

[Luedtke, Alex](#); [Carone, Marco](#); [Van der Laan, Mark J.](#)

An omnibus non-parametric test of equality in distribution for unknown functions. (English)

[Zbl 1407.62152](#)

[J. R. Stat. Soc., Ser. B, Stat. Methodol.](#) 81, No. 1, 75-99 (2019).

Summary: We present a novel family of non-parametric omnibus tests of the hypothesis that two unknown but estimable functions are equal in distribution when applied to the observed data structure. We developed these tests, which represent a generalization of the maximum mean discrepancy tests described by Gretton and colleagues, using recent developments from the higher order pathwise differentiability literature. Despite their complex derivation, the associated test statistics can be expressed quite simply as U-statistics. We study the asymptotic behaviour of the proposed tests under the null hypothesis and under both fixed and local alternatives. We provide examples to which our tests can be applied and show that they perform well in a simulation study. As an important special case, our proposed tests can be used to determine whether an unknown function, such as the conditional average treatment effect, is equal to zero almost surely.

Reviewer: [Reviewer \(Berlin\)](#)

MSC:

[62G10](#) Nonparametric hypothesis testing

[62G20](#) Asymptotic properties of nonparametric inference

Keywords:

[equality in distribution](#); [higher order pathwise differentiability](#); [infinite dimensional parameter](#); [maximum mean discrepancy](#); [omnibus test](#); [nonparametric test](#)

Software:

[np](#); [rARPACK](#); [SuperLearner](#)

Full Text: [DOI](#)

References:

- [1] [Berlinet, A. and Thomas-Agnan, C. \(2011\) Reproducing Kernel Hilbert Spaces in Probability and Statistics. New York: Springer Science and Business Media. · \[Zbl 1145.62002\]\(#\)](#)
- [2] [Carone, M., Dáz, I. and van der Laan, M. J. \(2014\) Higher-order targeted minimum loss-based estimation. \textit{Technical Report}. Division of Biostatistics, University of California at Berkeley, Berkeley.](#)
- [3] [Chakraborty, B. and Moodie, E. E. \(2013\) Statistical Methods for Dynamic Treatment Regimes. Berlin: Springer. · \[Zbl 1278.62169\]\(#\)](#)
- [4] [Franz, C. \(2006\) Discrete approximation of integral operators. \textit{Proc. Am. Math. Soc.}, 134, 2437-2446. · \[Zbl 1119.47067\]\(#\)](#)
- [5] [Gregory, G. G. \(1977\) Large sample theory for U-statistics and tests of fit. \textit{Ann. Statist.}, 5, 110-123. · \[Zbl 0371.62033\]\(#\)](#)
- [6] [Gretton, A., Borgwardt, M. M., Rasch, M., Schölkopf, B. and Smola, A. J. \(2006\) A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems 19* \(eds B. Schölkopf, J. C. Platt and T. Hoffman\), pp. 513-520. Cambridge: MIT Press.](#)
- [7] [Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. \(2012\) A kernel two-sample test. \textit{J. Mach. Learn. Res.}, 13, 723-773. · \[Zbl 1283.62095\]\(#\)](#)
- [8] [Gretton, A., Bousquet, O., Smola, A. and Schölkopf, B. \(2005\) Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory* \(eds S. Jain, H. U. Simon and E. Tomita\), pp. 63-77. New York: Springer. · \[Zbl 1168.62354\]\(#\)](#)
- [9] [Gretton, A., Fukumizu, K., Harchaoui, Z. and Sriperumbudur, B. K. \(2009\) A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems*, pp. 673-681.](#)
- [10] [Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K. and Sriperumbudur, B. K. \(2012\) Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems* \(eds F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger\), pp. 1205-1213. Red Hook: Curran Associates.](#)
- [11] [Hayfield, T. and Racine, J. S. \(2008\) Nonparametric econometrics: the np package. \textit{J. Statist. Softwr.}, 27, 1-32.](#)

- [12] van der Laan, M. J., Polley, E. and Hubbard, A. (2007) Super Learner. *Statist. Appl. Genet. Mol.*, 6, article 25. · [Zbl 1166.62387](#)
- [13] van der Laan, M. J. and Rose, S. (2011) Targeted Learning: Causal Inference for Observational and Experimental Data. New York: Springer.
- [14] Lavergne, P., Maistre, S. and Patilea, V. (2015) A significance test for covariates in nonparametric regression. *Electron. J. Statist.*, 9, 643-678. · [Zbl 1309.62076](#)
- [15] Nolan, D. and Pollard, D. (1987) U-processes: rates of convergence. *Ann. Statist.*, 15, 780-799. · [Zbl 0624.60048](#)
- [16] Nolan, D. and Pollard, D. (1988) Functional limit theorems for U -processes. *Ann. Probab.*, 16, 1291-1298. · [Zbl 0665.60037](#)
- [17] Pfanzagl, J. (1985) Asymptotic Expansions for General Statistical Models. Berlin: Springer. · [Zbl 0578.62003](#)
- [18] Pfanzagl, J. and Wefelmeyer, W. (1982) Contributions to a General Asymptotic Statistical Theory. Berlin: Springer.
- [19] Polley, E. and van der Laan, M. J. (2013) SuperLearner: super learner prediction. University of California at Berkeley, Berkeley. (Available from <http://cran.rproject.org/package=SuperLearner>.)
- [20] Qiu, Y., Mei, J. and authors of the ARPACK Library (2014) rARPACK: R wrapper of ARPACK for large scale eigenvalue/vector problems, on both dense and sparse matrices. Purdue University, West Lafayette. (Available from <http://cran.rproject.org/package=rARPACK>.)
- [21] Racine, J. S., Hart, J. and Li, Q. (2006) Testing the significance of categorical predictor variables in nonparametric regression models. *Econometr. Rev.*, 25, 523-544. · [Zbl 1106.62046](#)
- [22] Robins, J. M. (2004) Optimal structural nested models for optimal sequential decisions. In Proc. 2nd Seattle Symp. Biostatistics (eds D. Y. Lin and P. Heagerty), pp. 189-326. New York: Springer. · [Zbl 1279.62024](#)
- [23] Robins, J. M., Li, L., Tchetgen, E. and van der Vaart, A. W. (2008) Higher order influence functions and minimax estimation of non-linear functionals. In Essays in Honor of David A. Freedman (eds D. Nolan and T. Speed), pp. 335-421. New York: Springer. · [Zbl 05556804](#)
- [24] Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance based- and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, 41, 2263-2291. · [Zbl 1281.62117](#)
- [25] Steinwart, I. (2002) On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2, 67-93. · [Zbl 1009.68143](#)
- [26] van der Vaart, A. W. (2014) Higher order tangent spaces and influence functions. *Statist. Sci.*, 29, 679-686. · [Zbl 1331.62111](#)
- [27] van der Vaart, A.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.