Quantifying the cost of simultaneous non-parametric approximation of several samples

P. Laurie Davies, *Technical University of Eindhoven and University of Duisburg-Essen*
Arne Kovac, *University of Bristol*

Abstract

We consider the standard non-parametric regression model with Gaussian errors but where the data consist of different samples. The question to be answered is whether the samples can be adequately represented by the same regression function. To do this we define for each sample a universal, honest and non-asymptotic confidence region for the regression function. Any subset of the samples can be represented by the same function if and only if the intersection of the corresponding confidence regions is non-empty. If the empirical supports of the samples are disjoint then the intersection of the confidence regions is always non--empty and a negative answer can only be obtained by placing shape or quantitative smoothness conditions on the joint approximation, or by making additional assumptions about the support points. Alternatively, a simplest joint approximation function can be calculated which gives a measure of the cost of the joint approximation, for example, the number of extra peaks required.

Full Text: PDF

References

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proceedings of the American Mathematical Society, 6(2):170–176. MR0069229

Davies, P. L. (1995). Data features. Statistica Neerlandica, 49:185–245. MR1345378

Davies, P. L. (2004). The one-way table: In honour of John Tukey 1915-2000. Journal of Statistical Planning and Inference, 122:3–13. MR2057910

Davies, P. L. (2005). Universal principles, approximation and model choice. Invited talk, European Meeting of Statisticians, Oslo.

Davies, P. L. (2008). Approximating data (with discussion). Journal of the Korean Statistical Society, 37:191–240. MR2445029

Davies, P. L., Gather, U., Meise, M., Mergel, D., and Mildenberger, T. (2008a). Residual based localization and quantification of peaks in x-ray diffractograms. Annals of Applied Statistics, 2(3):861–886.

Davies, P. L., Gather, U., Nordman, D. J., and Weinert, H. (2008b). A comparison of automatic histogram constructions. EIMS: Probability and Statistics. to appear.

Davies, P. L., Gather, U., and Weinert, H. (2008c). Nonparametric regression as an example of model choice. Communications in Statistics - Simulation and Computation, 37:274 – 289. MR2422886

Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). Annals of Statistics, 29(1):1–65. MR1833958

Davies, P. L., Kovac, A., and Meise, M. (2009). Nonparametric regression, confidence regions and regularization. Annals of Statistics. To appear.

Delgado, M. A. (1992). Testing the equality of nonparametric regression curves. Statistics and Probability Letters, 17:199–204. MR1229937

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. Annals of Statistics, 29:1361–1400. MR1873335

Donoho, D. L. (1988). One-sided inference about functionals of a density. Annals of Statistics, 16:1390–1420. MR0964930

Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. Annals of Statistics, 26:288–314. MR1611768

Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. Bernoulli, 9(3):423–449. MR1997491

Dümbgen, L. (2006). Confidence bands for convex median curves using sign-tests. MR2459932

Dümbgen, L. (2007). Confidence bands for convex median curves using sign-tests. In Cator, E., Jongbloed, G., Kraaikamp, C., Lopuhaä, R., and Wellner, J., editors, Asymptotics: Particles, Processes and Inverse Problems, volume 55 of IMS Lecture Notes - Monograph Series 55, pages 85–100. IMS, Hayward, USA. MR2459932

Dümbgen, L. and Johns, R. (2004). Confidence bands for isotonic median curves using sign-tests. J. Comput. Graph. Statist., 13(2):519–533. MR2063998

Dümbgen, L. and Kovac, A. (2009). Extensions of smoothing via taut strings. Electronic Journal of Statistics, 3:41–75. MR2471586

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. Annals of Statistics, 29(1):124–152. MR1833961

Fan, J. and Lin, S. K. (1998). Test of significance when data are curves. Journal of American Statistical Association, 93:1007–1021. MR1649196

Hall, P. and Hart, D. H. (1990). Bootstrap test for difference between means in nonparametric regression. Journal of the American Statistical Association, 85(412):1039–1049. MR1134500

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust Statistics: The Approach Based on Influence Functions. Wiley, New York. MR0829458

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. Annals of Statistics, 18(1):63–89. MR1041386

Höhenrieder, C. (2008). Nichtparametrische Volatilitäts- und Trendapproximation von Finanzdaten. PhD thesis, Department of Mathematics, University Duisburg-Essen, Germany.

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

Kabluchko, Z. (2007). Extreme-value analysis of standardized Gaussian increments.

arXiv:0706.1849.

King, E., Hart, J. D., and Wehrly, T. E. (1990). Testing the equality of two regression curves using linear smoothers. Statistics and Probability Letters, 12:239–247. MR1130364

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. Journal of the American Statistical Association, 90(431):1085–1093. MR1354025

Kulasekera, K. B. and Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. Journal of the American Statistical Association, 92(438):500–511. MR1467844

Lavergne, P. (2001). An equality test across nonparametric regressions. Journal of Econometrics, 103:307–344. MR1838202

Lehmann, E. L. (1983). Theory of Point Estimation. Wiley. MR0702834

Li, K.-C. (1989). Honset confidence regions for nonparametric regression. Annals of Satistics, 17:1001–1008. MR1015135

Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: Exact and asymptotic theory. Annals of Statistics, 26(6):2339–2368. MR1700235

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves - an empirical process approach. Annals of Statistics, 31:880–920. MR1994734

Pötscher, B. M. and Leeb, H. (2008). Sparse estimators and the oracle property, or the return of Hodges. Journal of Econometrics, 142:201–211. MR2394290

Tukey, J. W. (1993). Issues relevant to an honest account of data-based inference, partially in the light of Laurie Davies's paper. Princeton University, Princeton.