Data from: Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations

Adams, Dean C., Iowa State University

Collyer, Michael L., Chatham University

Citation

Adams, Dean C.; Collyer, Michael L. (2017), Data from: Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations, Dataset, https://doi.org/10.5061/dryad.29722

Abstract

Recent years have seen increased interest in phylogenetic comparative analyses of multivariate datasets, but to date the varied proposed approaches have not been extensively examined. Here we review the mathematical properties required of any multivariate method, and specifically evaluate existing multivariate phylogenetic comparative methods in this context. Phylogenetic comparative methods based on the full multivariate likelihood are robust to levels of covariation among trait dimensions and are insensitive to the orientation of the dataset, but display increasing model misspecification as the number of trait dimensions increases. This is because the expected evolutionary covariance matrix (V) used in the likelihood calculations becomes more ill-conditioned as trait dimensionality increases, and as evolutionary models become more complex. Thus, these approaches are only appropriate for datasets with few traits and many species. Methods that summarize patterns across trait dimensions treated separately (e.g., SURFACE) incorrectly assume independence among trait dimensions, resulting in nearly a 100% model misspecification rate. Methods using pairwise composite likelihood are highly sensitive to levels of trait covariation, the orientation of the dataset, and the number of trait dimensions. The consequences of these debilitating deficiencies is that a user can arrive at differing statistical conclusions, and therefore biological inferences, simply from a dataspace rotation, like principal component analysis. By contrast, algebraic generalizations of the standard phylogenetic comparative toolkit that use the trace of covariance matrices are insensitive to levels of trait covariation, the number of trait dimensions, and the orientation of the dataset. Further, when appropriate permutation tests are used, these approaches display acceptable Type I error and statistical power. We conclude that methods summarizing information across trait dimensions, as well as pairwise composite likelihood methods should be avoided, while algebraic generalizations of the

phylogenetic comparative toolkit provide a useful means of assessing macroevolutionary patterns in multivariate data. Finally, we discuss areas in which multivariate phylogenetic comparative methods are still in need of future development; namely highly multivariate Ornstein-Uhlenbeck models and approaches for multivariate evolutionary model comparisons.

Usage Notes

Supplemental Material
Additional results from simulations supporting this manuscript
SuppMaterial-AdamsCollyer.docx

Simulation-RScripts
R scripts used to perform all simulations in this study

Simulation-Results
Results from all simulations in this study

Funding

National Science Foundation, Award: DEB-1556379

References

This dataset is supplement to https://doi.org/10.1093/sysbio/syx055

Keywords

high-dimensional data, multivariate, phylogenetic comparative methods

Files

4 files for this dataset

File	Size	Type
README_for_SuppMa...damsCollyer.docx	67.28 kB	application/vnd.openxmlformats-officedocument.wordprocessingml.document
Simulation-RScripts.zip	1.85 MB	application/zip
Simulation-Results.zip	399.92 kB	application/zip
SuppMaterial-AdamsCollyer.docx	67.28 kB	application/vnd.openxmlformats-officedocument.wordprocessingml.document

License

This releases your work to the public domain for any use.