

# **最尤系統樹推定 と系統樹の信頼性評価**



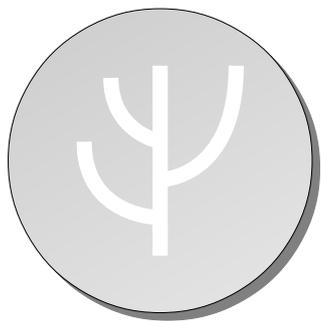
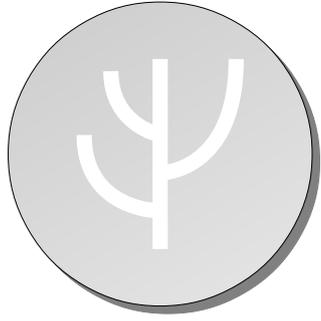
最尤法

って何？

最大尤度法

って何？

# コイントスの例





コイントスで裏が出る確率は  $X$  である

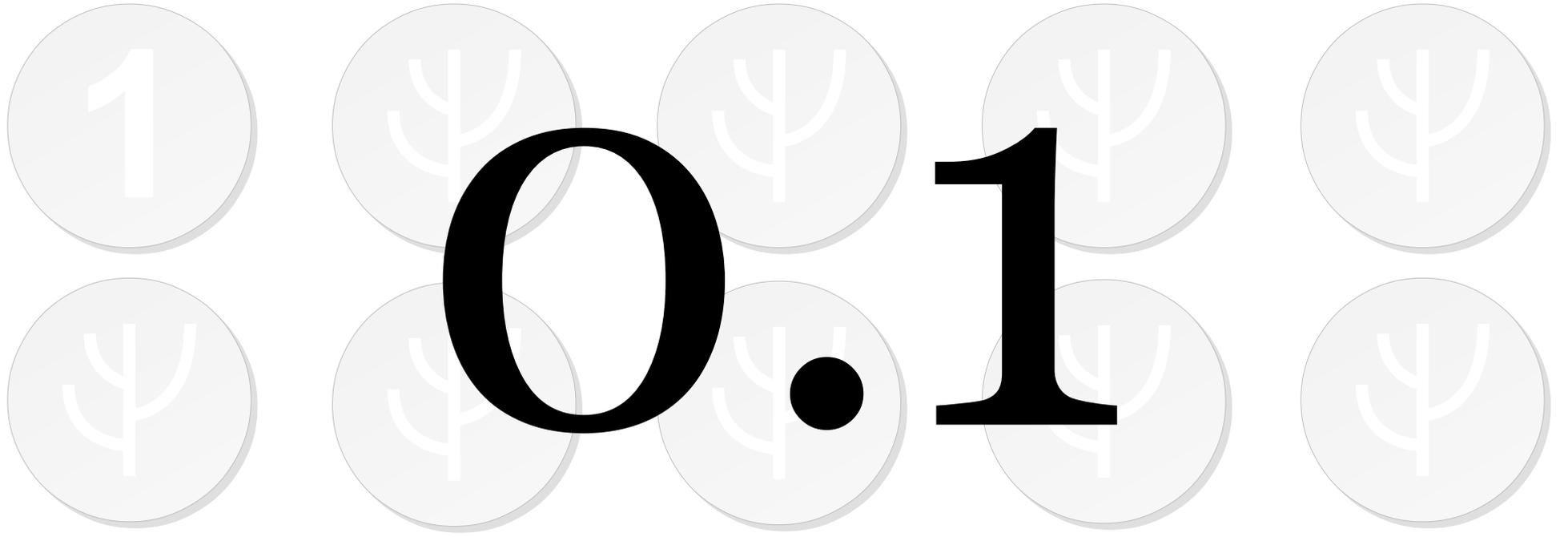


コイントスで裏が出る確率は  $X$  である

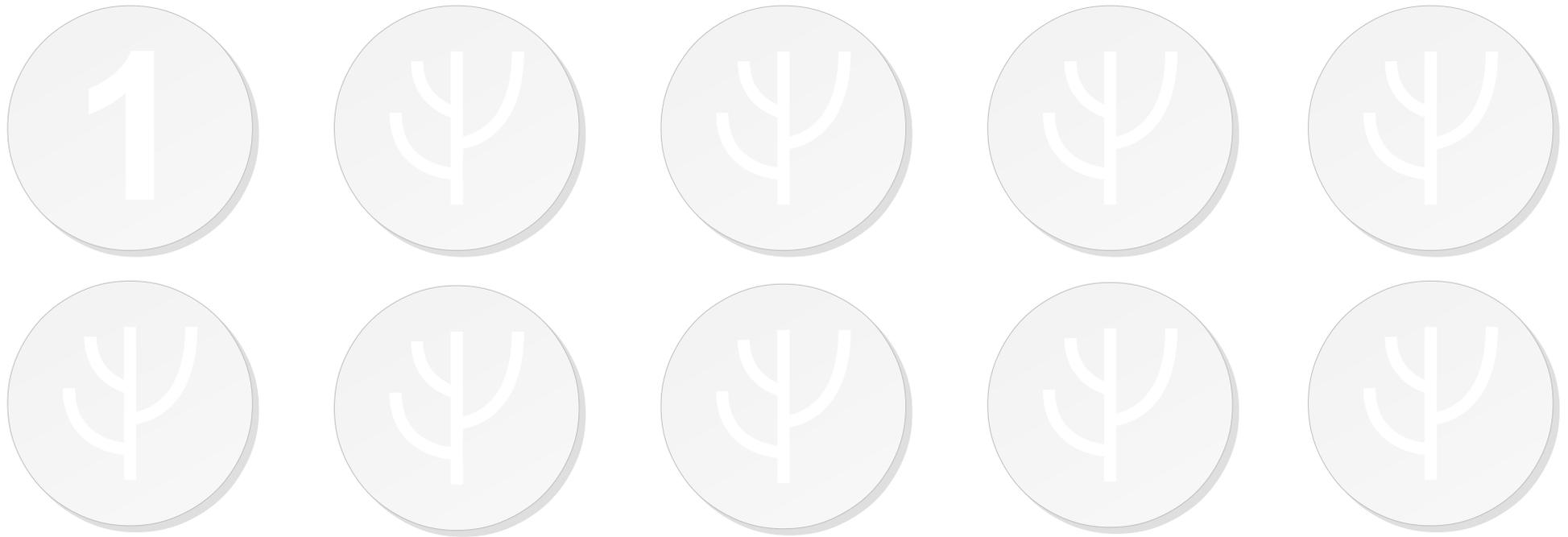


**$X$**

はいくつ？



# 10 回のコイントスで裏 1 回表 9 回が出る確率



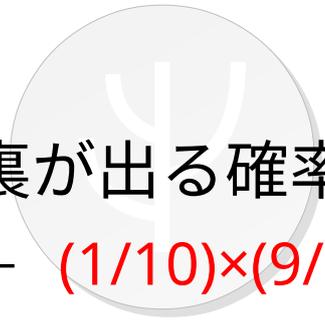
# 10 回のコイントスで裏 1 回表 9 回が出る確率



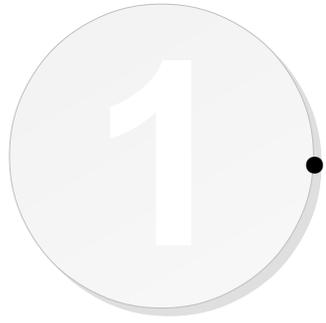
# 10 回のコイントスで裏 1 回表 9 回が出る確率



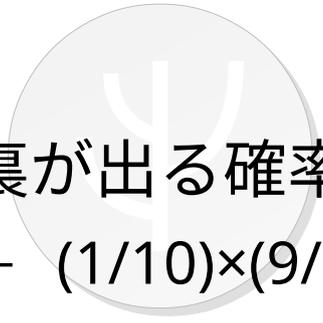
- 裏が出る確率は 0.1 であるというモデル  
-  $(1/10) \times (9/10)^9$



# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル  
-  $(1/10) \times (9/10)^9 = 0.03874205$



# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル  
-  $(1/10) \times (9/10)^9 = 0.03874205$



- 裏が出る確率は 0.5 であるというモデル



# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル
  - $(1/10) \times (9/10)^9 = 0.03874205$



- 裏が出る確率は 0.5 であるというモデル
  - $(1/2)^{10}$

# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル
  - $(1/10) \times (9/10)^9 = 0.03874205$



- 裏が出る確率は 0.5 であるというモデル
  - $(1/2)^{10} = 0.0009765625$

# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル
  - $(1/10) \times (9/10)^9 = \underline{\mathbf{0.03874205}}$



- 裏が出る確率は 0.5 であるというモデル
  - $(1/2)^{10} = \underline{\mathbf{0.0009765625}}$

# 10 回のコイントスで裏 1 回表 9 回が出る確率



- 裏が出る確率は 0.1 であるというモデル
  - $(1/10) \times (9/10)^9 = \underline{\mathbf{0.03874205}}$



- 裏が出る確率は 0.5 であるというモデル
  - $(1/2)^{10} = \underline{\mathbf{0.0009765625}}$

# 尤度

あるモデルを仮定したとき、  
データが実現する確率



あるモデルを仮定したとき、  
データが実現する確率

||

尤度

尤度が最大になるように  
パラメータ値を決める

尤度が最大になるように  
パラメータ値を決める

||

**最尤法**

# 系統樹推定 の場合

**樹形間の比較では分子進化モデルは固定**  
(パラメータ値は推定)

樹形間の比較では分子進化モデルは固定  
(パラメータ値は推定)

||

パラメータ数 ( $K$ ) は全樹形で等しい

樹形間の比較では分子進化モデルは固定  
(パラメータ値は推定)

||

パラメータ数 ( $K$ ) は全樹形で等しい

||

最大尤度樹形は AIC も最小

# 最尤系統樹推定

# 最尤系統樹推定

||

尤度が最大になる系統樹を  
採用する



問問

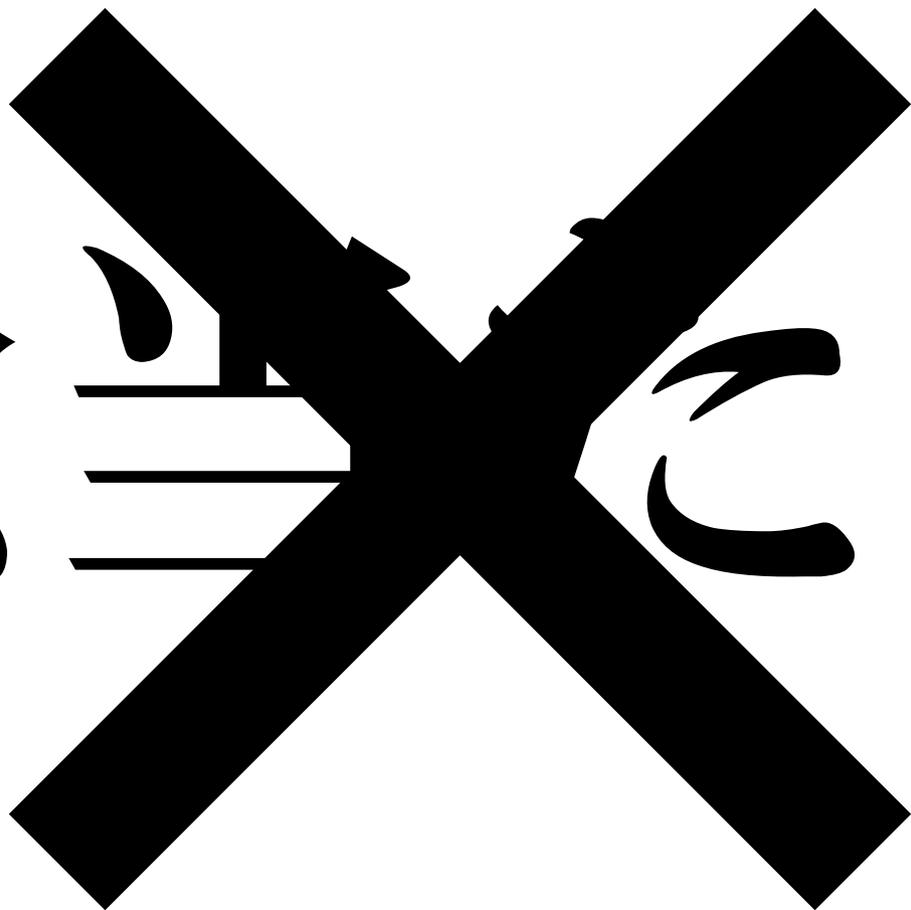
題題

尤度最大な樹形

を どう決める？

総当たり

総  
心



心

こ

り

魚 理 理

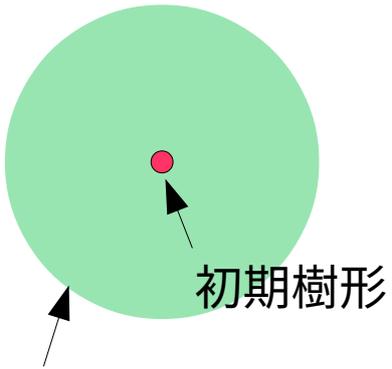
そこで、

局所探索



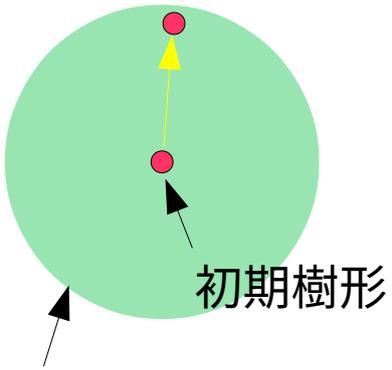


初期樹形



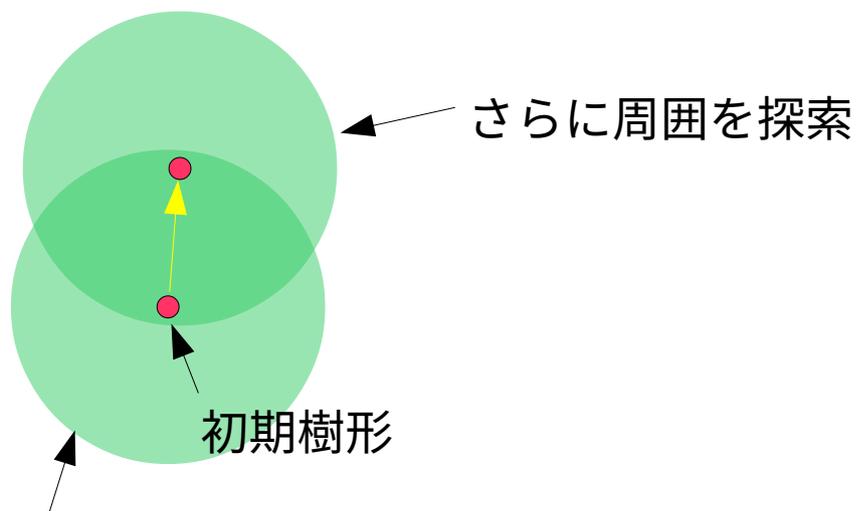
初期樹形

樹形改變によって探索される範圍

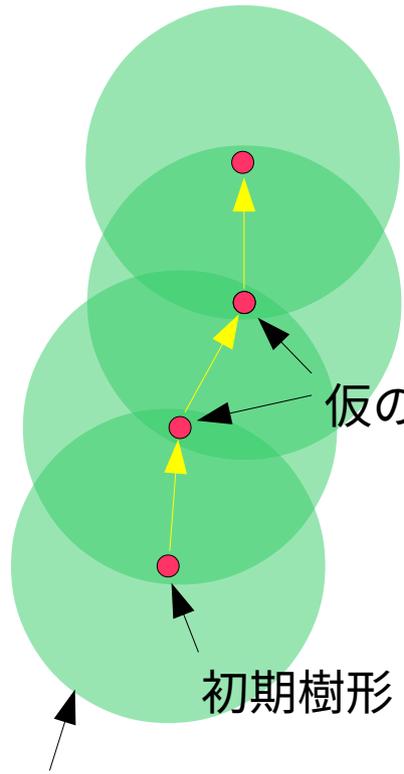


初期樹形

樹形改變によって探索される範囲



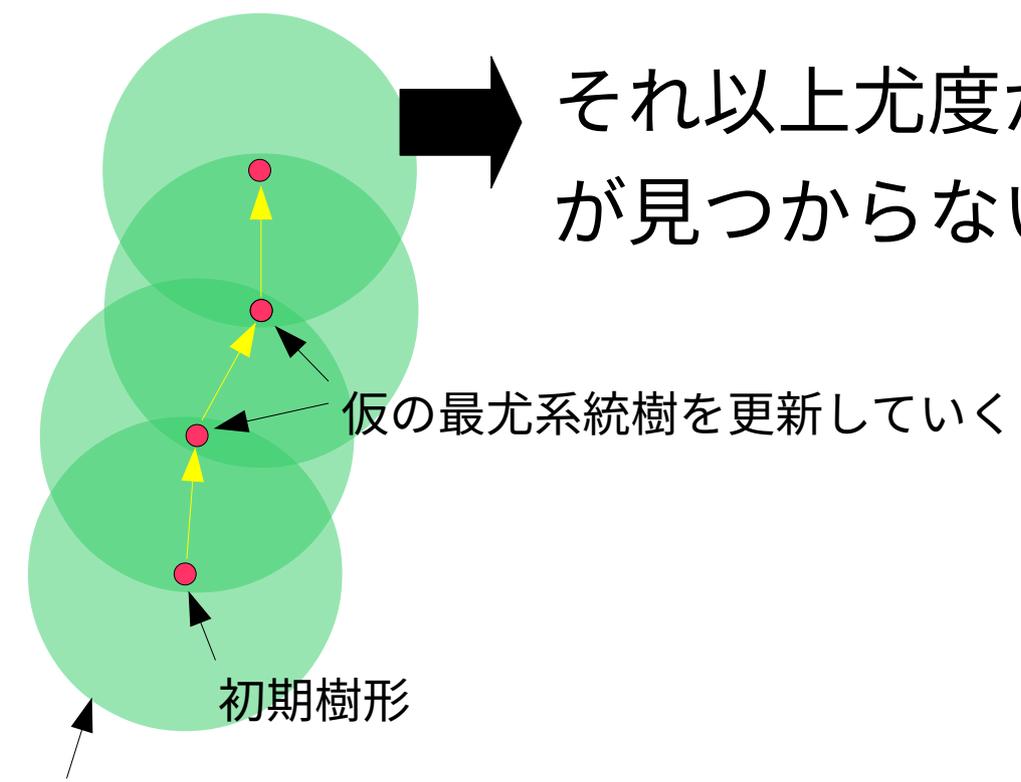
樹形改変によって探索される範囲



仮の最尤系統樹を更新していく

初期樹形

樹形変化によって探索される範囲

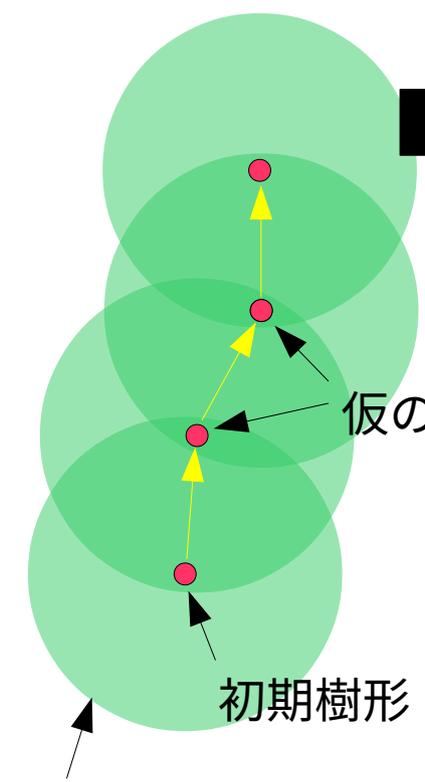


それ以上尤度が高いもの  
が見つからない

仮の最尤系統樹を更新していく

初期樹形

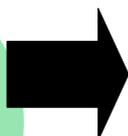
樹形改変によって探索される範囲



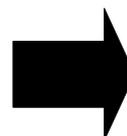
初期樹形

仮の最尤系統樹を更新していく

樹形改変によって探索される範囲



それ以上尤度が高いもの  
が見つからない



終了

# 局所探索

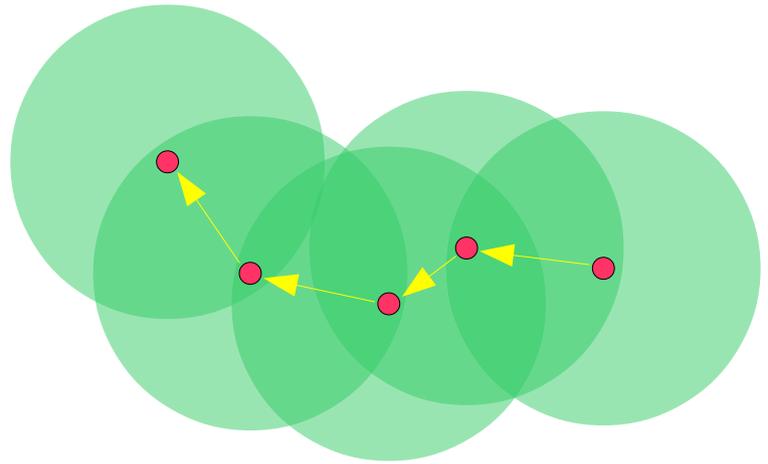
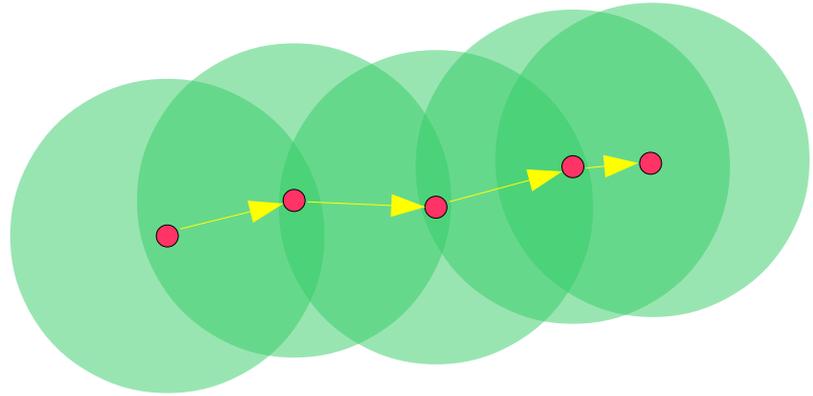
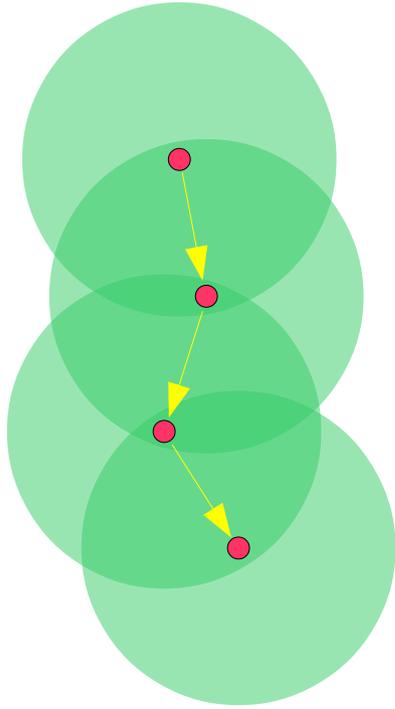
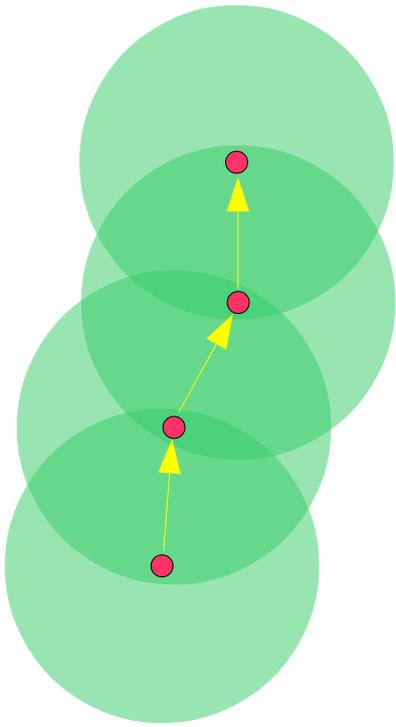
って不安だなあ

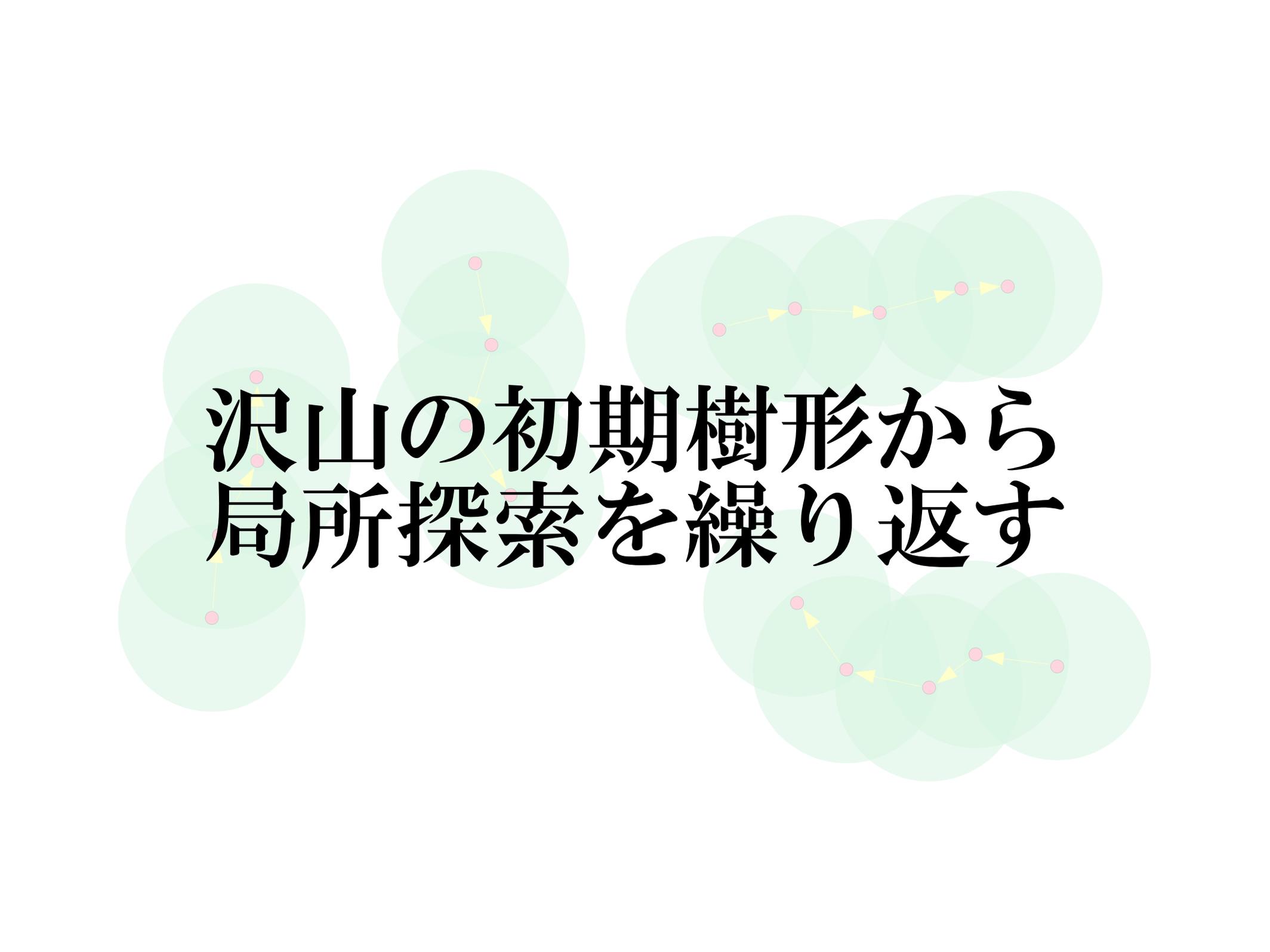
・ ・ ・

気休

め





The background features several overlapping light green circles. Within these circles, there are small diagrams showing paths of pink dots connected by yellow arrows. One path is a simple line, another is a zig-zag, and a third is a more complex path with multiple turns.

# 沢山の初期樹形から 局所探索を繰り返す

# 無作為配列付加による初期樹形生成

# 無作為配列付加による初期樹形生成

- 無作為に選んだ 3OTU で系統樹生成 ( 樹形 1 つ )

# 無作為配列付加による初期樹形生成

- 無作為に選んだ 30TU で系統樹生成 ( 樹形 1 つ )
- 無作為に選んだ 10TU の枝を最適な場所へ付ける

# 無作為配列付加による初期樹形生成

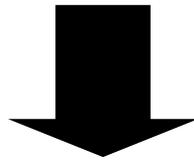
- 無作為に選んだ 30TU で系統樹生成 ( 樹形 1 つ )
- 無作為に選んだ 10TU の枝を最適な場所へ付ける
- さらに 10TU 足す、を繰り返す

# 無作為配列付加による初期樹形生成

- 無作為に選んだ 3OTU で系統樹生成 ( 樹形 1 つ )
- 無作為に選んだ 1OTU の枝を最適な場所へ付ける
- さらに 1OTU 足す、を繰り返す
- 全 OTU の系統樹を得る

# 無作為配列付加による初期樹形生成

- 無作為に選んだ 3OTU で系統樹生成 ( 樹形 1 つ )
- 無作為に選んだ 1OTU の枝を最適な場所へ付ける
- さらに 1OTU 足す、を繰り返す
- 全 OTU の系統樹を得る



反復すれば多数の樹形が得られる

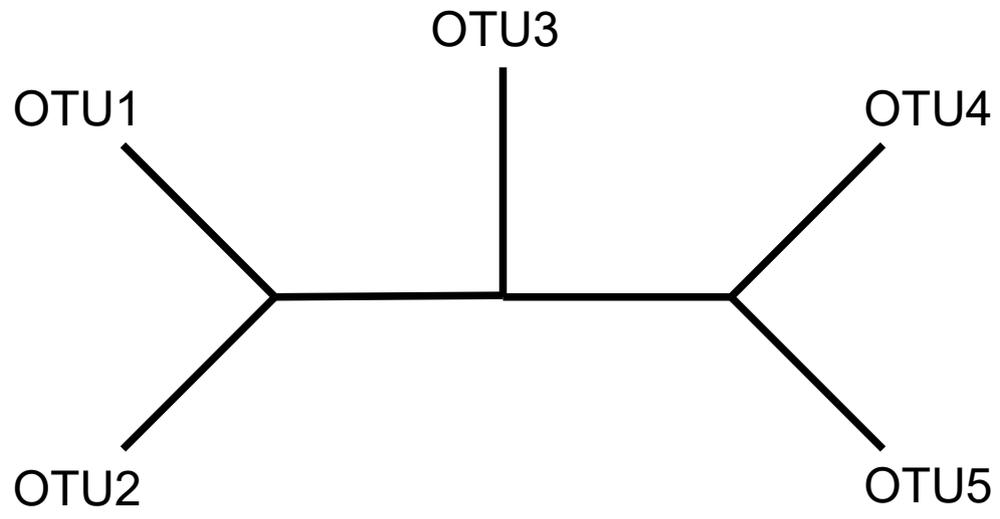


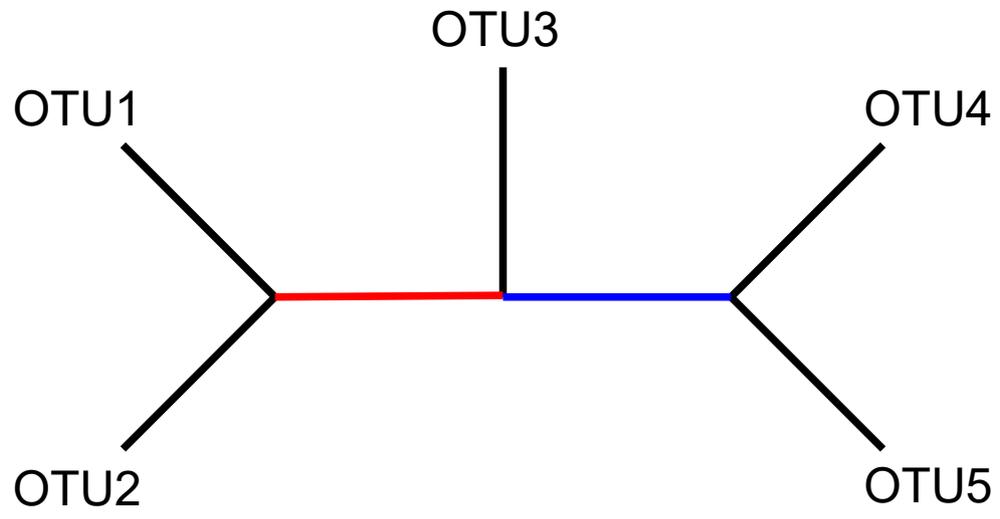
はたして

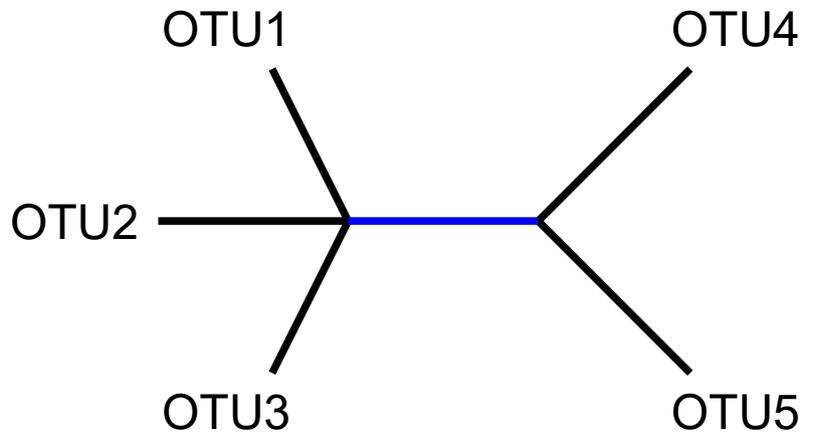
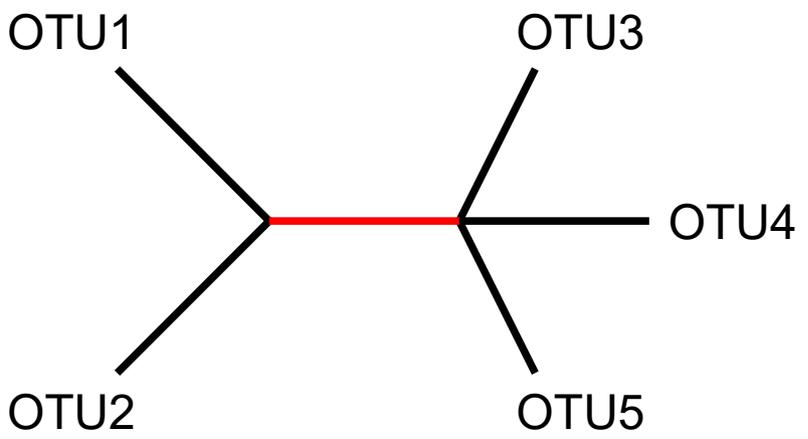
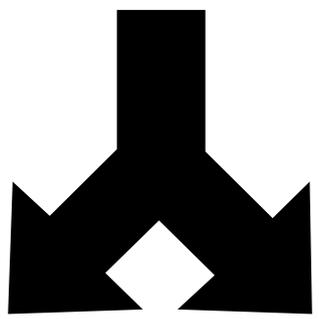
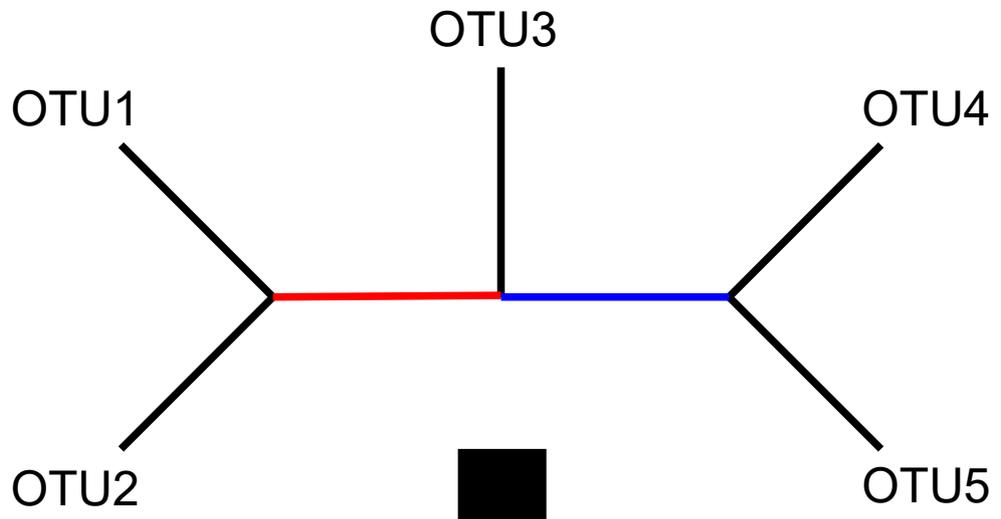
その系統樹は

どれくらい

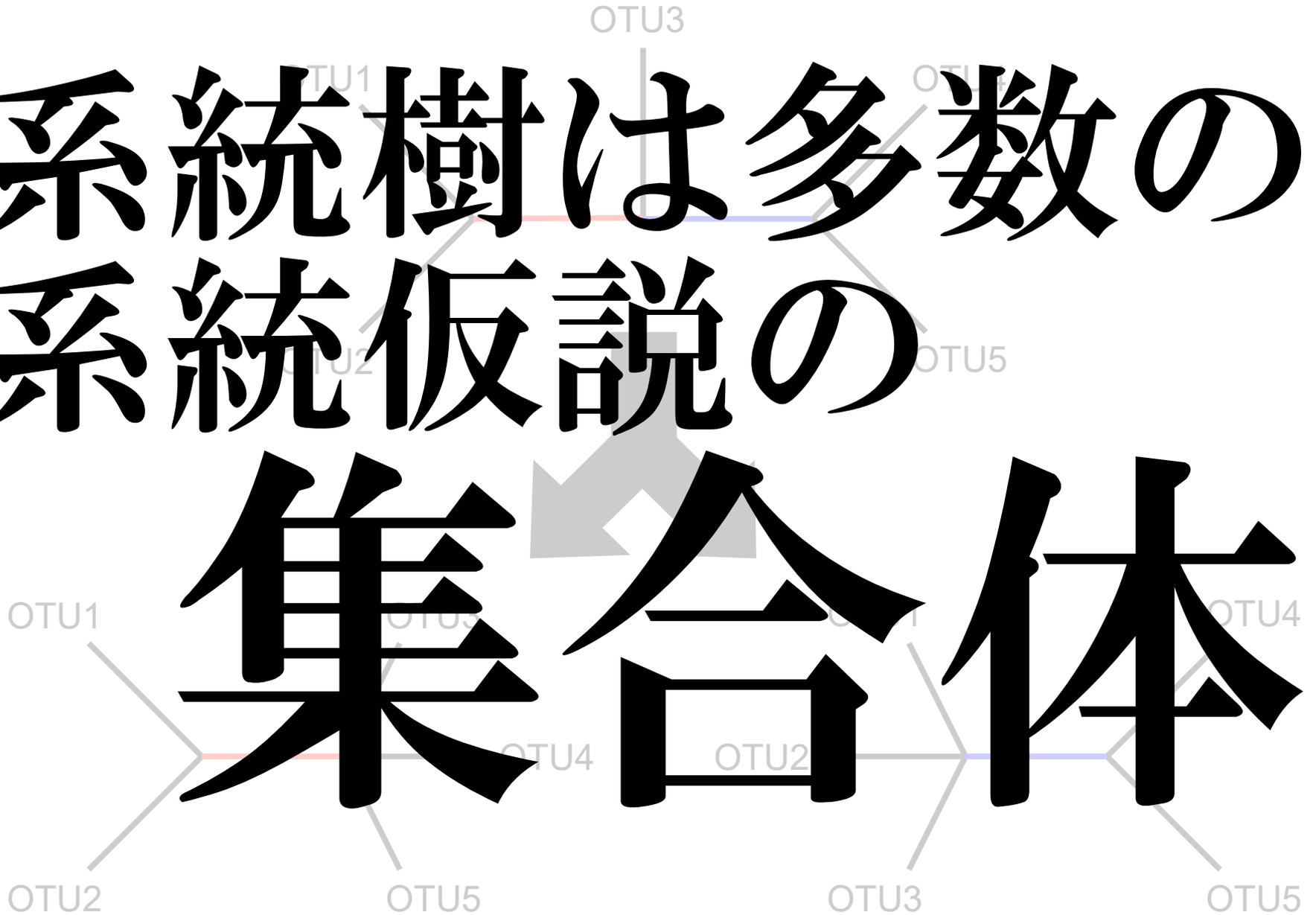
信頼できるのか？







# 系統樹は多数の 系統仮説の 集合体



はたして

その系統樹は

どれくらい

信頼できるのか？

はたして

各系統仮説は

どれくらい

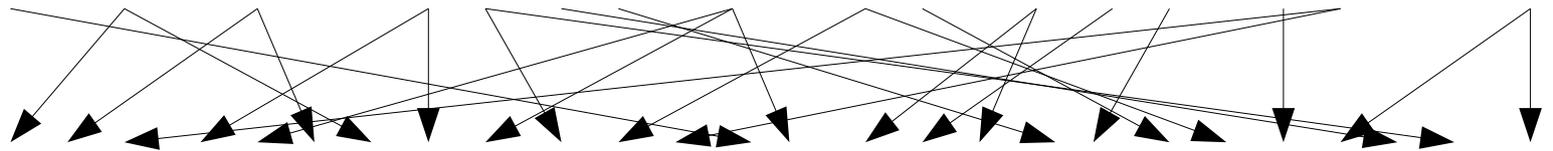
信頼できるのか？

ブートストラップ

解析

OTU1	A	C	C	G	T	T	A	C	C	G	A	T	T	A	C	C	G	A	T	T	A	C	T	T	A	C
OTU2	A	C	C	G	A	T	A	C	C	G	A	A	T	A	C	C	G	A	A	T	A	C	A	T	A	C
OTU3	T	C	C	C	A	T	T	C	C	C	A	A	T	T	C	C	C	A	A	T	T	C	A	T	T	C
OTU4	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU5	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU6	T	C	T	A	C	T	T	C	T	A	A	C	T	T	C	T	A	A	C	T	T	T	C	T	T	C
OTU7	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C
OTU8	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C
OTU9	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C

OTU1	A	C	C	G	T	T	A	C	C	G	A	T	T	A	C	C	G	A	T	T	A	C	T	T	A	C
OTU2	A	C	C	G	A	T	A	C	C	G	A	A	T	A	C	C	G	A	A	T	A	C	A	T	A	C
OTU3	T	C	C	C	A	T	T	C	C	C	A	A	T	T	C	C	C	A	A	T	T	C	A	T	T	C
OTU4	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU5	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU6	T	C	T	A	C	T	T	C	T	A	A	C	T	T	C	T	A	A	C	T	T	T	C	T	T	C
OTU7	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C
OTU8	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C
OTU9	T	C	T	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C



OTU1	C	T	T	C	T	T	C	C	T	C	C	T	A	T	A	T	A	A	T	C	C	C	C	G	C	C
OTU2	C	A	A	C	T	A	C	C	T	C	C	A	A	T	A	A	A	A	T	C	C	C	C	G	C	C
OTU3	C	A	A	C	T	A	C	C	T	C	C	A	T	T	A	A	A	A	T	C	C	C	C	C	C	C
OTU4	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	C	T	C
OTU5	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	C	T	C
OTU6	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	A	T	C
OTU7	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C
OTU8	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C
OTU9	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C

OTU1  
OTU2  
OTU3  
OTU4  
OTU5  
OTU6  
OTU7  
OTU8  
OTU9

A C C G T T A C C G A T T A C C G A T T A C T T A C T T A C  
A C C G T A C C G A A C G A A A C A A C A T A C  
T C T A T T C C C A T C A A A C A A T T C A T T C C  
T C T C A C T T C T A A T T C A A C T T C T T C C  
T C T A G T T C T A G T T C A A G T T G T T C C  
T C T A G T T C A A G T T C A A G T T G T T C C  
T C T A G T T C A A G T T C A A G T T G T T C C

データから  
データを作る

OTU1  
OTU2  
OTU3  
OTU4  
OTU5  
OTU6  
OTU7  
OTU8  
OTU9

C T T C T T C C T C C T A T A A T C C C C C G C C  
C A A C T A C C T C C A A T A A A T C C C C C G C C  
C A A C T A C C T C C A T T A A A T C C C C C C C C  
T C C C T C T C T T C C T T A C A A T T C C C T C C  
T C C C T C T C T T C C T T A C A A T T C C A T C C  
T G G C T G T C T T C G T T A G A A T T C C A T C C  
T G G C T G T C T T C G T T A G A A T T C C A T C C  
T G G C T G T C T T C G T T A G A A T T C C A T C C



OTU1	A	C	C	G	T	T	A	C	C	G	A	T	T	A	C	C	G	A	T	T	A	C	T	T	A	C
OTU2	A	C	C	G	A	T	A	C	C	G	A	A	T	A	C	C	G	A	A	T	A	C	A	T	A	C
OTU3	T	C	C	C	A	T	T	C	C	C	A	A	T	T	C	C	C	A	A	T	T	C	A	T	T	C
OTU4	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU5	T	C	T	C	C	T	T	C	T	C	A	C	T	T	C	T	C	A	C	T	T	T	C	T	T	C
OTU6	T	C	T	A	C	T	T	C	T	A	A	C	T	T	C	T	A	A	C	T	T	T	C	T	T	C
OTU7	T	C	A	G	T	T	C	T	A	A	C	T	T	C	T	A	A	G	T	T	T	G	T	T	C	
OTU8	T	C	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C	
OTU9	T	C	A	G	T	T	C	T	A	A	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C	

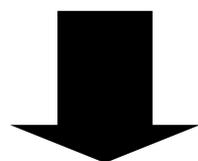
# リサンプリング

OTU1	C	T	T	C	T	T	C	C	T	C	C	T	A	T	A	T	A	A	T	C	C	C	C	G	C	C
OTU2	C	A	A	C	T	A	C	C	T	C	C	A	A	T	A	A	A	A	T	C	C	C	C	G	C	C
OTU3	C	A	A	C	T	A	C	C	T	C	C	A	T	T	A	A	A	A	T	C	C	C	C	C	C	C
OTU4	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	C	T	C
OTU5	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	C	T	C
OTU6	T	C	C	C	T	C	T	C	T	T	C	C	T	T	A	C	A	A	T	T	C	T	C	A	T	C
OTU7	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C
OTU8	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C
OTU9	T	G	G	C	T	G	T	C	T	T	C	G	T	T	A	G	A	A	T	T	C	T	C	A	T	C



リサンプル

リサンプル

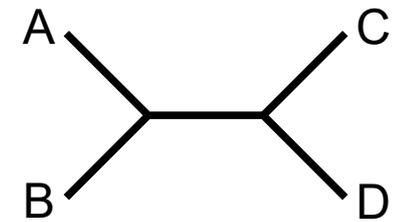
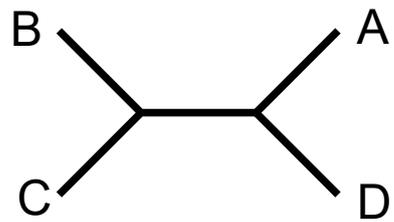
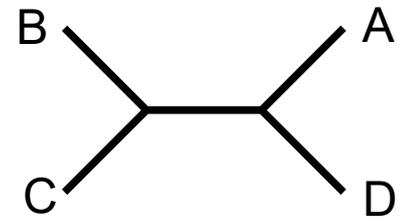
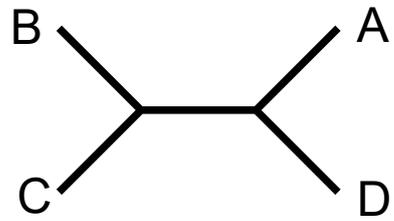


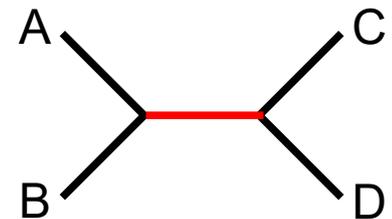
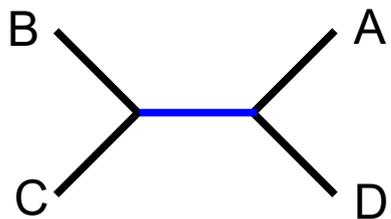
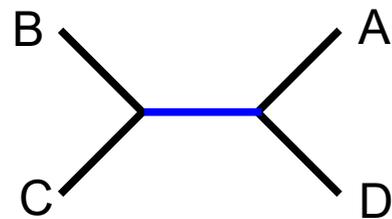
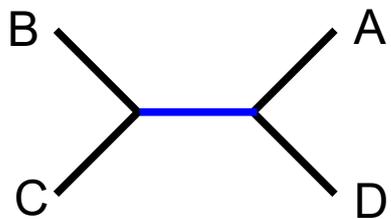
最尤系統樹推定

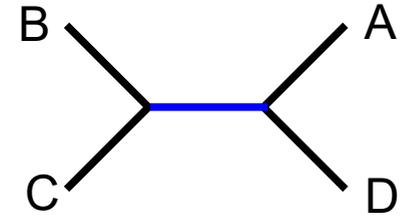
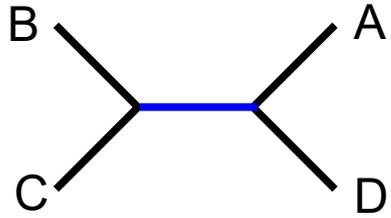
ひたすら

↓  
反復

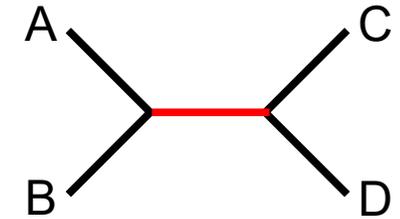
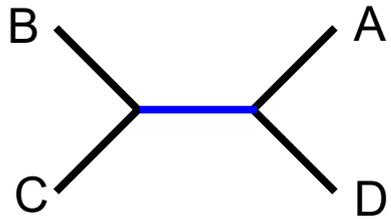
最元祖統御推定







**75%**



**25%**

# 出現頻度を 信頼性の指標に

B A

B A

D

B A

A C

C D

B D

25%