Nonpar MANOVA via Independence Testing

Sambit Panda[1], Ronan Perry[1], Cencheng Shen[2], Jelle Zorn[3], Antoine Lutz[3], Carey E. Priebe[4], Joshua T. Vogelstein[1*]

[1]Department of Biomedical Engineering, Johns Hopkins University, [2]Department of Applied Economics and Statistics, University of Delaware, [3]Lyon Neuroscience Research Center, INSERM U1028, CNRS UMR5292, Lyon 1 University, [4]Department of Applied Mathematics and Statistics, Johns Hopkins University

Introduction: The *k*-sample testing problem tests whether k groups of data points are sampled from the same distribution. Multivariate analysis of variance (Manova) is currently the gold standard for *k*-sample testing but makes strong, often inappropriate, parametric assumptions. Moreover, independence testing and *k*-sample testing are tightly related. There are many nonparametric multivariate independence tests with strong theoretical and empirical properties, including distance correlation (Dcorr) and Hilbert-Schmidt-Independence-Criterion (Hsic).

Goal: We prove that universally consistent independence tests achieve universally consistent *k*-sample testing, and that *k*-sample statistics like Energy and Maximum Mean Discrepancy (MMD) are exactly equivalent to Dcorr. We can extend this framework to perform multilevel and multiway tests.

Discussion: Empirically evaluating these tests for *k*-sample-scenarios demonstrates that these nonparametric independence tests typically out-perform Manova, even for Gaussian distributed settings. We thus illustrate the existence of many theoretically motivated and empirically per-formant *k*-sample tests.

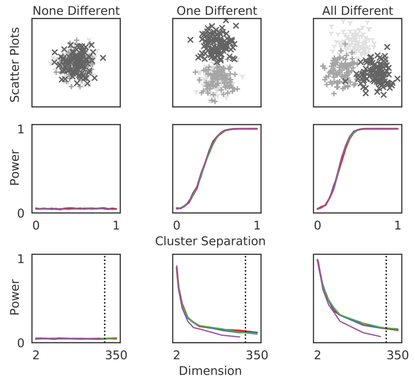

Fig 1: Power versus epsilon and dimension. The top row shows a scatter plot of each simulation for a given cluster separation. Nonpar Manova performs as well or better than Manova in all settings.

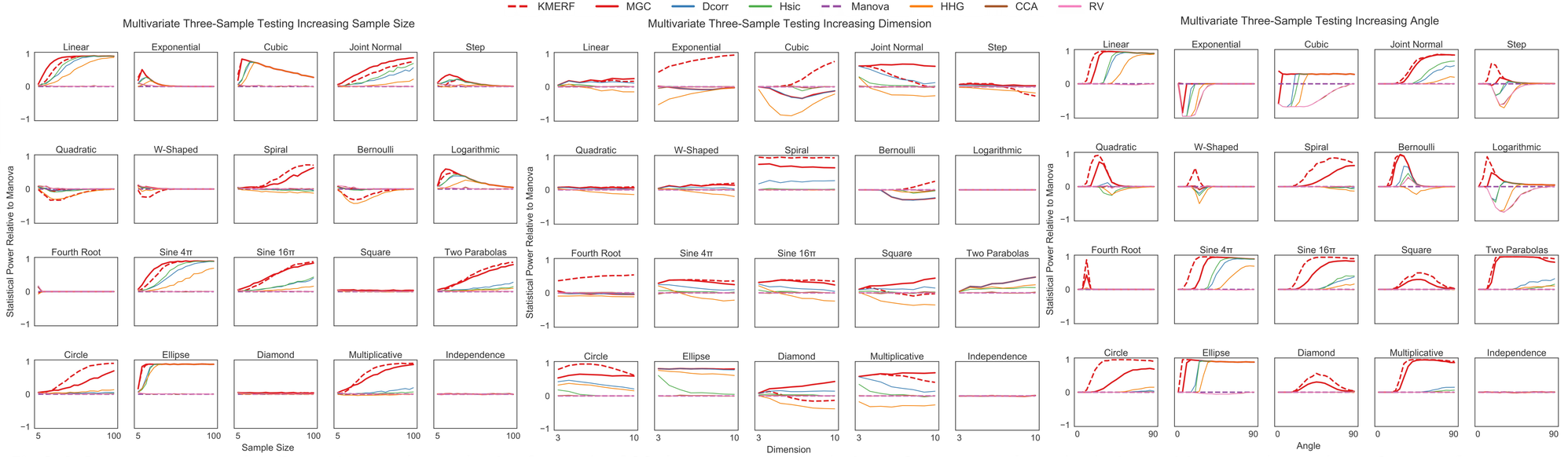

Fig 2-4: Power versus angle, dimension, and sample size for each of 20 three-sample simulations. Curves are plotted relative to Manova: those above 0 outperform Manova and those below 0 perform worse than Manova. Nonpar Manova implementations empirically dominate MANOVA in most settings.

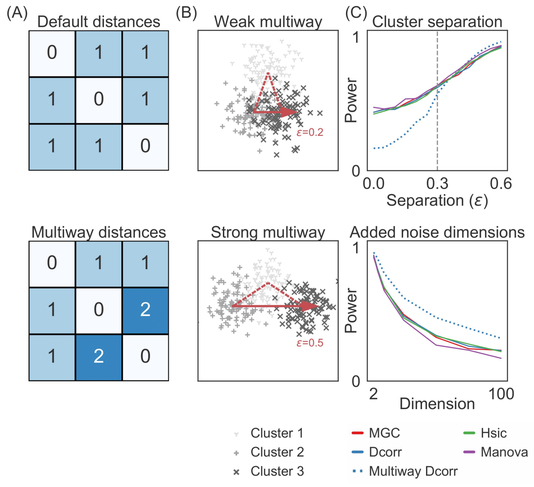

Fig 5: (A) Multiway tests are manipulated using the usual one-hot encoding for label matrices. (B) Scatter plots for multiway simulations. (C) Multiway Dcorr performs worse than the other tests at low epsilon. It performs as good or better than other tests at all other settings.

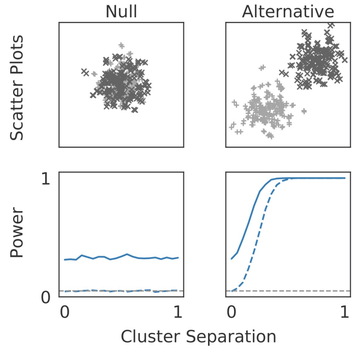

Fig 6: 100 means were sampled from each of two, two-dimensional Gaussians. Two samples were generated from Gaussians centered at each mean and with lower variance. Only multilevel Dcorr is. Dcorr is invalid and its greater power under the alternative is an artifact of its invalidity.

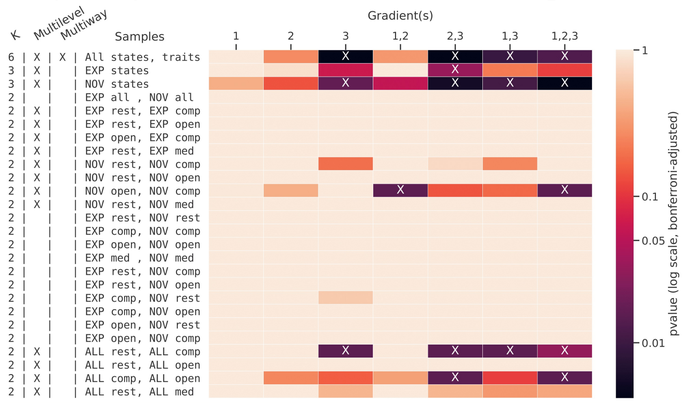

Fig 7: Omnibus and pairwise tests on combinations of states (open, rest, comp) and traits (NOV, EXP) reveal significant corrected p-values (denoted with a white X) at 0.05.