An Adaptive Association Test for Microbiome Data

Chong Wu[1], Jun Chen[2], Junghi[1] Kim and Wei Pan[1]

[1] Division of Biostatistics, School of Public Health, University of Minnesota;
[2] Division of Biomedical Statistics and Informatics, Mayo Clinic

Aug. 4th, 2016

UNIVERSITY OF MINNESOTA

Table of Contents

Human Microbiome Data

- Microbe: Tiny living organism, such as bacterium, fungus, or virus
- Microbiome: The genomes of human microbes; the way they interact with the human host
- Why human microbiome is important?
  - More than 10 times the number of microbes lives in the human body than cells.
  - Play an important part in our overall health.

Human microbiome association studies

- Seen as an ”extended” human genome
- Detect an association of the human microbiome diversity with a phenotype of interest
- Improve our understanding of the non-genetic component of complex traits and diseases

Goal

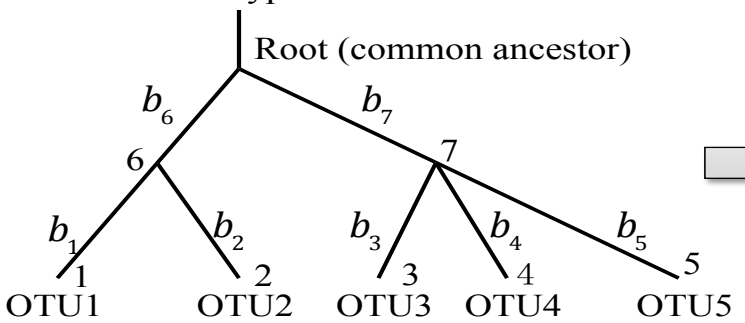
Testing the association between the whole microbiome composition and the outcome of interest

The Feature of The Human Microbiome Data

- Operational taxonomic units (OTUs): surrogates for biological taxa
  - High dimensional: the number of OTUs are usually much larger than the sample size
  - Overdispersion: most OTUs are rare
- Phylogenetic Tree
  - A branching diagram or ”tree” showing the inferred evolutionary relationships among various biological species

Outline

Null hypothesis: No association between any taxa and the outcome



Sample \ OTU	1	2	3	4	5
1	0	0	2	0	1
2	2	5	0	0	1
3	0	0	0	1	0
4	1	0	0	0	3

Taxon proportion: $p_{ik} = Z_{ik} \Big/ \sum_{j=1}^{q} Z_{ij}$ $p_{13} = 2/(1+2) \approx 0.7$ $p_{26} = (2+5)/(2+5+1) \approx 0.9$

Generalized taxon proportion: $Q_{ik}^u = b_k I\left(p_{ik} > 0\right)$; $Q_{ik}^w = b_k p_{ik}$ e.g. $Q_{13}^u = b_3$; $Q_{13}^w = 0.7 b_3$

Test Statistics: $U = \sum_{i=1}^{n} \left(Y_i - \hat{\mu}_{i,0}\right) Q_{i\cdot}$; $T_{\text{MiSPU}(\gamma)} = \sum_{k=1}^{m} U_k^{\gamma}$; $T_{a\text{MiSPU}_u} = \min_{\gamma \in \Gamma} P_{\text{MiSPU}(\gamma)}$

Residual Permutation

P value

MiSPU

- Generalized taxon proportion Q_{ik}:

$$Q_{ik}^{w} = b_k p_{ki}; \quad Q_{ik}^{u} = b_k I(p_{ki})$$

- Raw weighted UniFrac distance is exactly the same as the L_1 distance of the Q_{ik}^{w}
- For a binary outcome, we use a logistic regression model:

$$\text{Logit}[Pr(Y_i = 1)] = \beta_0 + \beta' X_i + \sum_{k=1}^{m} Q_{ik}\varphi_k$$

- $H_0 : \varphi = (\varphi_1, \ldots, \varphi_m)' = 0$; that is, there is no association between any taxa and the outcome of interest

MiSPU

- Score:

$$U = \sum_{i=1}^{n} (Y_i - \hat{\mu}_{i,0}) Q_i.$$

- MiSPU test statistic:

$$T_{\text{MiSPU}(\gamma)} = w'U = \sum_{k=1}^{m} U_k^{\gamma}$$

- Use permutation scheme (Pan et al., 2014) to calculate the p value

The choice of γ

- As γ goes to infinity, we have

$$T_{\text{MiSPU}(\infty)} \propto ||U||_\infty = \max_{k=1}^{m} |U_k|$$

- Intuition in the choice of γ:
  - the more sparse the signals, the larger γ
  - if (most) associations in one direction, the use an odd γ
- In practice, how to choose γ?

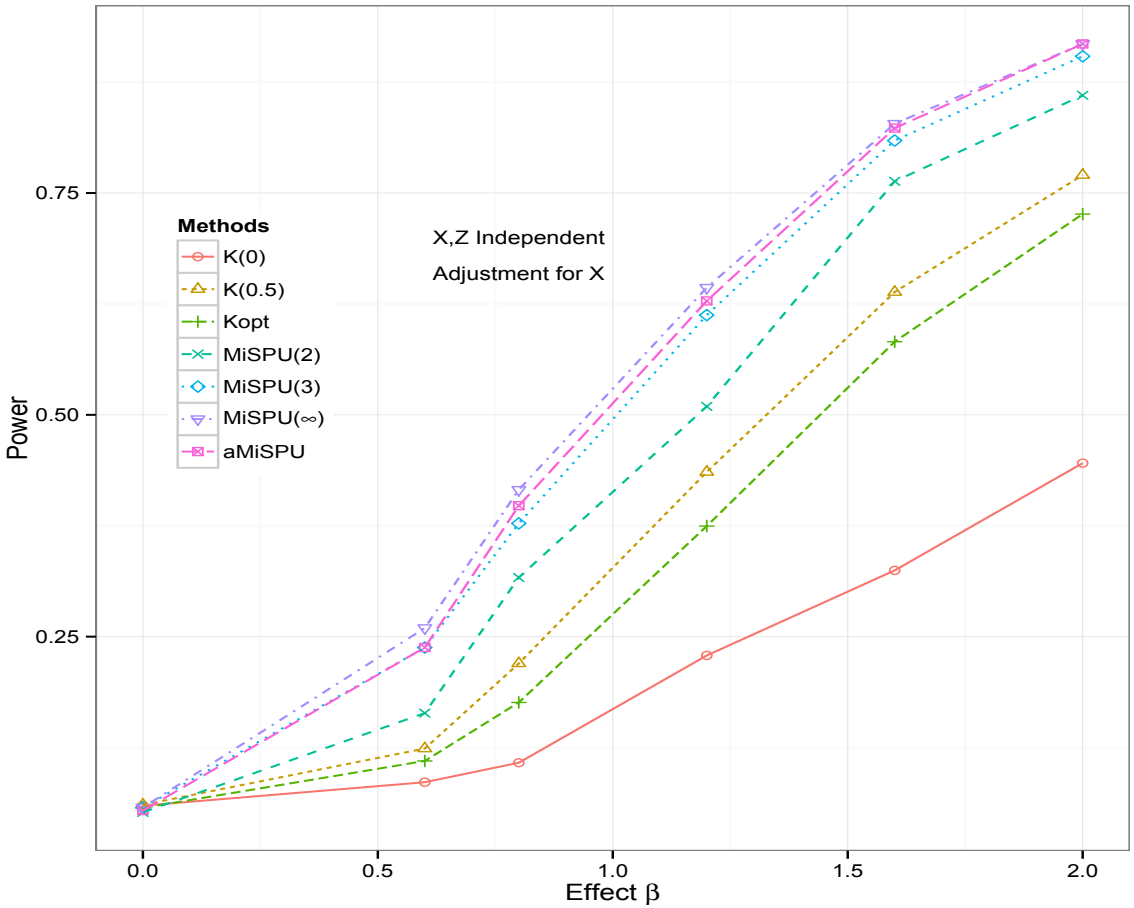
aMiSPU

- Choose the one giving the most significant p value
- Use an adaptive test idea (Pan et al., 2014)
- aMiSPU test statistic:

$$T_{\text{aMiSPU}} = \min_{\gamma \in \Gamma} P_{\text{MiSPU}(\gamma)}.$$

- Use permutation or parametric bootstrap to estimate its p value

Simulation Results



Real Data: Gut Microbiome

- Diet strongly affects human health, partly by modulating gut microbiome composition
- In one cross-sectional study, 98 healthy volunteers were enrolled and habitual long-term diet information was collected using food frequency questionnaire
- Original study failed to detect the gender effect (p value **0.080**)
- Increasing evidence suggests that there is sex difference in the human gut microbiome (Bolnick et al., 2014)
- Our new method can detect it (p value **0.0058**)

Real Data: Gut Microbiome

- A taxon in *Bacteroides* explains more than 90% relative contributions
- Top 4 taxa all come from the *Bacteroides*
- Gender status is likely associated with *Bacteroides*, but independent with other enterotypes