Supplementary Materials

Quang P. Nguyen

September 24, 2020
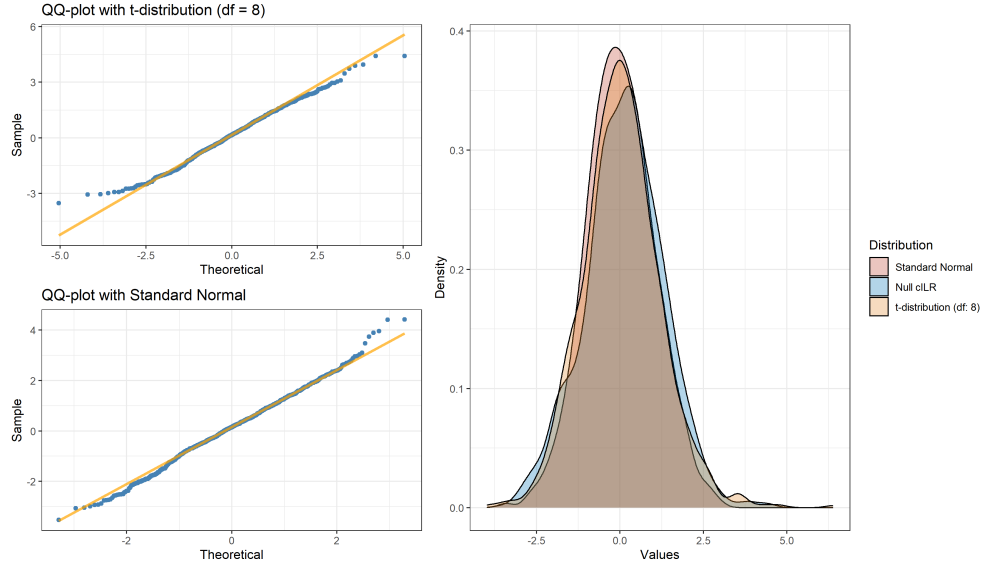
1 Distribution of cILR

Figure 1: The distribution of cILR statistic under the null. We compared the null distribution of the test statistic and compare it with the standard normal distribution, and the t-distribution with degrees of freedom fitted to cILR scores using the maximum likelihood method

2 Simulation Design

We simulated microbiome relative abundance data using the NorTA similar to [1]. Using this method, we can generate synthetic microbial counts that incorporates a complex correlation structure and multiple types of marginals. Specifically, we chose our marginals to be the zero-inflated negative binomial distribution based on results by Kurtz et al. [2]

We fit the the parameters to our marginal model based on 16S rRNA sequencing of the V3-V5 region from stool samples in the Human Microbiome Project (HMP). This data was accquired via the *HMP16S* package in R [3].
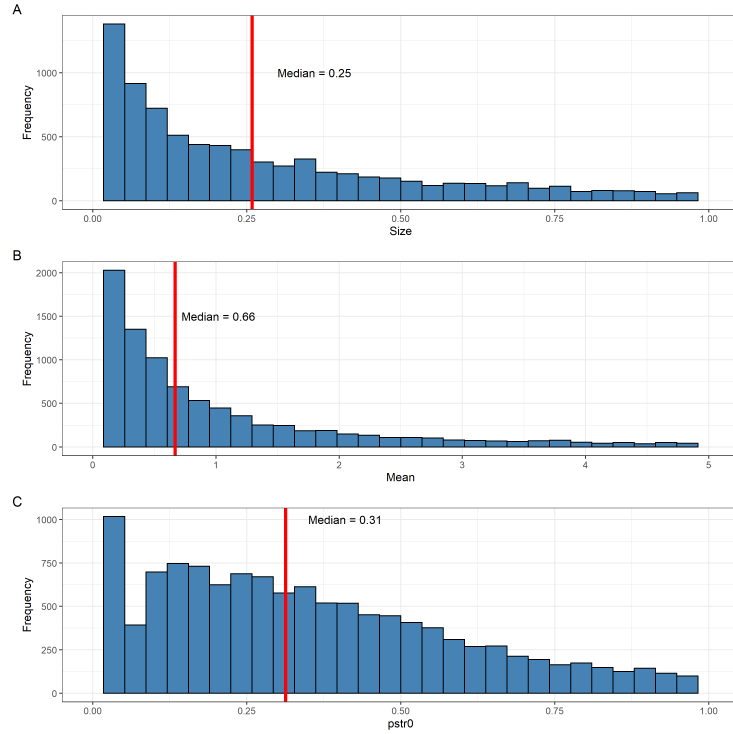
Figure 2: Distribution of each parameter of the zero inflated negative binomial distribution fitted to HMP16S data. The parameters are size (panel A), mean (panel B) and probability of 0 (panel C)

References

[1] Marne C Cario. Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix. page 19.

[2] Zachary D. Kurtz, Christian L. Müller, Emily R. Miraldi, Dan R. Littman, Martin J. Blaser, and Richard A. Bonneau. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Computational Biology*, 11(5):e1004226, May 2015.

[3] Lucas Schiffer, Rimsha Azhar, Lori Shepherd, Marcel Ramos, Ludwig Geistlinger, Curtis Huttenhower, Jennifer B Dowd, Nicola Segata, and Levi Waldron. HMP16SData: Efficient access to the human microbiome project through bioconductor. *American Journal of Epidemiology*, 2019.