Robust regression with compositional covariates.  (English)  Zbl 07422846

Summary: Many biological high-throughput datasets, such as targeted amplicon-based and metagenomic sequencing data, are compositional. A common exploratory data analysis task is to infer robust statistical associations between high-dimensional microbial compositions and habitat- or host-related covariates. To address this, a general robust statistical regression framework RobRegCC (Robust Regression with Compositional Covariates) is proposed, which extends the linear log-contrast model by a mean shift formulation for capturing outliers. RobRegCC includes sparsity-promoting convex and non-convex penalties for parsimonious model estimation, a data-driven robust initialization procedure, and a novel robust cross-validation model selection scheme. The procedure is implemented in the R package robregcc. Extensive simulation studies show the RobRegCC’s ability to perform simultaneous sparse log-contrast regression and outlier detection over a wide range of settings. To demonstrate the seamless applicability of the workflow to real data, the gut microbiome dataset from HIV patients are analyzed and robust associations between a sparse set of microbial species and host immune response from soluble CD14 measurements are inferred.

MSC:
62-XX  Statistics

Keywords:
compositional data; robust; mean shift; sparsity; microbiome

Software:
robustbase; UNLocBoX; RobStatTM; R

Full Text: DOI arXiv

References:


[38] Rocafort, M.; Noguera-Julian, M.; Rivera, J.; Pastor, L.; Guillén, Y.; Langhorst, J.; Parera, M.; Mandomando, I.; Carrillo, J.; Ureña, E., Evolution of the gut microbiome following acute hiv-1 infection, Microbiome, 7, 73 (2019)

[39] Rousseau, P.; Yohai, V., Robust regression by means of s-estimators, (Robust and Nonlinear Time Series Analysis (1984), Heidelberg Academy of Sciences and Humanities · Zbl 07263531


[42] She, Y., Selective factor extraction in high dimensions, Biometrika, 104, 97-110 (2017) · Zbl 07072184

[43] She, Y.; Chen, K., Robust reduced-rank regression, Biometrika, 104, 633-647 (2017) · Zbl 07072232

Edited by FIZ Karlsruhe, the European Mathematical Society and the Heidelberg Academy of Sciences and Humanities © 2022 FIZ Karlsruhe GmbH