Null Models

Null Models

- If X is responsible for coexistence in communities, then species should be *more* dissimilar *than expected*

- Why do we care are what is expected?

- How do we calculate an expectation?

Null Models

- Why do we care are what is expected?
  - In general, in statistics we measure something for a sample (e.g. a mean), but we do not know whether that value is any different than what is expected
  - In other words we don't know if our descriptive statistic for a sample is "significant" or how it deviates from expected or another sample
  - Thus we compare our result to an expected distribution (e.g. a *t* distribution)

Null Models

- Why do we care are what is expected?
  - In the case of phylogenetic or functional diversity analyses we care because species diversity is often correlated with phylogenetic or functional diversity. Therefore we cannot state whether or not there is a unusual level (high or low) in the community.
  - It may just be low because there are more or less species not because any biological process relating to phylogenetic relatedness or functional traits

Null Models

- ## How do we calculate an expectation?

  - In frequentist statistics we use things like a *t* distribution for our expectation.

  - In many cases there is no expected distribution available and we must generate one.

  - Null models randomize the observed data to generate a null distribution

# A General Philosophy for Null Models

- Keep everything in the observed data constant *except* the one thing you are interested in

- In other words do not randomize everything (this inflates type I error)

- This is easy to say, but can be more difficult to do.

A Phylogenetic Null Model

- Let us say we are trying to solve Cam Webb's problem
    - Is the MPD in a community higher or lower than expected
    - Cam did this is a simple way by comparing the observed to the maximum possible MPD given the data
        - This approach suffers from several potential biases

Phylogeny

Community Phylogeny
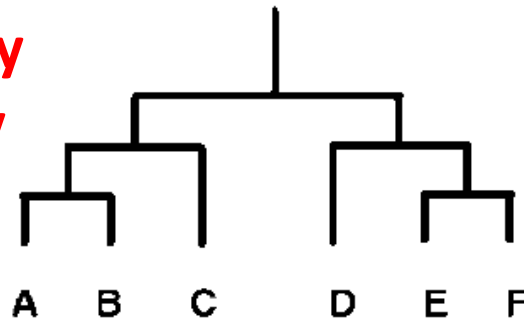
Greatest possible mean *pairwise* nodal distance for a community of 4 taxa (given this phylogeny) = 3.66 nodes (for A, B, E, F)

Greatest possible mean *nearest* nodal distance for a community of 4 taxa (given this phylogeny) = 2.00 nodes (for A, C, D, F)

Community 1: A, B, C, D

Community 2: A, B, E, F

Phylogenetic Distance Matrix

Nodal distances:

	A	B	C	D
A		1	2	4
B			2	4
C				3

Nodal distances:

	A	B	E	F
A		1	5	5
B			5	5
E				1

Mean *pairwise* nodal distance =
$(1 + 2 + 4 + 2 + 4 + 3) / 6$
$= 2.66$

MPD

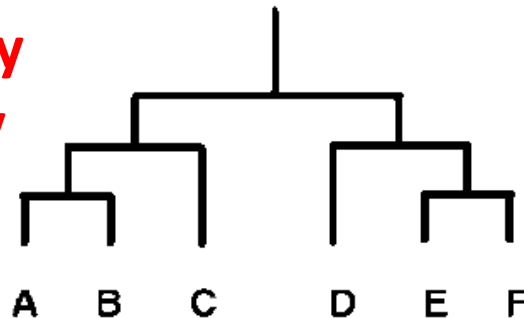Mean *pairwise* nodal distance =
$(1 + 5 + 5 + 5 + 5 + 1) / 6$
$= 3.66$

Net Relatedness Index =
$1 - (2.66 / 3.66) = 0.273$

NRI aka SES.MPD

Net Relatedness Index =
$1 - (3.66 / 3.66) = 0.0$

Phylogeny

Community Phylogeny

Greatest possible mean *pairwise* nodal distance for a community of 4 taxa (given this phylogeny) = 3.66 nodes (for A, B, E, F)

Greatest possible mean *nearest* nodal distance for a community of 4 taxa (given this phylogeny) = 2.00 nodes (for A, C, D, F)

Phylogenetic Distance Matrix

Community 1: A, B, C, D

Nodal distances:

	A	B	C	D
A		1	2	4
B			2	4
C				3

Community 2: A, B, E, F

Nodal distances:

	A	B	E	F
A		1	5	5
B			5	5
E				1

MNTD

Mean *nearest* nodal distance = (1 + 1 + 2 + 3) / 4 = 1.75

Mean *nearest* nodal distance = (1 + 1 + 1 + 1) / 4 = 1.0

NTI aka SES.MNTD

Nearest Taxa Index = 1 - (1.75 / 2.0) = 0.125

**Nearest Taxa Index =** 1 - (1.0 / 2.0) = 0.5
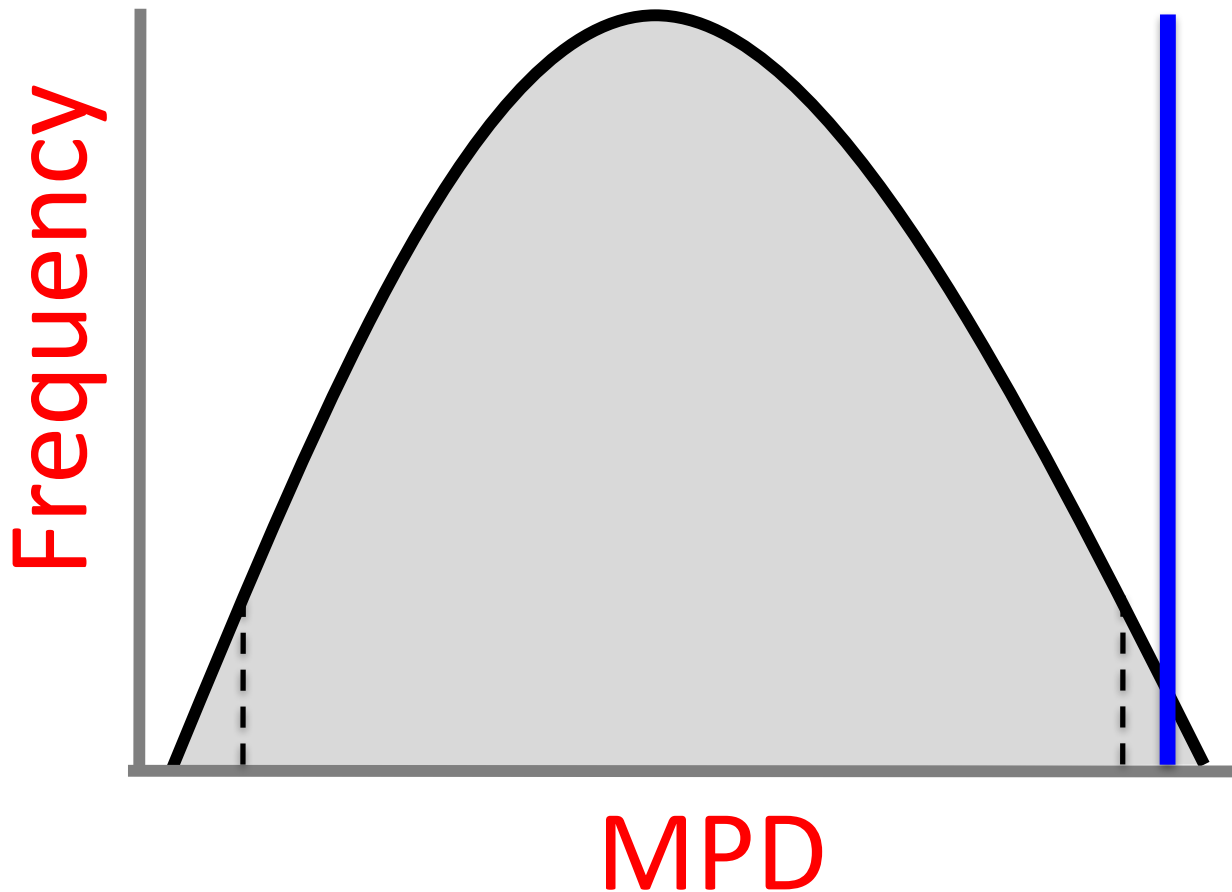
A Phylogenetic Null Model

- Let us say we are trying to solve Cam Webb's problem
    - Is the MPD in a community higher or lower than expected
    - Cam did this is a simple way by comparing the observed to the maximum possible MPD given the data
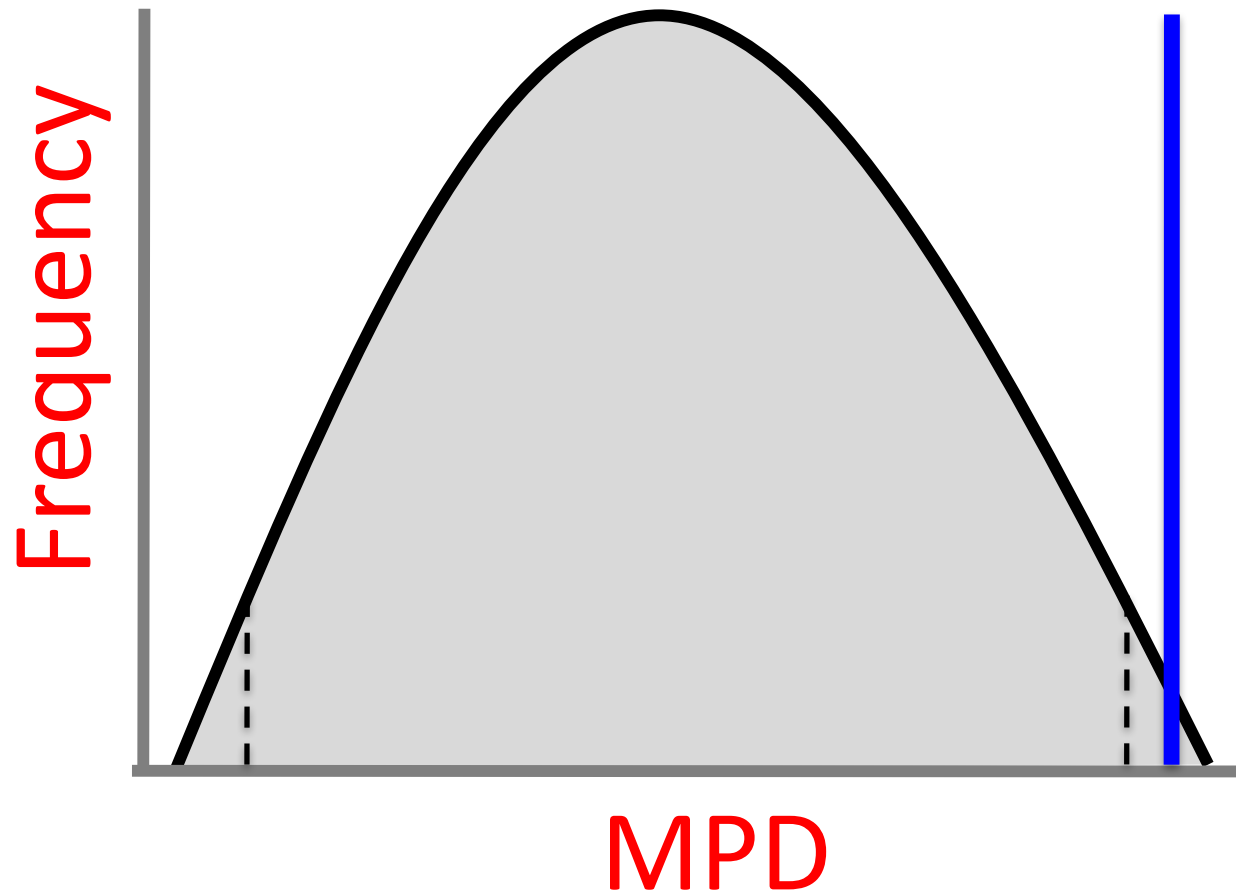   - An alternative approach would be to use a null model

A NULL MODEL

	Observed	Random	
A		H	A
B		C	H
C		G	C
D		B	B
E		E	F
F		F	G
G		D	H
H		A	D
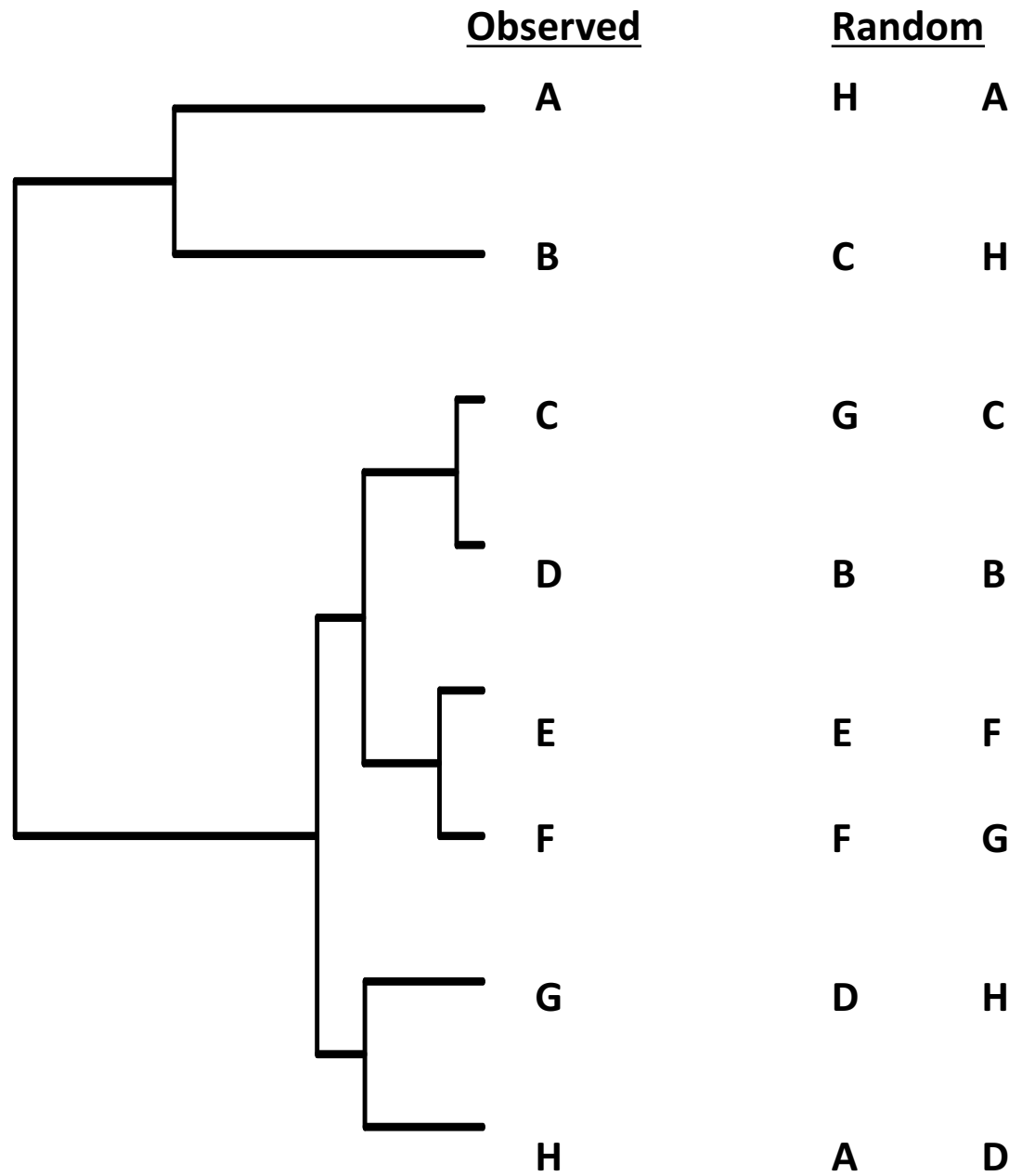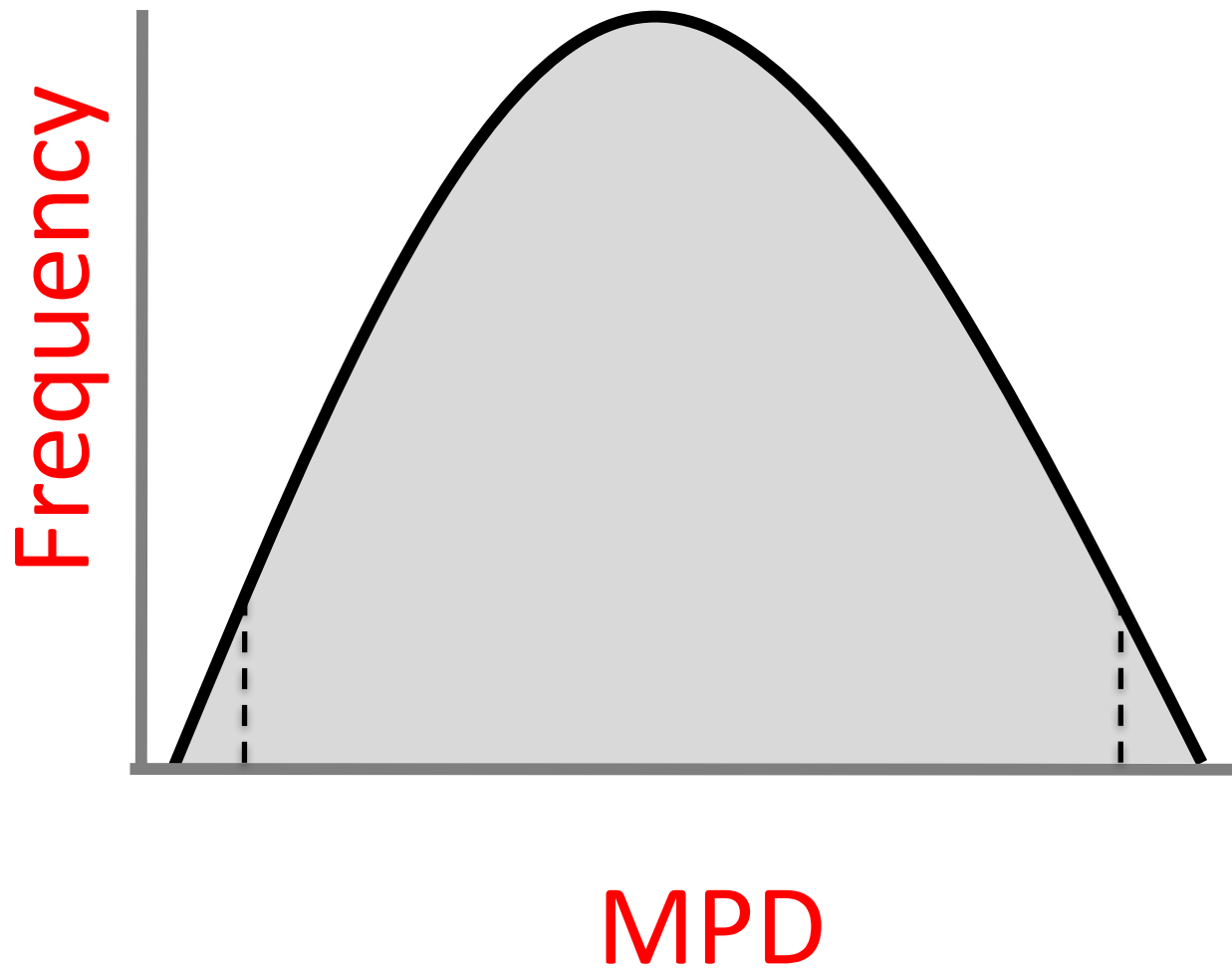
Standardized Effect Size = (Observed MPD − Mean of null distribution)/Standard deviation of null deviation

P-value = calculated directly from null distribution

What Other Null Models Could I Do?

- Many!
- The question you must always ask yourself though is what is really being randomized and what is being constrained and you must be able to explain that in your Methods section.
- A null that randomizes the names of species on a phylogeny randomizes the relatedness of species, but it does not randomize community data (species richness, abundance distributions, etc)

Randomizing Community Data – The Independent Swap Algorithm

- ## Community Data Matrix (CDM):
  - Rows = communities
  - Columns = species
  - Cells = pres/abs

- ## An Independent Swap randomizes the CDM with 2 rules/constraints:
  - All row sums must be maintained
    - Therefore the observed species richness of a community is fixed
  - All column sums must be maintained
    - Therefore the observed frequency of a species across communities is fixed

Matrix A

$$\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{matrix}$$

Matrix B

$$\begin{matrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$$

Matrix C

$$\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{matrix}$$