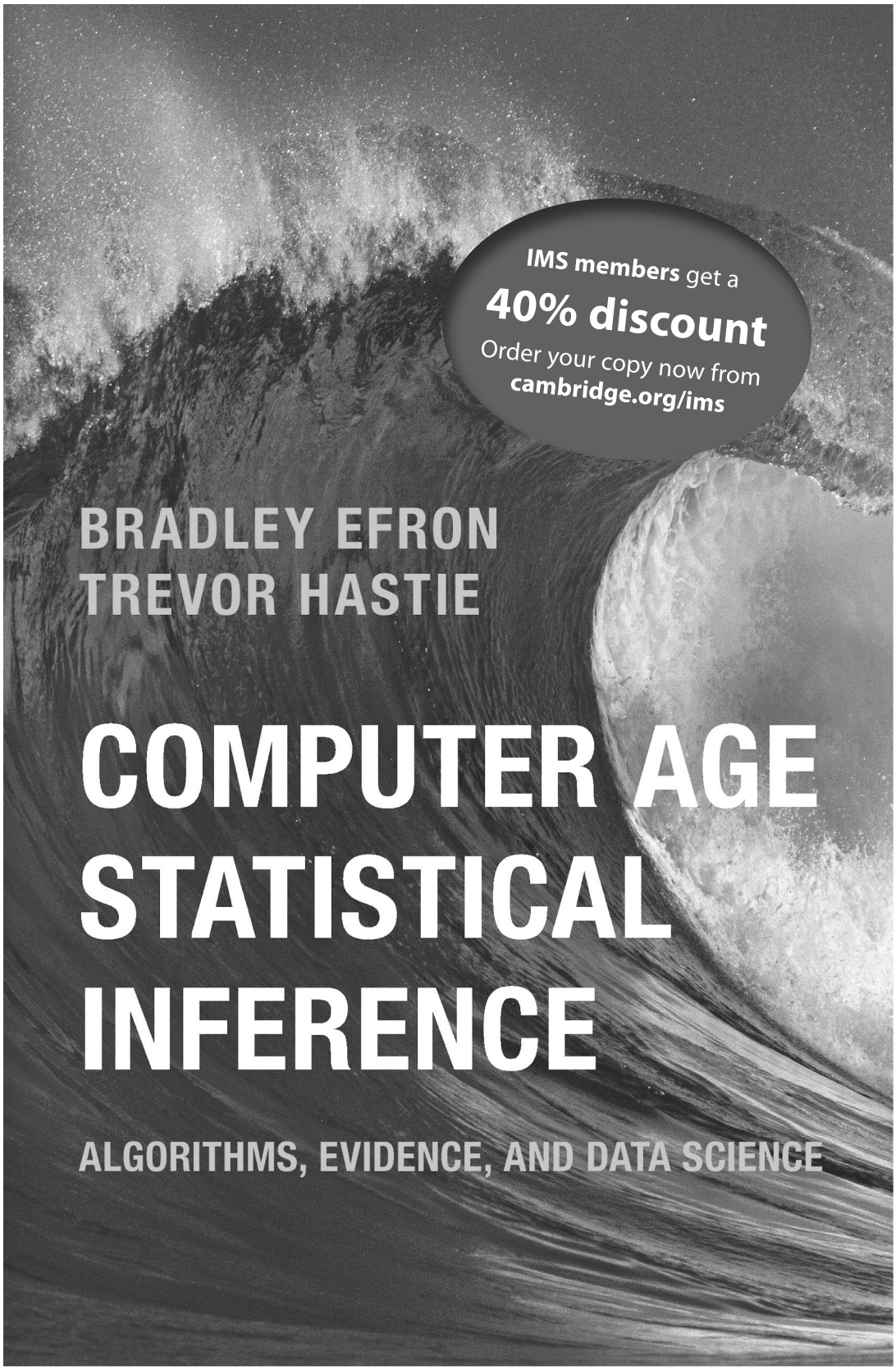
The Annals of Applied Statistics. *Editor-in-Chief*: Tilmann Gneiting, Heidelberg Institute for Theoretical Studies, HITS gGmbH, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

The Annals of Probability. *Editor*: Amir Dembo, Department of Statistics and Department of Mathematics, Stanford University, Stanford, California 94305, USA

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

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

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


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