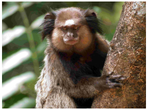
Change-point Detection on Phylogenetic Trees from Present-day Data

C. Ané[1], Paul Bastide[2], M. Mariadassou[3], S. Robin[4]

[1] Department of Statistics and Botany, University of Wisconsin–Madison, USA
[2] Evolutionary and Computational Virology, Rega Institute, KU Leuven, Belgium
[3] MaIAGE, INRA, Jouy-en-Josas, France.
[4] MIA-Paris, INRA - AgroParisTech, Paris, France

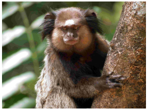
20 November 2017

New World Monkeys (Aristide et al., 2016)



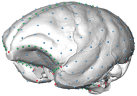
Callithrix penicillata

New World Monkeys (Aristide et al., 2016)

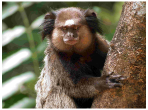


Callithrix penicillata

$\longrightarrow$

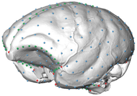


New World Monkeys (Aristide et al., 2016)

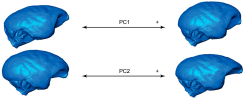


Callithrix penicillata

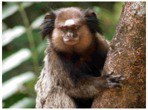
⟶



⟶

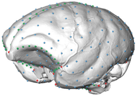


New World Monkeys (Aristide et al., 2016)

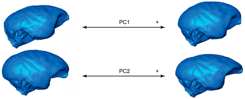


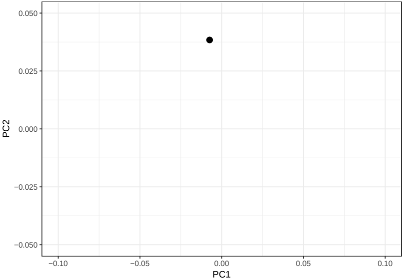
Callithrix penicillata

$\longrightarrow$

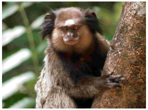


$\longrightarrow$



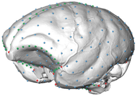


New World Monkeys (Aristide et al., 2016)

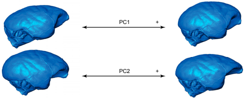


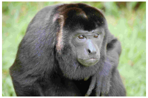
Callithrix penicillata

→

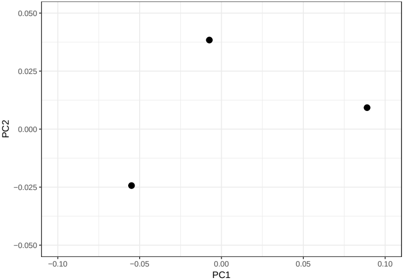


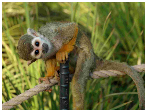
→





Alouatta palliata

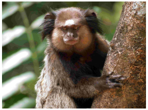




Saimiri sciureus

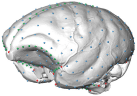
New World Monkeys

(Aristide et al., 2016)

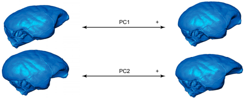


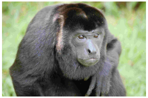
Callithrix penicillata

→

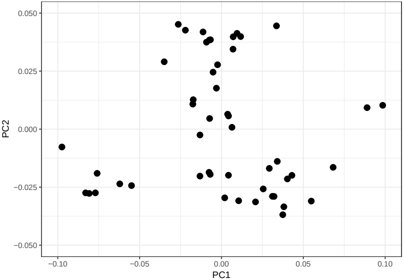


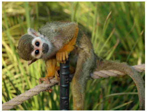
→





Alouatta palliata

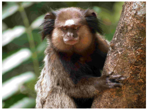




Saimiri sciureus

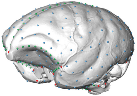
New World Monkeys

(Aristide et al., 2016)

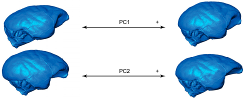


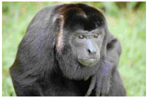
Callithrix penicillata

$\longrightarrow$

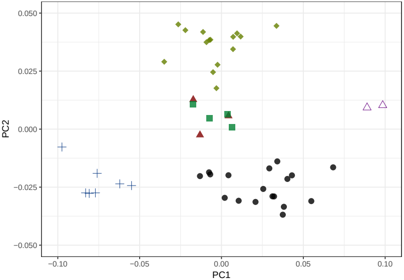


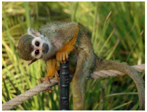
$\longrightarrow$





Alouatta palliata

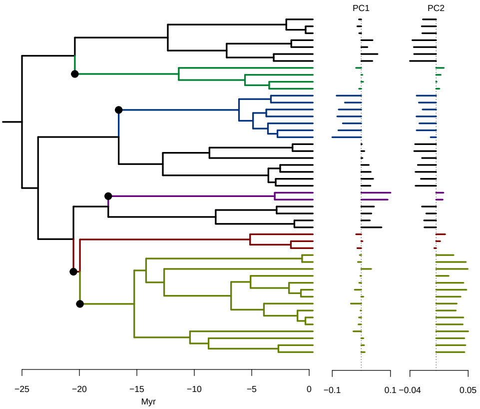


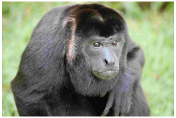


Saimiri sciureus

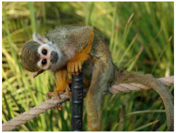
New World Monkeys

(Aristide et al., 2016)

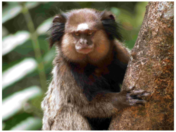




Alouatta palliata

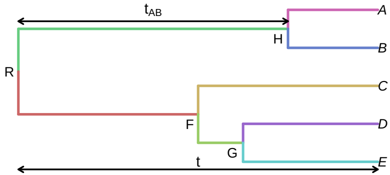


Saimiri sciureus



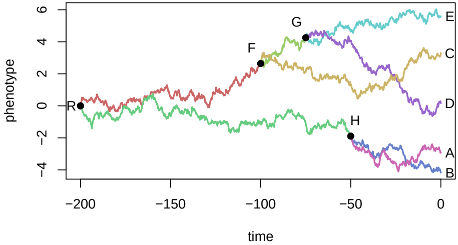
Callithrix penicillata

Shifted BM on a Tree (Felsenstein, 1985)



Known tree.

Only tip values observed.

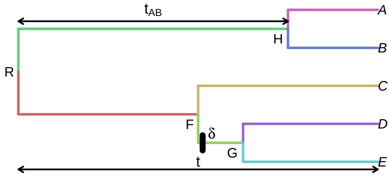


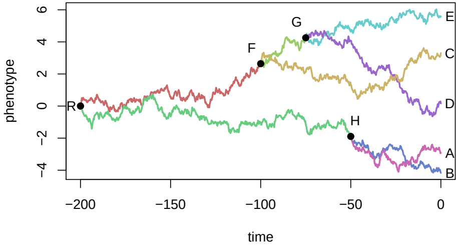
Brownian Motion:

$$\mathbb{V}\mathrm{ar}\left[A \mid R\right] = \sigma^2 t$$
$$\mathbb{C}\mathrm{ov}\left[A; B \mid R\right] = \sigma^2 t_{AB}$$

Shifted BM on a Tree

(Felsenstein, 1985)





Known tree.

Only tip values observed.

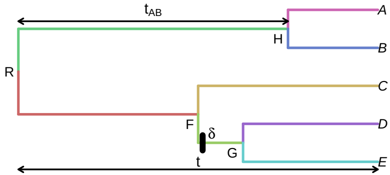
Brownian Motion:

$$\mathbb{V}\text{ar}\,[A \mid R] = \sigma^2 t$$
$$\mathbb{C}\text{ov}\,[A; B \mid R] = \sigma^2 t_{AB}$$

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

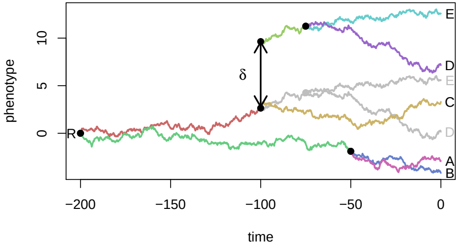
Shifted BM on a Tree

(Felsenstein, 1985)



Known tree.

Only tip values observed.



Brownian Motion:

$$\mathbb{V}\mathrm{ar}\left[A \mid R\right] = \sigma^2 t$$
$$\mathbb{C}\mathrm{ov}\left[A; B \mid R\right] = \sigma^2 t_{AB}$$

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

Outline

1. Shifted BM on a Tree
2. Shifted OU on a Tree
3. Multivariate Trait

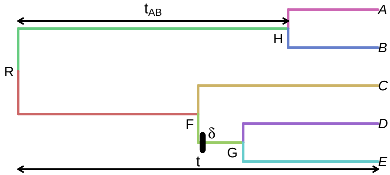
Outline

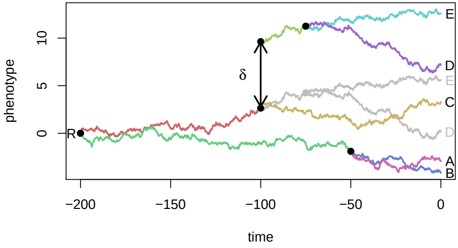
1. Shifted BM on a Tree
   - Identifiability
   - Incomplete Data Model
   - Linear Regression Model

2. Shifted OU on a Tree

3. Multivariate Trait

Shifted BM on a Tree (Felsenstein, 1985)





Known tree.

Only tip values observed.

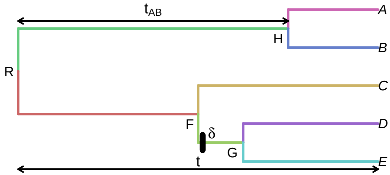
Brownian Motion:

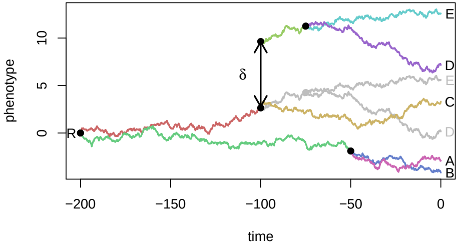
$$\mathbb{V}\mathrm{ar}\left[A \mid R\right] = \sigma^2 t$$
$$\mathbb{C}\mathrm{ov}\left[A; B \mid R\right] = \sigma^2 t_{AB}$$

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

Shifted BM on a Tree

(Felsenstein, 1985)





Known tree.

Only tip values observed.

Goal: Find shifts position.

Brownian Motion:

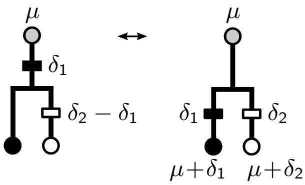
$$\mathbb{V}\text{ar}\,[A \mid R] = \sigma^2 t$$

$$\mathbb{C}\text{ov}\,[A;\, B \mid R] = \sigma^2 t_{AB}$$

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

Equivalencies

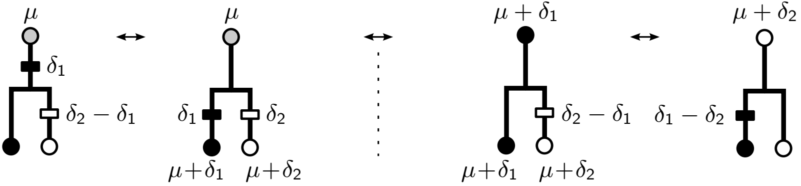
- Equivalent configurations:



- Over-parametrization: parsimonious configurations.

Equivalencies

- Equivalent configurations:



- Over-parametrization: parsimonious configurations.

Parsimonious Solution: Definition

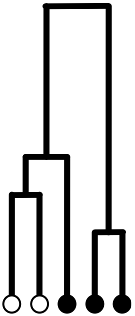
Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

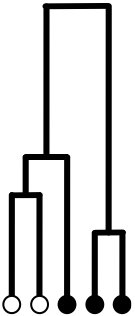
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

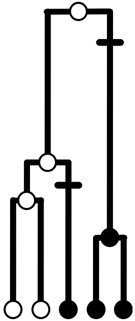


Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

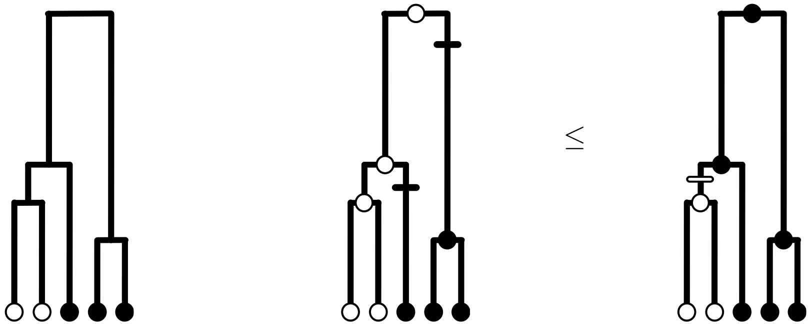




Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

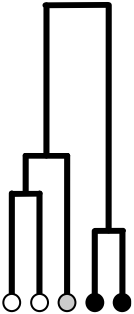
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

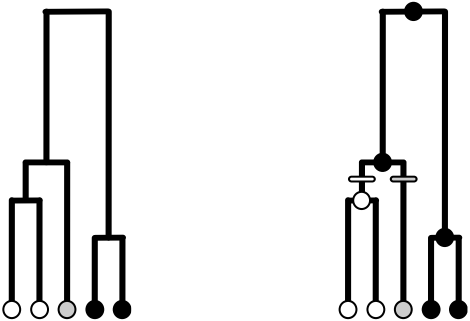
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

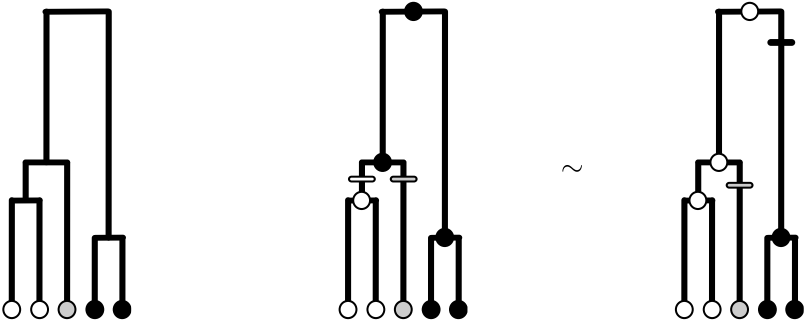
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution: Definition

Definition (Parsimonious Allocation)

A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Equivalent Parsimonious Allocations

Definition (Equivalency)

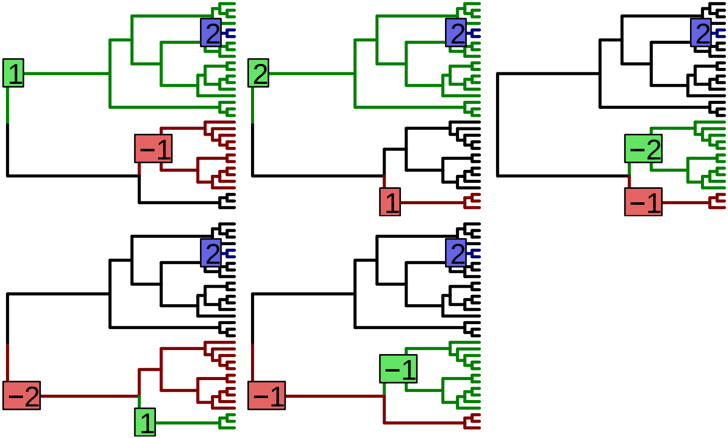
Two allocations are said to be *equivalent* (noted $\sim$) if they are both parsimonious and give the same colors at the tips.

Find one solution Existing Dynamic Programming algorithms (Fitch, Sankoff, see Felsenstein, 2004).

Enumerate all solutions New adapted recursive algorithm (implemented in `PhylogeneticEM`).

+

Equivalent Parsimonious Solutions



Equivalent allocations and values of the shifts - BM.

Collection of Models

New Problem Number of Equivalence Classes: $\left|\mathcal{S}_K^{PI}\right|$?

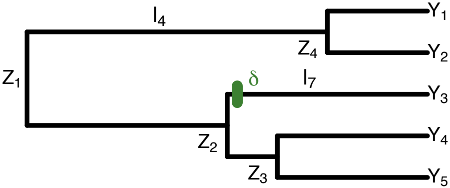
- $\left|\mathcal{S}_K^{PI}\right| \leq \binom{m+n-1}{K} = \binom{\#\text{ of edges}}{\#\text{ of shifts}}$
- Recursive algorithm to compute $\left|\mathcal{S}_K^{PI}\right|$ (implemented in `PhylogeneticEM`).

$\mapsto$ Generally dependent on the topology of the tree. +

- Binary tree: $\left|\mathcal{S}_K^{PI}\right| = \binom{2n-2-K}{K} = \binom{\#\text{ of edges}-\#\text{ of shifts}}{\#\text{ of shifts}}$

$\mapsto$ See convex characters: Semple and Steel (2003)

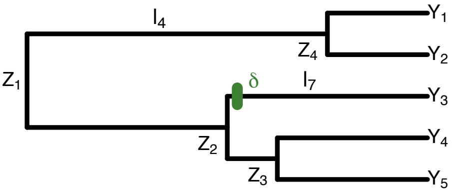
Incomplete Data Model



Y : observed traits

Z : latent variables

Incomplete Data Model

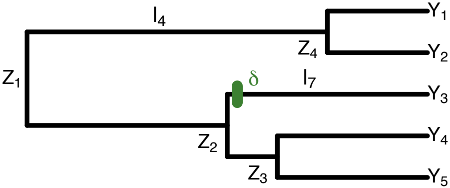


Y : observed traits

Z : latent variables

$$\textit{BM}: \quad \begin{aligned} Z_4 | Z_1 &\sim \mathcal{N}\left(Z_1 \qquad , \sigma^2 \ell_4\right) \\ Y_3 | Z_2 &\sim \mathcal{N}\left(Z_2 + \delta, \sigma^2 \ell_7\right) \end{aligned}$$

Incomplete Data Model



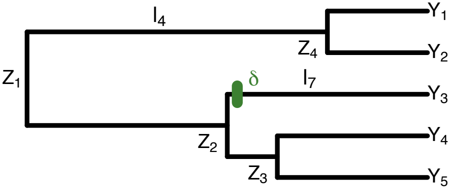
Y : observed traits

Z : latent variables

$$\begin{aligned} BM: \quad Z_4|Z_1 &\sim \mathcal{N}\left(Z_1 \qquad , \sigma^2\ell_4\right) \\ Y_3|Z_2 &\sim \mathcal{N}\left(Z_2 + \delta, \sigma^2\ell_7\right) \end{aligned}$$

$$p_\theta(\mathbf{Z}, \mathbf{Y}) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j|Z_{\text{parent}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i|Z_{\text{parent}(i)})$$

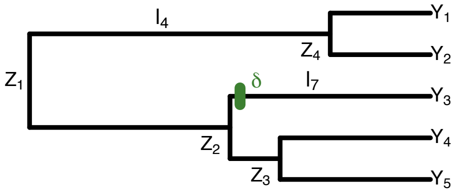
EM Algorithm: K fixed



$$BM: \quad \begin{aligned} Z_4 | Z_1 &\sim \mathcal{N}\left(Z_1 \qquad, \sigma^2 \ell_4\right) \\ Y_3 | Z_2 &\sim \mathcal{N}\left(Z_2 + \delta, \sigma^2 \ell_7\right) \end{aligned}$$

$$\text{Goal:} \quad \hat{\boldsymbol{\theta}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\boldsymbol{\theta}}_\eta}(\mathbf{Y})$$

EM Algorithm: K fixed

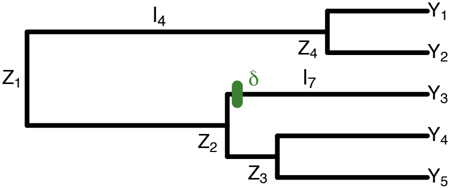


$$BM: \quad \begin{array}{l} Z_4 | Z_1 \sim \mathcal{N}\left(Z_1 \qquad, \sigma^2 \ell_4\right) \\ Y_3 | Z_2 \sim \mathcal{N}\left(Z_2 + \delta, \sigma^2 \ell_7\right) \end{array}$$

Goal: $\hat{\boldsymbol{\theta}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\boldsymbol{\theta}}_\eta}(\mathbf{Y})$

EM Maximize $\log p_{\boldsymbol{\theta}}(\mathbf{Y})$ through $\mathbb{E}_{\boldsymbol{\theta}}[\log p_{\boldsymbol{\theta}}(\mathbf{Z}, \mathbf{Y}) \mid \mathbf{Y}]$.

EM Algorithm: K fixed



$$BM: \quad Z_4|Z_1 \sim \mathcal{N}\left(Z_1 \quad\quad, \sigma^2\ell_4\right)$$
$$Y_3|Z_2 \sim \mathcal{N}\left(Z_2 + \delta, \sigma^2\ell_7\right)$$

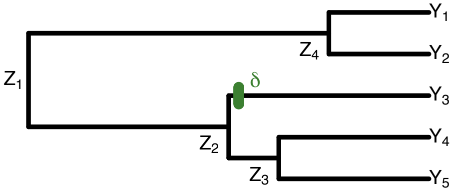
$$\text{Goal:} \quad \hat{\boldsymbol{\theta}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\boldsymbol{\theta}}_\eta}(\mathbf{Y})$$

EM Maximize $\log p_{\boldsymbol{\theta}}(\mathbf{Y})$ through $\mathbb{E}_{\boldsymbol{\theta}}[\log p_{\boldsymbol{\theta}}(\mathbf{Z}, \mathbf{Y}) \mid \mathbf{Y}]$.

E step Given $\boldsymbol{\theta}^h$, compute $p_{\boldsymbol{\theta}^h}(\mathbf{Z} \mid \mathbf{Y})$

M step $\boldsymbol{\theta}^{h+1} = \operatorname{argmax}_{\boldsymbol{\theta}} \left\{\mathbb{E}_{\boldsymbol{\theta}^h}[\log p_{\boldsymbol{\theta}}(\mathbf{Z}, \mathbf{Y}) \mid \mathbf{Y}]\right\}$

E step



Compute the following quantities:

$$\mathbb{E}^{(h)}[Z_j \mid \mathbf{Y}],\ \mathbb{V}\mathrm{ar}^{(h)}\left[Z_j \mid \mathbf{Y}\right],\ \mathbb{C}\mathrm{ov}^{(h)}\left[Z_j, Z_{\mathrm{parent}(j)} \mid \mathbf{Y}\right]$$

- Gaussian properties: $O(n^3)$.
- Gaussian properties + Tree structure: $O(n)$.
  $\mapsto$ "Upward-Downward" algorithm.

M Step

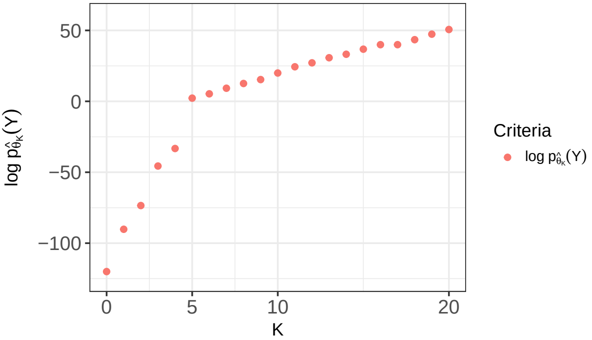
Maximize:

$$\mathbb{E}\left[\log p_\theta(\mathbf{Z}, \mathbf{Y}) \mid \mathbf{Y}\right] = -\sum_{j=2}^{m+n} C_j(\mathbf{\Delta}) + \mathcal{F}^{(h)}\left(\mu, \sigma^2\right)$$

- μ, σ^2: simple maximization
- Discrete location of K shifts
  - $\mapsto$ Exact and fast for the BM

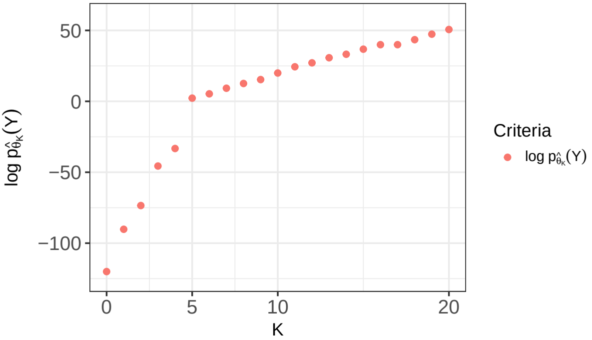
+

Model Selection: Penalized Likelihood



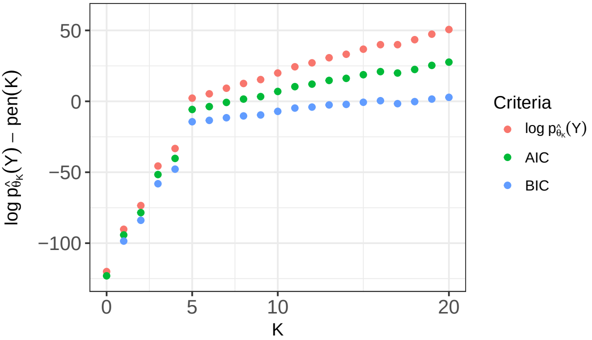
Model Selection: Penalized Likelihood

Idea $\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmax}} \left\{ \log \; p_{\hat{\theta}_K}(\mathbf{Y}) - \operatorname{pen}(K) \right\}$



Model Selection: Penalized Likelihood

Idea $$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmax}} \left\{ \log \ p_{\hat{\theta}_K}(\mathbf{Y}) - \operatorname{pen}(K) \right\}$$



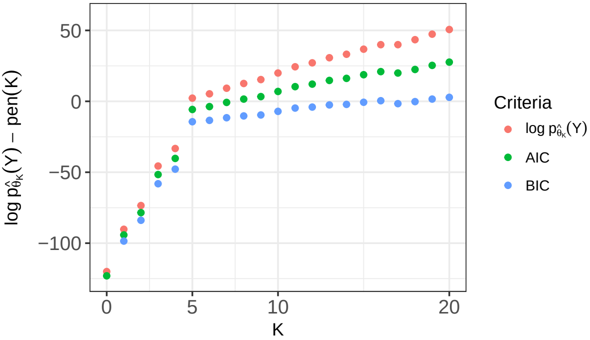
Penalties:

AIC $K + 3$

BIC $\frac{1}{2}(K+3)\log(n)$

Model Selection: Penalized Likelihood

Idea $$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmax}} \left\{ \log\ p_{\hat{\theta}_K}(\mathbf{Y}) - \operatorname{pen}(K) \right\}$$



Penalties:

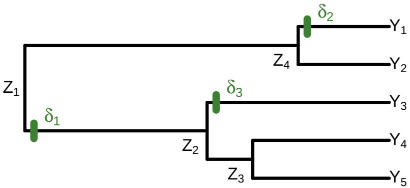
AIC $K + 3$

BIC $\frac{1}{2}(K+3)\log(n)$

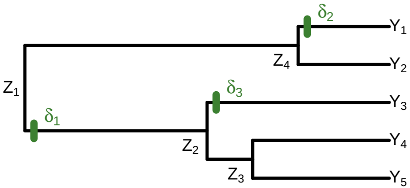
Solution

- Use $\left|\mathcal{S}_K^{PI}\right|$.
- Linear Regression Model.

Linear Regression Model

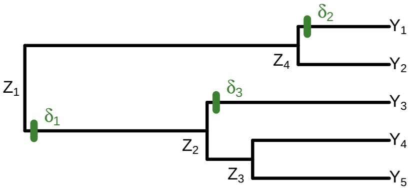


Linear Regression Model



$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu \\ \delta_1 \\ \cdot \\ \cdot \\ \delta_2 \\ \cdot \\ \delta_3 \\ \cdot \\ \cdot \end{pmatrix}$$

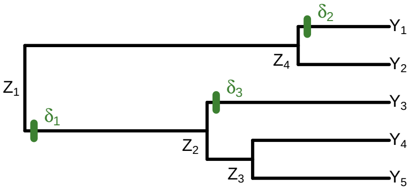
Linear Regression Model



$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu \\ \delta_1 \\ \cdot \\ \cdot \\ \delta_2 \\ \cdot \\ \delta_3 \\ \cdot \\ \cdot \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

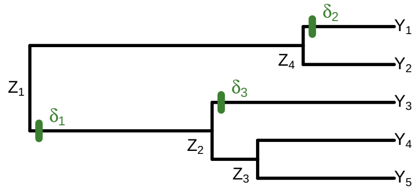
Linear Regression Model



$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu \\ \delta_1 \\ \cdot \\ \cdot \\ \delta_2 \\ \cdot \\ \delta_3 \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T\Delta} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu + \delta_2 \\ \mu \\ \mu + \delta_1 + \delta_3 \\ \mu + \delta_1 \\ \mu + \delta_1 \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

Linear Regression Model



$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu \\ \delta_1 \\ \cdot \\ \cdot \\ \delta_2 \\ \cdot \\ \delta_3 \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T\Delta} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{pmatrix} \mu + \delta_2 \\ \mu \\ \mu + \delta_1 + \delta_3 \\ \mu + \delta_1 \\ \mu + \delta_1 \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

$$BM: \quad \mathbf{Y} = \mathbf{T\Delta} + \sigma \mathbf{E}^{BM}$$

$$\mathbf{E}^{BM} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} + \hat{\sigma}^2_K \operatorname{pen}\left(n, K, \left|\mathcal{S}^{PI}_K\right|\right) \right\}$$

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|_{\mathbf{V}^{-1}}^2 + \hat{\sigma}_K^2 \operatorname{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right) \right\}$$

$$\hat{\sigma}_K^2 = \frac{\left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|_{\mathbf{V}^{-1}}^2}{n - K - 1}$$

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\operatorname{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\text{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Oracle

$$\inf_{\eta \in \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI}} \left\| \mathbb{E}\left[\mathbf{Y}\right] - \mathbf{Y}^*_\eta \right\|^2_{\mathbf{V}^{-1}}$$

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\operatorname{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Oracle

$$\inf_{\eta \in \bigcup_{K=0}^{p-1} \mathcal{S}_K^{PI}} \left\| \mathbb{E}\left[\mathbf{Y}\right] - \mathbf{Y}^*_\eta \right\|^2_{\mathbf{V}^{-1}}$$

Definition (Baraud et al. (2009))

Let D, $N > 0$, and $X_D \sim \chi^2(D)$, $X_N \sim \chi^2(N)$, $X_D \perp X_N$.

$$\operatorname{Dkhi}[D, N, x] = \frac{1}{\mathbb{E}\left[X_D\right]} \mathbb{E}\left[\left(X_D - x \frac{X_N}{N} \right)_+ \right], \quad \forall x > 0$$

$$\operatorname{Dkhi}[D, N, \operatorname{EDkhi}[D, N, q]] = q, \quad \forall 0 < q \leq 1$$

LINselect: Oracle Inequality

Proposition (Form of the Penalty and guarantees)

Under our setting: $\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \sigma\mathbf{E}$ *with* $E \sim \mathcal{N}(0, \mathbf{V})$, *define the penalty:*

$$\text{pen}(K) = A\frac{n-K-1}{n-K-2}\,\text{EDkhi}\left[K+2, n-K-2, exp\left(-\log\left|\mathcal{S}_K^{PI}\right| - 2\log(K+2)\right)\right]$$

If $\kappa < 1$, *and* $p \leq \min\left(\frac{\kappa n}{2+\log(2)+\log(n)}, n-7\right)$, *we get:*

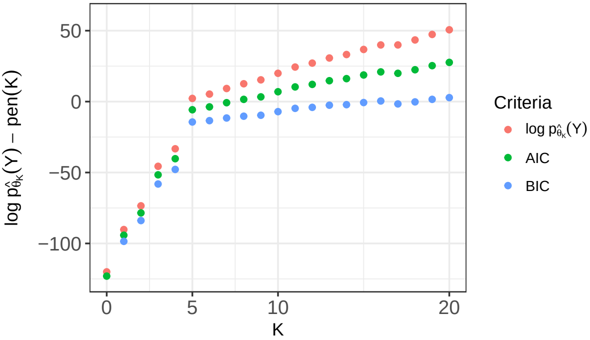
$$\mathbb{E}\left[\frac{\left\|\mathbb{E}\left[\mathbf{Y}\right] - \hat{\mathbf{Y}}_{\hat{K}}\right\|^2_{\mathbf{V}^{-1}}}{\sigma^2}\right] \leq C(A,\kappa)\inf_{\eta\in\mathcal{M}}\left\{\frac{\left\|\mathbb{E}\left[\mathbf{Y}\right] - \mathbf{Y}^*_\eta\right\|^2_{\mathbf{V}^{-1}}}{\sigma^2} + (K_\eta + 2)\left(3 + \log(n)\right)\right\}$$

with $C(A,\kappa)$ *a constant depending on* A *and* κ *only.*

Based on Baraud et al. (2009) +

Model Selection: Penalized Likelihood

Idea $\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmax}} \left\{ \log \ p_{\hat{\theta}_K}(\mathbf{Y}) - \operatorname{pen}(K) \right\}$



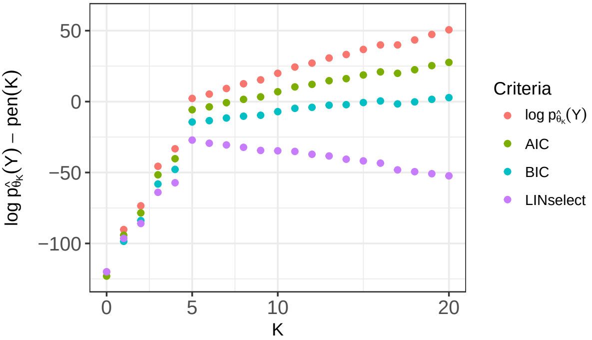
Penalties:

AIC $K + 3$

BIC $\frac{1}{2}(K + 3)\log(n)$

Model Selection: Penalized Likelihood

Idea $$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmax}} \left\{ \log \; p_{\hat{\theta}_K}(\mathbf{Y}) - \operatorname{pen}(K) \right\}$$



Penalties:

AIC $K + 3$

BIC $\frac{1}{2}(K + 3)\log(n)$

LINselect $\operatorname{pen}(n, K, |\mathcal{S}_K^{PI}|)$

LINselect Model Selection: Important Points

Based on Baraud, Giraud, and Huet (2009)

- Non-asymptotic bound.
- Unknown variance.
- No constant to be calibrated.

Note

- Non iid variance.
- Penalty depends on the tree topology (through $\left|\mathcal{S}_K^{PI}\right|$).

LASSO Regression

Lasso regression:

$$\hat{\mathbf{\Delta}} = \underset{\mathbf{\Delta}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \mathbf{T}\mathbf{\Delta} \right\|_{\mathbf{V}^{-1}}^{2} + \lambda \left\| \mathbf{\Delta}_{-1} \right\|_{1} \right\}$$

LASSO Regression

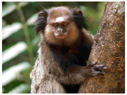
Lasso regression:

$$\hat{\mathbf{\Delta}} = \underset{\mathbf{\Delta}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \mathbf{T}\mathbf{\Delta} \right\|_{\mathbf{V}^{-1}}^{2} + \lambda \left\| \mathbf{\Delta}_{-1} \right\|_{1} \right\}$$

Initialization: For K fixed

- Choose λ to get K shifts
- Estimate $\mathbf{\Delta}$ with a Gauss Lasso

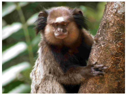
New World Monkey Dataset



We have:

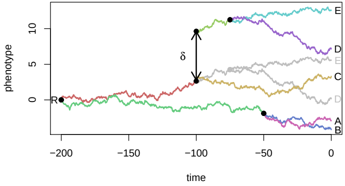
- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)

New World Monkey Dataset

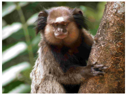


We have:

- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)

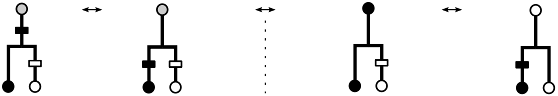


New World Monkey Dataset

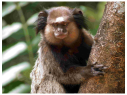


We have:

- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)

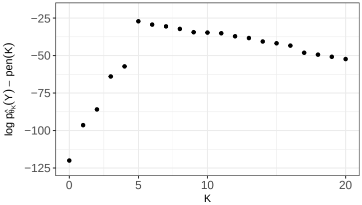


New World Monkey Dataset

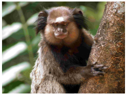


We have:

- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)



New World Monkey Dataset



We have:

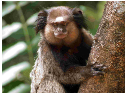
- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)

But...

- The BM is not realistic in many cases.
  - No selection.
  - Unbounded variance.

$\mapsto$ Use the Ornstein-Uhlenbeck instead.

New World Monkey Dataset



We have:

- A model of trait evolution
- A way to asses identifiability
- An inference strategy (EM + LINselect)

But...

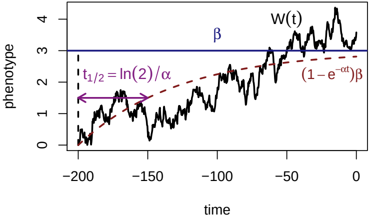
- The BM is not realistic in many cases.
    - No selection.
    - Unbounded variance.

$\mapsto$ Use the Ornstein-Uhlenbeck instead.

Outline

1. Shifted BM on a Tree
2. Shifted OU on a Tree
    - Ornstein-Uhlenbeck
    - Re-scaling
3. Multivariate Trait

Ornstein-Uhlenbeck Modeling (Hansen, 1997)



$$dW(t) = \alpha[\beta - W(t)]dt + \sigma dB(t)$$

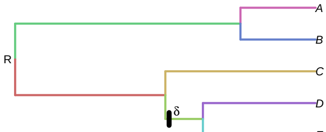
Deterministic part:

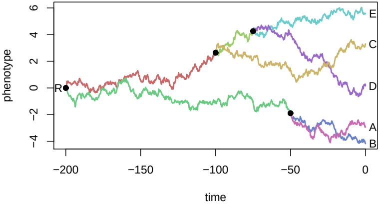
- β: primary optimum (mechanistically defined).
- $\ln(2)/\alpha$: phylogenetic half live.

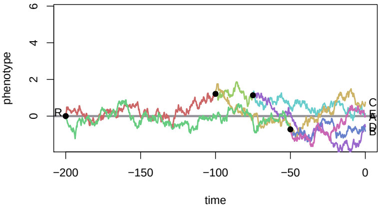
Stochastic part:

- $W(t)$: trait value (actual optimum).
- $\sigma dB(t)$: Brownian fluctuations.

Shifts







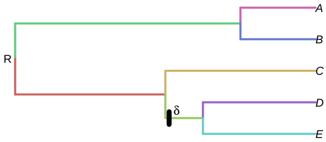
BM Shifts in the **mean**:

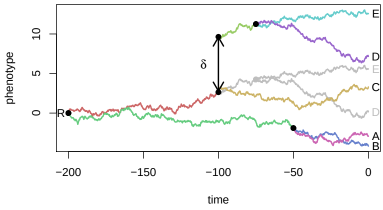
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

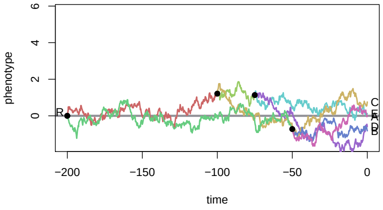
OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Shifts







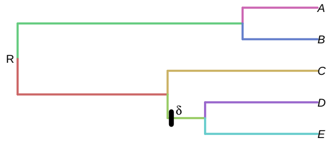
BM Shifts in the **mean**:

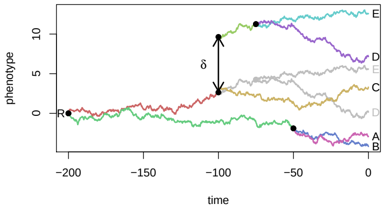
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

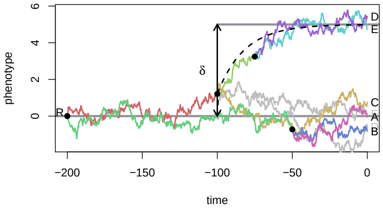
OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

Shifts







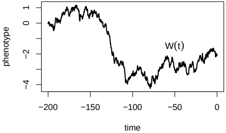
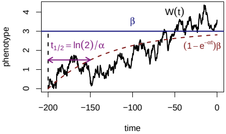
BM Shifts in the **mean**:

$$m_{\text{child}} = m_{\text{parent}} + \delta$$

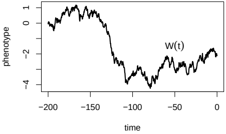
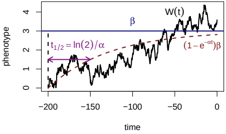
OU Shifts in the **optimal value**:

$$\beta_{\text{child}} = \beta_{\text{parent}} + \delta$$

BM vs OU

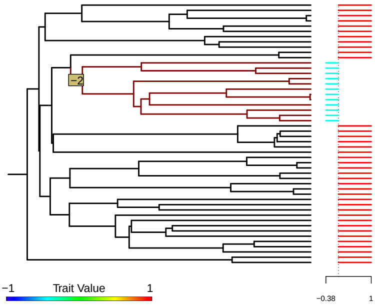
	Equation	Stationary State	$\mathbb{C}\text{ov}[Y_i; Y_j]$
	$dW(t) = \sigma dB(t)$	None.	$t_{ij} \times \sigma^2$
	$dW(t) = \sigma dB(t)$ $+\alpha[\beta - W(t)]dt$	$\begin{cases} \mu = \beta \\ \gamma^2 = \dfrac{\sigma^2}{2\alpha} \end{cases}$	$\dfrac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t_{ij}} - 1) \times \sigma^2$

BM vs OU

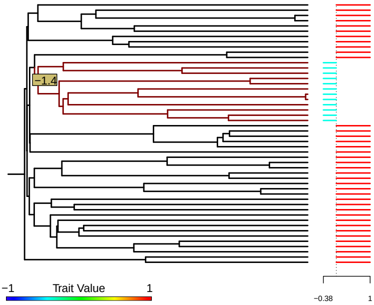
	Equation	Stationary State	$\mathbb{C}\text{ov}\left[Y_i; Y_j\right]$
	$dW(t) = \sigma dB(t)$	None.	$t_{ij} \times \sigma^2$
	$dW(t) = \sigma dB(t)$ $+\alpha[\beta - W(t)]dt$	$\begin{cases} \mu = \beta \\ \gamma^2 = \dfrac{\sigma^2}{2\alpha} \end{cases}$	$\underbrace{\frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t_{ij}} - 1)}_{t'_{ij}(\alpha)} \times \sigma^2$

OU $\iff$ BM

$$\text{OU} \iff \text{BM on a re-scaled tree with } t' = \frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t} - 1)$$



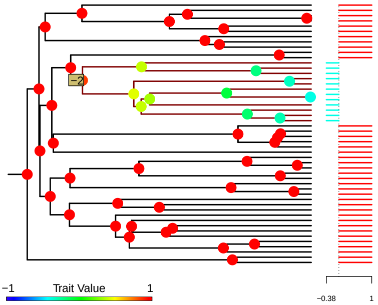
Original tree.



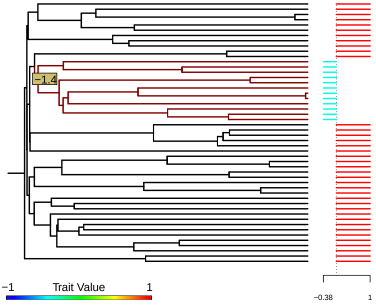
Re-scaled tree.

OU $\iff$ BM

$$\text{OU} \iff \text{BM on a re-scaled tree with } t' = \frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t} - 1)$$



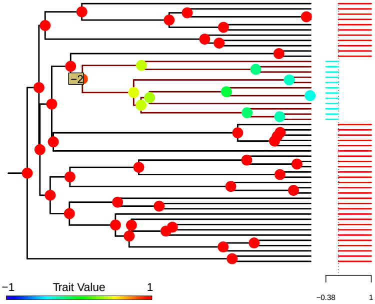
OU: $\beta_0 = \mu = 1$ and $t_{1/2} = 0.5$



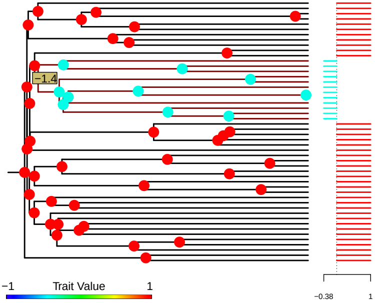
Re-scaled tree.

OU $\iff$ BM

$$\text{OU} \iff \text{BM on a re-scaled tree with } t' = \frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t} - 1)$$



OU: $\beta_0 = \mu = 1$ and $t_{1/2} = 0.5$



Re-scaled tree, equivalent BM.

OU $\iff$ BM

OU $\iff$ BM on a re-scaled tree with $t' = \frac{1}{2\alpha} e^{-2\alpha h}(e^{2\alpha t} - 1)$

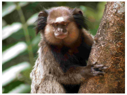
Remarks:

- This only works for an *ultrametric* tree.
- The laws of the internal nodes is changed.
- This is **not** the following standard time transformation

$$X_t = X_0 e^{-\alpha t} + \beta(1 - e^{-\alpha t}) + \frac{\sigma}{\sqrt{2\alpha}} e^{-\alpha t} B_{e^{2\alpha t}-1}$$

to get the BM solution of the OU.

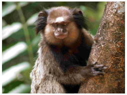
New World Monkey Dataset



We have:

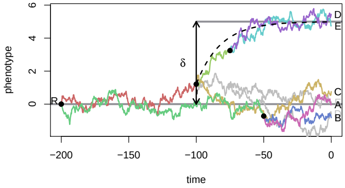
- A better model of trait evolution.
- A way to assess identifiability.
- An inference strategy (grid on α + EM + LINselect).

New World Monkey Dataset

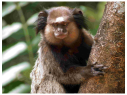


We have:

- A better model of trait evolution.
- A way to assess identifiability.
- An inference strategy (grid on α + EM + LINselect).

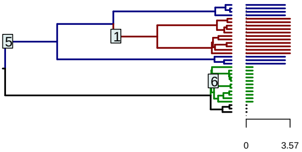


New World Monkey Dataset

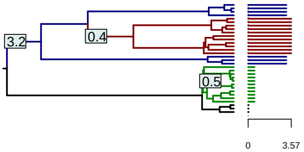


We have:

- A better model of trait evolution.
- A way to assess identifiability.
- An inference strategy (grid on α + EM + LINselect).



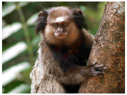
$\longrightarrow$



OU Groups are means, not regimes

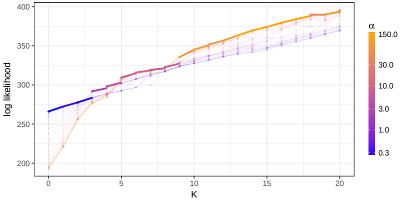
Defined from the equivalent BM

New World Monkey Dataset

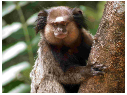


We have:

- A better model of trait evolution.
- A way to assess identifiability.
- An inference strategy (grid on α + EM + LINselect).



New World Monkey Dataset



We have:

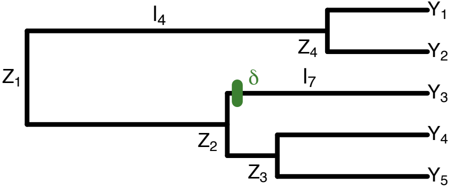
- A better model of trait evolution.
- A way to assess identifiability.
- An inference strategy (grid on α + EM + LINselect).

But...

- Brains are multivariate.

Outline

Multivariate BM



Data Vectors of p traits:

$$\mathbf{Y}_i^T = (Y_{i1}, \cdots, Y_{ip})$$

Shifts $\boldsymbol{\delta}$ vector size p.

$\hookrightarrow$ All traits shift together.

Incomplete Data Representation

$$\mathbf{Y}_3 \mid \mathbf{Z}_2 \sim \mathcal{N}\left(\mathbf{Z}_2 + \boldsymbol{\delta},\ \ell_7 \mathbf{R}\right)$$

Linear Model Representation

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Model Selection

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Problem: Can we do Model Selection when $\mathbf{R}$ is unknown ?

Model Selection

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Problem: Can we do Model Selection when $\mathbf{R}$ is unknown ?

- Slope Heuristic based method

$$\text{vec}(\mathbf{Y}) = (\mathbf{I}_p \otimes \mathbf{T})\,\text{vec}(\boldsymbol{\Delta}) + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{R} \otimes \mathbf{V})$$

  - Massart (2007)
    - oracle inequality with **known variance**
    - penalty up to a **multiplicative constant**
  - Baudry et al. (2012)
    - Slope-heuristic method to **calibrate the constant**
    - Implemented in `capushe` (Brault et al., 2012)

Model Selection

$$\mathbf{Y} = \mathbf{T}\mathbf{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Problem: Can we do Model Selection when $\mathbf{R}$ is unknown ?

- Slope Heuristic based method

$$\text{vec}(\mathbf{Y}) = (\mathbf{I}_p \otimes \mathbf{T})\,\text{vec}(\mathbf{\Delta}) + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{R} \otimes \mathbf{V})$$

  - Massart (2007)
    - oracle inequality with **known variance**
    - penalty up to a **multiplicative constant**
  - Baudry et al. (2012)
    - Slope-heuristic method to **calibrate the constant**
    - Implemented in `capushe` (Brault et al., 2012)

$\hookrightarrow$ Does not work well in practice.

Model Selection

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Problem: Can we do Model Selection when $\mathbf{R}$ is unknown ?

- Slope Heuristic based method

  $\hookrightarrow$ Does not work well in practice.

- phylogenetic BIC method Khabbazian et al. (2016)

  $\hookrightarrow$ Independant traits only.

Model Selection

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Problem: Can we do Model Selection when $\mathbf{R}$ is unknown ?

- Slope Heuristic based method

  $\hookrightarrow$ Does not work well in practice.

- phylogenetic BIC method Khabbazian et al. (2016)

  $\hookrightarrow$ Independant traits only.

- LINselect-based method

  $\hookrightarrow$ Idea: ML = LSQ for $\hat{\boldsymbol{\Delta}}$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \mathrm{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y})$$

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\mathrm{argmin}} \left\| \mathbf{Y} - \hat{\mathbf{Y}}_\eta \right\|^2_{\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\mathrm{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\mathrm{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \operatorname{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y})$$

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmin}} \left\| \mathbf{Y} - \hat{\mathbf{Y}}_\eta \right\|^2_{\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\operatorname{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \left(\mathrm{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_1) \cdots \mathrm{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_p)\right)$$

Independent !

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmin}} \left\| \mathbf{Y} - \hat{\mathbf{Y}}_\eta \right\|^2_{\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\operatorname{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\operatorname{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \left(\mathrm{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_1) \cdots \mathrm{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_p)\right) \quad \text{Independent !}$$

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\mathrm{argmin}} \sum_{l=1}^{p} \left\| \mathbf{Y}_l - [\hat{\mathbf{Y}}_\eta]_l \right\|^2_{\mathbf{V}^{-1}} = \underset{\eta \in \mathcal{S}_K^{PI}}{\mathrm{argmin}} \left\| \mathbf{Y} - \hat{\mathbf{Y}}_\eta \right\|^2_{F,\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\mathrm{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{\mathbf{V}^{-1}} \left(1 + \frac{\mathrm{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1} \right) \right\}$$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \left(\text{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_1) \cdots \text{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_p)\right)$$

Independent !

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta\in\mathcal{S}_K^{PI}}{\text{argmin}} \sum_{l=1}^{p} \left\| \mathbf{Y}_l - [\hat{\mathbf{Y}}_\eta]_l \right\|^2_{\mathbf{V}^{-1}} = \underset{\eta\in\mathcal{S}_K^{PI}}{\text{argmin}} \left\| \mathbf{Y} - \hat{\mathbf{Y}}_\eta \right\|^2_{F,\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0\leq K\leq K_{\max}}{\text{argmin}} \left\{ \left\| \mathbf{Y} - \hat{\mathbf{Y}}_K \right\|^2_{F,\mathbf{V}^{-1}} \left(1 + \frac{\text{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n-K-1} \right) \right\}$$

Model Selection: LINselect

$$\mathbf{Y} = \mathbf{T}\boldsymbol{\Delta} + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{MN}_{n\times p}(\mathbf{0}, \mathbf{V}, \mathbf{R})$$

Projectors

$$\hat{\mathbf{Y}}_\eta = \left(\text{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_1) \cdots \text{Proj}^{\mathbf{V}}_{S_\eta}(\mathbf{Y}_p)\right)$$

Independent !

EM Estimators

$$\hat{\mathbf{Y}}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\text{argmin}} \sum_{l=1}^{p} \left\|\mathbf{Y}_l - [\hat{\mathbf{Y}}_\eta]_l\right\|^2_{\mathbf{V}^{-1}} = \underset{\eta \in \mathcal{S}_K^{PI}}{\text{argmin}} \left\|\mathbf{Y} - \hat{\mathbf{Y}}_\eta\right\|^2_{F,\mathbf{V}^{-1}}$$

Goal

$$\hat{K} = \underset{0 \leq K \leq K_{\max}}{\text{argmin}} \left\{ \text{tr}(\hat{\mathbf{R}}_K)\left(1 + \frac{\text{pen}\left(n, K, \left|\mathcal{S}_K^{PI}\right|\right)}{n - K - 1}\right)\right\}$$

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case $\mathbf{A}$ and $\mathbf{\Sigma}$ must commute

- $\mathbf{A}$ and $\mathbf{\Sigma}$ diagonal $\rightarrow$ independent traits
  - $\hookrightarrow$ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - $\hookrightarrow$ Justification: de-correlate the traits with a pPCA
  - $\hookrightarrow$ With shifts: not justified
- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and $\mathbf{\Sigma}$ full $\rightarrow$ scOU
  - $\hookrightarrow$ Same tree re-scaling trick $\rightarrow$ scBM

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case $\mathbf{A}$ and $\mathbf{\Sigma}$ must commute

- $\mathbf{A}$ and $\mathbf{\Sigma}$ diagonal $\rightarrow$ independent traits
  - $\hookrightarrow$ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - $\hookrightarrow$ Justification: de-correlate the traits with a pPCA
  - × With shifts: not justified
- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and $\mathbf{\Sigma}$ full $\rightarrow$ scOU
  - $\hookrightarrow$ Same tree re-scaling trick $\rightarrow$ BM

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case **A** and **Σ** must commute

- **A** and **Σ** diagonal → independent traits
  - ↪ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - ↪ Justification: de-correlate the traits with a pPCA
  - × With shifts: not justified
- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and **Σ** full → scOU
  - ↪ Same tree re-scaling trick → BM

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case **A** and **Σ** must commute

- **A** and **Σ** diagonal → independent traits
  - ↪ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - ↪ Justification: de-correlate the traits with a pPCA
  - × With shifts: not justified

- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and **Σ** full → scOU
  - ↪ Same tree re-scaling trick → BM

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case $\mathbf{A}$ and $\mathbf{\Sigma}$ must commute

- $\mathbf{A}$ and $\mathbf{\Sigma}$ diagonal $\rightarrow$ independent traits
  - $\hookrightarrow$ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - $\hookrightarrow$ Justification: de-correlate the traits with a pPCA
  - $\times$ With shifts: not justified
- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and $\mathbf{\Sigma}$ full $\rightarrow$ scOU
  - $\hookrightarrow$ Same tree re-scaling trick $\rightarrow$ BM

Multivariate OU

SDE $$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Good Case **A** and **Σ** must commute

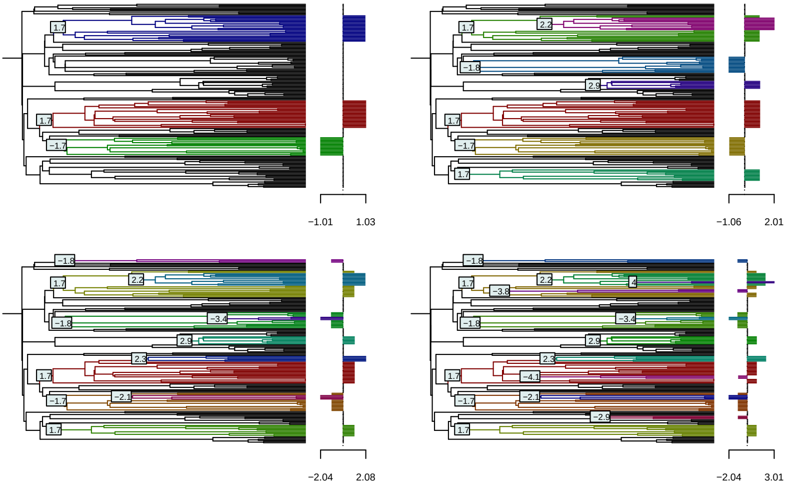
- **A** and **Σ** diagonal → independent traits
  - ↪ Ingram and Mahler (2013); Khabbazian et al. (2016)
  - ↪ Justification: de-correlate the traits with a pPCA
  - × With shifts: not justified

  \+

- $\mathbf{A} = \alpha \mathbf{I}_p$ *scalar* and **Σ** full → scOU
  - ↪ Same tree re-scaling trick → BM

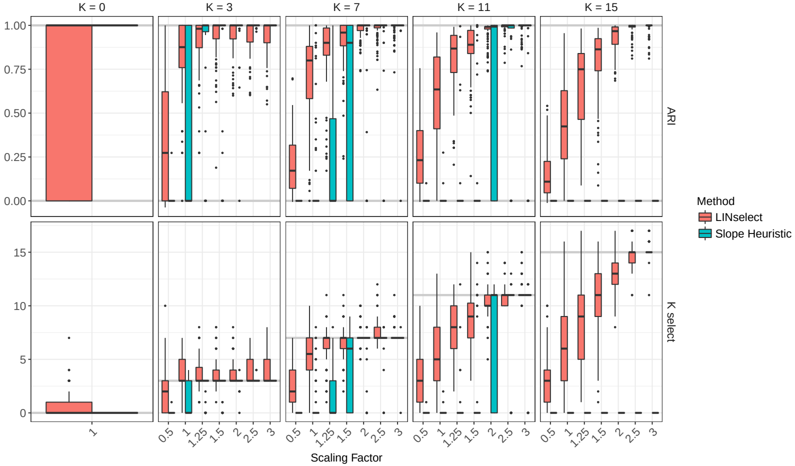
  \+

Simulations: Experimental Design



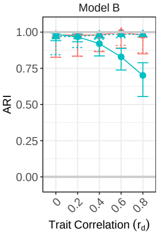
Tree Used

Simulations: Model Selection

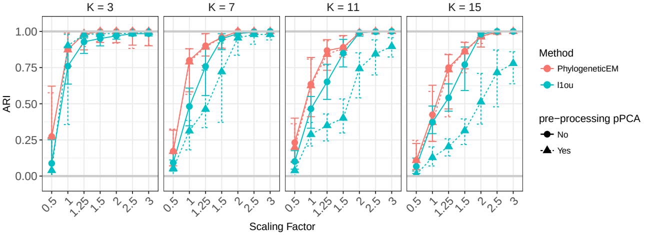


K selected and ARI

Simulations: pPCA



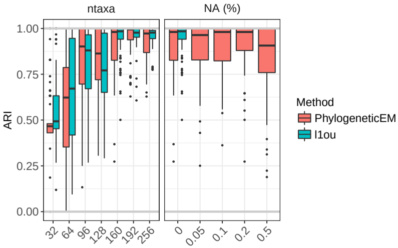
(a) Correlation



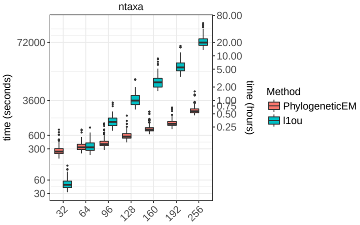
(b) Number of shifts

Pre-processing pPCA

Simulations: Scalability



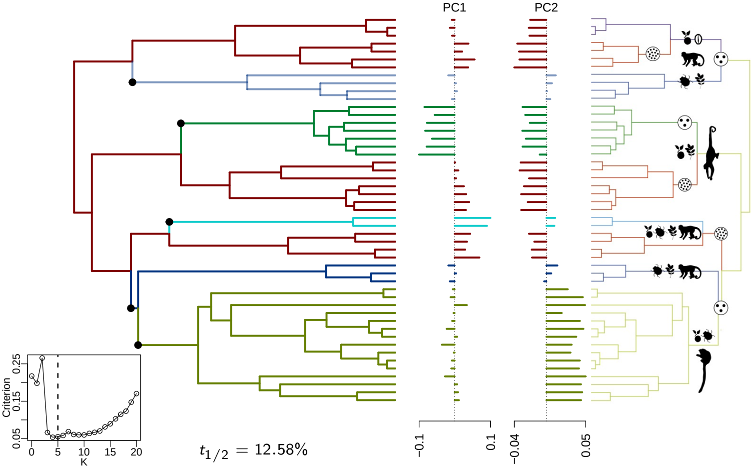
ARI



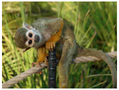
Timing

New World Monkeys

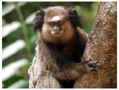
(Aristide et al., 2016)



Alouatta palliata



Saimiri sciureus



Callithrix penicillata

$t_{1/2} = 12.58\%$

$\text{Cor}[PC1, PC2] = -0.13$

Contributions

Statistical Inference, Univariate

Bastide, Mariadassou, Robin (2017). Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4), 1067–1093.

Multivariate

Bastide, Ané, Robin, Mariadassou (2017). Inference of Adaptive Shifts for Multivariate Correlated Traits. *Systematic Biology, under minor revisions*.

R package

- `PhylogeneticEM`, available on the CRAN.
  - ↦ Univariate and multivariate.
  - ↦ `Rcpp`, continuous integration, unitary tests, online doc.
  - ↦ GitHub: `https://github.com/pbastide/PhylogeneticEM`

Conclusion and Perspectives

A general inference framework for trait evolution models.

Literature
- **Model**: Felsenstein (1985); Butler and King (2004).
- **Shift detection**: Ingram and Mahler (2013); Uyeda and Harmon (2014); Khabbazian et al. (2016).

Contributions
- **Univariate**: Identifiability, EM, Model selection.
- **Multivariate**: OU with correlations.

Perspectives
- Deal with uncertainty (data, tree). + +
- Non-ultrametric trees (fossils). +
- Patterns in missing data. +
- Phylogenetic Networks. +

Bibliography

Aristide L, dos Reis SF, Machado AC, Lima I, Lopes RT, Perez SI. 2016. Brain shape convergence in the adaptive radiation of New World monkeys. *Proceedings of the National Academy of Sciences*. 113:2158–2163.

Baraud Y, Giraud C, Huet S. 2009. Gaussian model selection with an unknown variance. *Annals of Statistics*. 37:630–672.

Baudry JP, Maugis C, Michel B. 2012. Slope heuristics: overview and implementation. *Statistics and Computing*. 22:455–470.

Brault V, Baudry JP, Maugis C, Michel B. 2012. capushe: Capushe, Data-Driven Slope Estimation and Dimension Jump. R package version 1.0.

Butler MA, King AA. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist*. 164:683–695.

Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013. Phylogennomics Reveals Extensive Reticulate Evolution in Xiphophorus Fishes. *Evolution*. 67:2166–2179.

Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist*. 125:1–15.

Felsenstein J. 2004. Inferring Phylogenies.

Hansen TF. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution*. 51:1341.

Ingram T, Mahler DL. 2013. SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*. 4:416–425.

Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*. 7:811–824.

Massart P. 2007. *Concentration Inequalities and Model Selection*. 1896.

Semple C, Steel M. 2003. Phylogenetics. Oxford University Press, oxford lec edition.

Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*. 12:e1005896.

Uyeda JC, Harmon LJ. 2014. A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Systematic Biology*. 63:902–918.

Thank you for listening

pbastide.github.io

Appendices

Xiphophorus Fish Dataset (Cui et al., 2013)

X. Montezumae

Xiphophorus Fish Dataset

(Cui et al., 2013)

X. Montezumae

Two traits

- Sword index
- Female preference

Xiphophorus Fish Dataset (Cui et al., 2013)

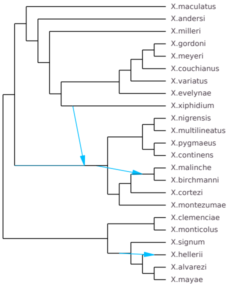
X. Montezumae

Two traits

- Sword index
- Female preference

Problem There are hybrids !

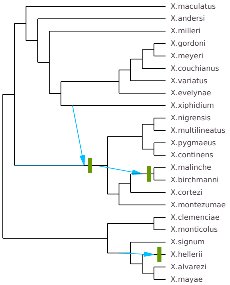
Phylogenetic "Networks" (Solís-Lemus and Ané, 2016)



X. Couchianus

X. Montezumae

Phylogenetic "Networks" (Solís-Lemus and Ané, 2016)



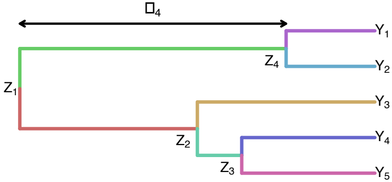
X. Couchianus

X. Montezumae

Question:

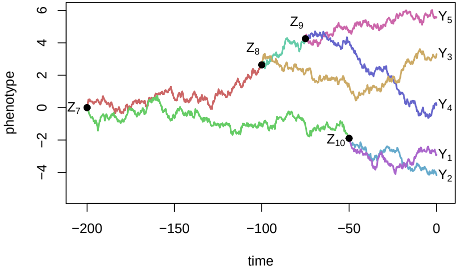
- Can we see the effects of ancestral transgressive evolution ?

Shifted BM on a Network



Known network.

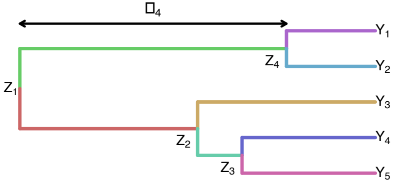
Only tip values observed.



Brownian Motion:

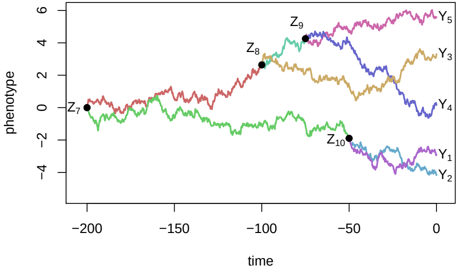
$$\mathbb{C}\text{ov}\left[Y_1; Y_2\right] = \sigma^2 \ell_4$$

Shifted BM on a Network



Known network.

Only tip values observed.



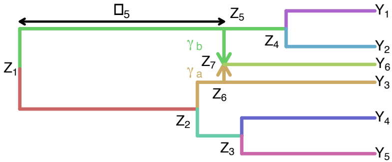
Brownian Motion:

$$V_{ij}^{\text{tree}} = \sum_{e \in p_i \cap p_j} \ell_e$$

Sum over shared edges.

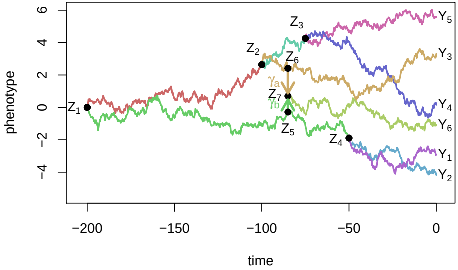
p_i: path from root to tip i

Shifted BM on a Network



Known network.

Only tip values observed.

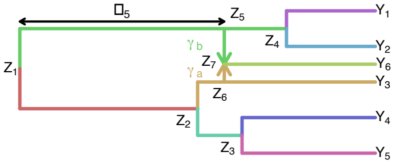


Brownian Motion:

$$Z_7 = \gamma_a Z_6 + \gamma_b Z_5$$

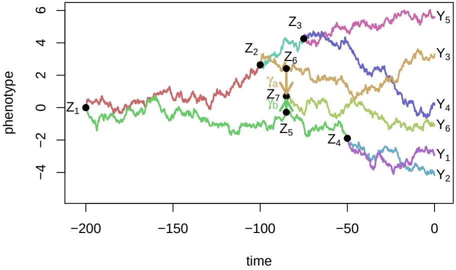
$$\gamma_a + \gamma_b = 1$$

Shifted BM on a Network



Known network.

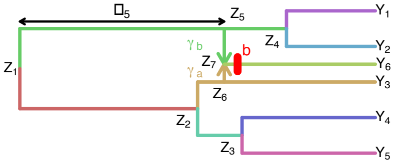
Only tip values observed.



Brownian Motion:

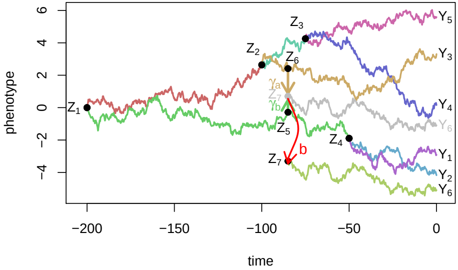
$$V_{ij}^{net} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e \right) \left(\prod_{e \in p_j} \gamma_e \right) \sum_{e \in p_i \cap p_j} \ell_e$$

Shifted BM on a Network



Known network.

Only tip values observed.

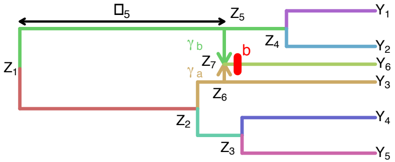


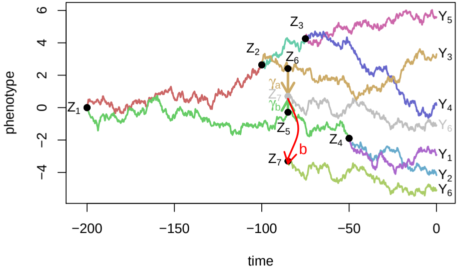
Brownian Motion:

$$Z_7 = \gamma_a Z_6 + \gamma_b Z_5 + b$$

b : Transgressive evolution.

Shifted BM on a Network





Known network.

Only tip values observed.

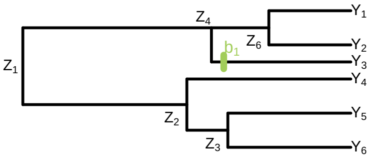
Goal: Test for transgressive evolution.

Brownian Motion:

$$Z_7 = \gamma_a Z_6 + \gamma_b Z_5 + b$$

b : Transgressive evolution.

Linear Regression Model

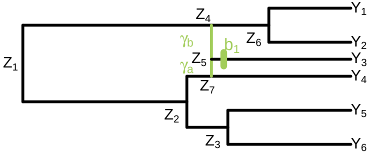


$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_6 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T\Delta} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \mu \\ \mu + b_1 \\ \mu \\ \mu \\ \mu \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{array}{c} \begin{array}{ccccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Z_6 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Y} = \mathbf{T\Delta} + \sigma \mathbf{E}^{\text{net}}$$

Linear Regression Model



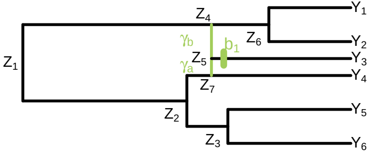
$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T\Delta} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \mu \\ \mu + b_1 \\ \mu \\ \mu \\ \mu \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 & Z_6 & Z_7 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \gamma_b & 1 & \cdot & \gamma_a & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Y} = \mathbf{T\Delta} + \sigma \mathbf{E}^{\text{net}}$$

$$T_{ij} = \sum_{p \in \mathcal{P}_{j \to i}} \prod_{e \in p} \gamma_e$$

Linear Regression Model



$$\mathbf{\Delta} = \begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{T\Delta} = \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{pmatrix} \mu \\ \mu \\ \mu + b_1 \\ \mu \\ \mu \\ \mu \end{pmatrix}$$

$$\mathbf{T} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & \mathbf{Z_5} & Z_6 & Z_7 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{array} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \gamma_b & \mathbf{1} & \cdot & \gamma_a & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Y} = \mu \mathbf{1} + \mathbf{N}\mathbf{b} + \sigma \mathbf{E}^{\text{net}}$$

$$T_{ij} = \sum_{p \in \mathcal{P}_{j \to i}} \prod_{e \in p} \gamma_e$$

Transgressive Evolution: Testing Effect(s)

Model:

$$\mathbf{Y} = \mu\mathbf{1} + \mathbf{N}\mathbf{b} + \sigma^2\mathbf{E} \quad , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

Tests:

$\mathcal{H}_0$:	No TE	$\mathbf{b} = \mathbf{0}$
$\mathcal{H}_1$:	TE with one single effect	$\mathbf{b} = b.\mathbf{1}$
$\mathcal{H}_2$:	TE with heterogeneous effects	$\mathbf{b} \in \mathbb{R}^h$

Fisher:

$$F_{10} \sim \mathcal{F}_{1,n-2}\left(\Delta_{10}(b, \sigma^2)\right)$$
$$F_{21} \sim \mathcal{F}_{h-1,n-h-1}\left(\Delta_{21}(\mathbf{b}, \sigma^2)\right)$$

+

Xiphophorus fishes (Cui et al., 2013)

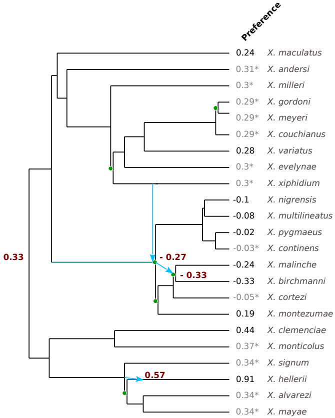
X. Montezumae

Sword Index

No evidence for TE.

Xiphophorus fishes (Cui et al., 2013)

X. Montezumae



Sword Index

No evidence for TE.

Female Preference

Heterogeneous TE.

Xiphophorus fishes (Cui et al., 2013)

X. Montezumae

Sword Index

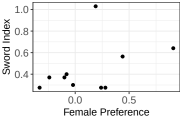
No evidence for TE.

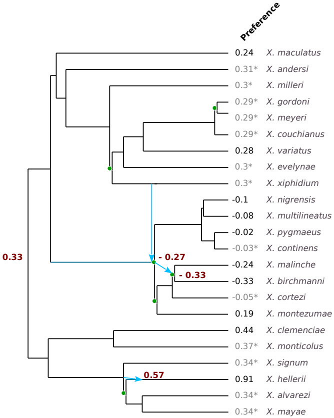
Female Preference

Heterogeneous TE.

Regression

Positive correlation (Non-significant).





Contributions

Preprint

Bastide, Solís-Lemus, Kriebel, Sparks, Ané (submitted). Phylogenetic Comparative Methods for Phylogenetic Networks with Reticulations.

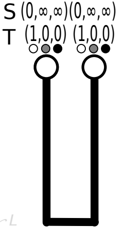
`Julia` package

Solís-Lemus, **Bastide**, Ané (2017). `PhyloNetworks`: a package for phylogenetic networks. *Molecular Biology and Evolution*, msx235.

- $\mapsto$ Network inference and use.
- $\mapsto$ Continuous integration, unitary tests, online doc.

back

Cardinal of Equivalence Classes



Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1 \leq p \leq K}{\operatorname{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \ldots p_L) \in \mathcal{K}_k^1 \times \ldots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1, \ldots p_L) \in \mathcal{K}_k^1 \times \ldots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$

Termination Sum on the root vector

back

Cardinal of Equivalence Classes

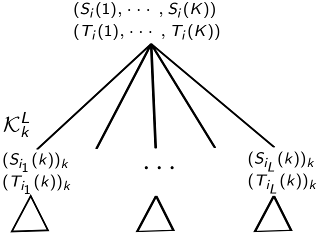
Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1\le p\le K}{\text{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \ldots p_L) \in \mathcal{K}_k^1 \times \ldots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1,\ldots p_L)\in\mathcal{K}_k^1\times\ldots\times\mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l\in\mathcal{K}_k^l} T_{i_l}(p_l)$$



Termination Sum on the root vector

back

Cardinal of Equivalence Classes

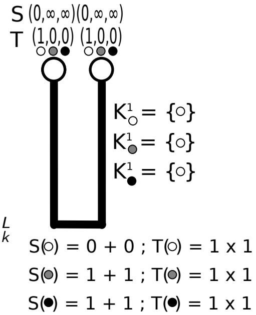
Initialization For tips

Propagation

$$\mathcal{K}_k^l = \underset{1 \le p \le K}{\operatorname{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$



Termination Sum on the root vector

back

Cardinal of Equivalence Classes

Initialization For tips

Propagation

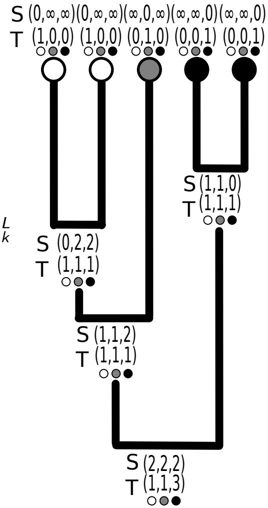
$$\mathcal{K}_k^l = \underset{1\le p\le K}{\text{argmin}} \left\{ S_{i_l}(p) + \mathbb{I}\{p \neq k\} \right\}$$

$$S_i(k) = \sum_{l=1}^{L} S_{i_l}(p_l) + \mathbb{I}\{p_l \neq k\} \ , \ \forall (p_1, \dots p_L) \in \mathcal{K}_k^1 \times \dots \times \mathcal{K}_k^L$$

$$T_i(k) = \sum_{(p_1,\dots p_L)\in \mathcal{K}_k^1\times\dots\times\mathcal{K}_k^L} \prod_{l=1}^{L} T_{i_l}(p_l) = \prod_{l=1}^{L} \sum_{p_l \in \mathcal{K}_k^l} T_{i_l}(p_l)$$

Termination Sum on the root vector

back



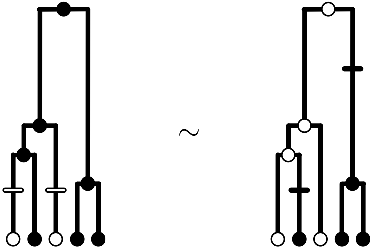
Linking Shifts and Clustering

Assumption "No Homoplasy": 1 shift $=$ 1 new color

Proposition "K shifts $\iff$ $K+1$ clusters"

Linking Shifts and Clustering

Assumption "No Homoplasy": 1 shift = 1 new color

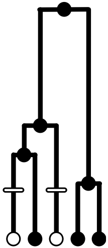


The No Homoplasy hypothesis is not respected.

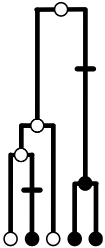
Proposition "K shifts $\iff$ $K+1$ clusters"

Linking Shifts and Clustering

Assumption "No Homoplasy": 1 shift = 1 new color



$\sim$

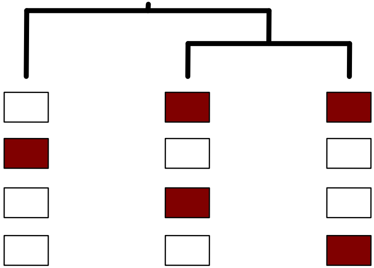


The No Homoplasy hypothesis is not respected.

Proposition "K shifts $\iff$ $K+1$ clusters"

Definitions

- $\mathcal{T}$ a rooted tree with n tips
- $N_K^{(\mathcal{T})} = |\mathcal{C}_K|$ the number of possible partitions of the tips in K clusters
- $A_K^{(\mathcal{T})}$ the number of possible *marked* partitions



Partitions in two groups for a binary tree with 3 tips

Difference between $N_2^{(\mathcal{T}_3)}$ and $A_2^{(\mathcal{T}_3)}$:

- $N_2^{(\mathcal{T}_3)} = 3$: partitions 1 and 2 are equivalent
- $A_2^{(\mathcal{T}_3)} = 4$: one marked color ("white = ancestral state")

General Formula (Binary Case)

If $\mathcal{T}$ is a binary tree, consider T_ℓ and $\mathcal{T}_r$ the left and right sub-trees of $\mathcal{T}$. Then:

$$\begin{cases} N_K^{(\mathcal{T})} = \sum_{k_1+k_2=K} N_{k_1}^{(\mathcal{T}_\ell)} N_{k_2}^{(\mathcal{T}_r)} + \sum_{k_1+k_2=K+1} A_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} \\ A_K^{(\mathcal{T})} = \sum_{k_1+k_2=K} A_{k_1}^{(\mathcal{T}_\ell)} N_{k_2}^{(\mathcal{T}_r)} + N_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} + \sum_{k_1+k_2=K+1} A_{k_1}^{(\mathcal{T}_\ell)} A_{k_2}^{(\mathcal{T}_r)} \end{cases}$$

We get:

$$N_{K+1}^{(\mathcal{T})} = N_{K+1}^{(n)} = \binom{2n-2-K}{K} \quad \text{and} \quad A_{K+1}^{(\mathcal{T})} = A_{K+1}^{(n)} = \binom{2n-1-K}{K}$$

Recursion Formula (General Case)

If we are at a node defining a tree $\mathcal{T}$ that has p daughters, with sub-trees $\mathcal{T}_1, \ldots, \mathcal{T}_p$, then we get the following recursion formulas:

$$\begin{cases} N_K^{(\mathcal{T})} = \displaystyle\sum_{\substack{k_1+\cdots+k_p=K \\ k_1,\ldots,k_p\geq 1}} \prod_{i=1}^{p} N_{k_i}^{(\mathcal{T}_i)} + \sum_{\substack{I\subset [\![1,p]\!] \\ |I|\geq 2}} \sum_{\substack{k_1+\cdots+k_p=K+|I|-1 \\ k_1,\ldots,k_p\geq 1}} \prod_{i\in I} A_{k_i}^{(\mathcal{T}_i)} \prod_{i\notin I} N_{k_i}^{(\mathcal{T}_i)} \\ A_K^{(\mathcal{T})} = \displaystyle\sum_{\substack{I\subset [\![1,p]\!] \\ |I|\geq 1}} \sum_{\substack{k_1+\cdots+k_p=K+|I|-1 \\ k_1,\ldots,k_p\geq 1}} \prod_{i\in I} A_{k_i}^{(\mathcal{T}_i)} \prod_{i\notin I} N_{k_i}^{(\mathcal{T}_i)} \end{cases}$$

No general formula. The result depends on the topology of the tree.

back

Cholesky Decomposition

The problem is:

$$\hat{\boldsymbol{\Delta}} = \underset{\boldsymbol{\Delta}}{\operatorname{argmin}} \left\{ \|\mathbf{Y} - \mathbf{T}\boldsymbol{\Delta}\|^2_{\mathbf{V}^{-1}} + \lambda \left|\boldsymbol{\Delta}_{-\mathbf{1}}\right|_1 \right\}$$

Cholesky decomposition of $\mathbf{V}$:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T \ , \ \mathbf{L} \text{ a lower triangular matrix}$$

Then:

$$\|\mathbf{Y} - \mathbf{T}\boldsymbol{\Delta}\|^2_{\mathbf{V}^{-1}} = \left\|\mathbf{L}^{-1}\mathbf{Y} - \mathbf{L}^{-1}\mathbf{T}\boldsymbol{\Delta}\right\|^2$$

And if $\mathbf{Y}' = \mathbf{L}^{-1}\mathbf{Y}$ and $\mathbf{T}' = \mathbf{L}^{-1}\mathbf{T}$, the problem becomes:

$$\hat{\boldsymbol{\Delta}} = \underset{\boldsymbol{\Delta}}{\operatorname{argmin}} \left\{ \left\|\mathbf{Y}' - \mathbf{T}'\boldsymbol{\Delta}\right\|^2 + \lambda \left|\boldsymbol{\Delta}{-1}\right|_1 \right\}$$

Gauss Lasso

Let $\hat{m}_\lambda$ be the set of selected variables (including the root). Then:

$$\hat{\mathbf{\Delta}}^{\text{Gauss}} = \Pi_{\hat{F}_\lambda}(\mathbf{Y}') \text{ with } \hat{F}_\lambda = \text{Span}\{\mathbf{T}'_j : j \in \hat{m}_\lambda\}$$

back

Goal and Notations

Data A process on a tree with the following structure:

$$\forall j > 1, \quad X_j | X_{\mathrm{pa}(j)} \sim \mathcal{N}\left(m_j(X_{\mathrm{pa}(j)}) = q_j X_{\mathrm{pa}(j)} + r_j, \sigma_j^2\right)$$

$$\text{BM:} \begin{cases} q_j = 1 \\ r_j = \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \\ \sigma_j^2 = \ell_j \sigma^2 \end{cases} \qquad \text{OU:} \begin{cases} q_j = e^{-\alpha \ell_j} \\ r_j = \beta^{\mathrm{pa}(j)}(1 - e^{-\alpha \ell_j}) + \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \left(1 - e^{-\alpha(1-\nu_k)\ell_j}\right) \\ \sigma_j^2 = \dfrac{\sigma^2}{2\alpha}(1 - e^{-2\alpha \ell_j}) \end{cases}$$

Goal Compute the following quantities, at every node j:

$$\mathbb{V}\mathrm{ar}^{(h)}\left[Z_j \mid \mathbf{Y}\right], \mathbb{C}\mathrm{ov}^{(h)}\left[Z_j, Z_{\mathrm{pa}(j)} \mid \mathbf{Y}\right], \mathbb{E}^{(h)}[Z_j \mid \mathbf{Y}]$$

Upward

Goal Compute for a vector of tips, given their common ancestor:

$$f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j; a) = A_j(\mathbf{Y}^j)\Phi_{M_j(\mathbf{Y}^j), S_j^2(\mathbf{Y}^j)}(a)$$

Initialization For tips: $f_{Y_i|Y_i}(Y_i; a) = \Phi_{Y_i,0}(a)$

Propagation

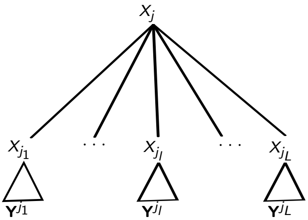
$$f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j; a) = \prod_{l=1}^{L} f_{\mathbf{Y}^{j_l}|X_j}(\mathbf{Y}^{j_l}; a)$$

$$f_{\mathbf{Y}^{j_l}|X_j}(\mathbf{Y}^{j_l}; a) = \int_{\mathbb{R}} f_{\mathbf{Y}^{j_l}\big|X_{j_l}}(\mathbf{Y}^{j_l}; b) f_{X_{j_l}\big|X_j}(b; a) db$$

Root Node and Likelihood At the root:

$$f_{X_1|\mathbf{Y}}(a; \mathbf{Y}) \propto f_{\mathbf{Y}|X_1}(\mathbf{Y}; a) f_{X_1}(a)$$

$$\begin{cases} \mathbb{V}\mathrm{ar}\left[X_1 \mid \mathbf{Y}\right] = \left(\dfrac{1}{\gamma^2} + \dfrac{1}{S_1^2(\mathbf{Y})}\right)^{-1} \\ \mathbb{E}\left[X_1 \mid \mathbf{Y}\right] = \mathbb{V}\mathrm{ar}\left[X_1 \mid \mathbf{Y}\right]\left(\dfrac{\mu}{\gamma^2} + \dfrac{M_1(\mathbf{Y})}{S_1^2(\mathbf{Y})}\right) \end{cases}$$



Downward

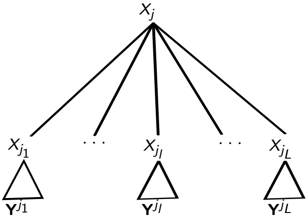
Compute $E_j = \mathbb{E}\left[X_j \mid \mathbf{Y}\right]$, $V_j^2 = \mathbb{V}\mathrm{ar}\left[X_j \mid \mathbf{Y}\right]$, $C_{j,\mathrm{pa}(j)}^2 = \mathbb{C}\mathrm{ov}\left[X_j; X_{\mathrm{pa}(j)} \mid \mathbf{Y}\right]$

Initialization Last step of Upward.

Propagation

$$f_{X_{\mathrm{pa}(j)},X_j|\mathbf{Y}}(a,b;\mathbf{Y}) = f_{X_{\mathrm{pa}(j)}|\mathbf{Y}}(a;\mathbf{Y}) f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}}(b;a,\mathbf{Y})$$

$$\begin{aligned} f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}}(b;a,\mathbf{Y}) &= f_{X_j|X_{\mathrm{pa}(j)},\mathbf{Y}^j}(b;a,\mathbf{Y}^j) \\ &\propto f_{X_j|X_{\mathrm{pa}(j)}}(b;a) f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j;b) \end{aligned}$$



Formulas

Upward

$$\begin{cases} S_j^2(\mathbf{Y}^j) = \left(\sum_{l=1}^{L} \frac{q_{j_l}^2}{S_{j_l}^2(\mathbf{Y}^{j_l}) + \sigma_{j_l}^2} \right)^{-1} \\ M_j(\mathbf{Y}^j) = S_j^2(\mathbf{Y}^j) \sum_{l=1}^{L} q_{j_l} \frac{M_{j_l}(\mathbf{Y}^{j_l}) - r_{j_l}}{S_{j_l}^2(\mathbf{Y}^{j_l}) + \sigma_{j_l}^2} \end{cases}$$

Downward

$$\begin{cases} C_{j,\mathrm{pa}(j)}^2 = q_j \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} V_{\mathrm{pa}(j)}^2 \\ E_j = \frac{S_j^2(\mathbf{Y}^j)(q_j E_{\mathrm{pa}(j)} + r_j) + \sigma_j^2 M_j(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} \\ V_j^2 = \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} \left(\sigma_j^2 + p_j^2 \frac{S_j^2(\mathbf{Y}^j)}{S_j^2(\mathbf{Y}^j) + \sigma_j^2} V_{\mathrm{pa}(j)}^2 \right) \end{cases}$$

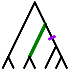
back

M Step: Segmentation

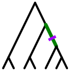
$$C_j(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \,\middle|\, Y\right] - r_j - s_j \Delta_j\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0 = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \,\middle|\, Y\right]\right)^2$$



$$C_j^1(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \,\middle|\, Y\right] - s_j \Delta_j\right)^2$$



Algorithm:

1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\boldsymbol{\Delta}) = 0$

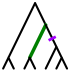
back

M Step: Segmentation

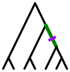
$$C_j(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j\Delta_j\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0 = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] \qquad\right)^2$$



$$C_j^1(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j\Delta_j\right)^2$$



Algorithm:

1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\boldsymbol{\Delta}) = 0$

back

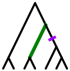
M Step: Segmentation

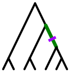
$$C_j(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j\Delta_j\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0 = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] \qquad\quad \right)^2$$

$$C_j^1(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j\Delta_j\right)^2$$





Algorithm:

1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\boldsymbol{\Delta}) = 0$

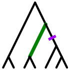
back

M Step: Segmentation

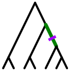
$$C_j(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j\Delta_j\right)^2$$

BM : $r_j = 0$, each cost is independent.

$$C_j^0 = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] \qquad\right)^2$$



$$C_j^1(\boldsymbol{\Delta}) = \sigma_j^{-2}\left(\mathbb{E}\left[X_j \mid Y\right] - q_j\mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j\Delta_j\right)^2$$



Algorithm:

1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\boldsymbol{\Delta}) = 0$

back

M Step: Segmentation

$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

OU : $r_j = \beta^{\mathrm{pa}(j)}$, a cost depends on all its parents.

- Exact minimization: too costly.
- Need of an heuristic.
- Idea: rewrite as a least square:

$$\|D - AU\Delta\|^2$$

  with D a vector of size $n + m$, A a diagonal matrix of size $n + m$, Δ the vector of shifts and U the incidence matrix of the tree.
- Then use Stepwise selection or LASSO.
- Other idea: binary segmentation.

back

Model Selection with Unknown Variance

Theorem (Baraud et al. (2009))

Under the following setting:

$$Y' = \mathbb{E}\left[Y'\right] + \gamma E' \quad \textit{with} \quad E' \sim \mathcal{N}(0, I_n) \quad \textit{and} \quad \mathcal{S}' = \{S'_\eta, \eta \in \mathcal{M}\}$$

If $D_\eta = \mathrm{Dim}(S'_\eta)$, $N_\eta = n - D_\eta \geq 7$, $\max(L_\eta, D_\eta) \leq \kappa n$, *with* $\kappa < 1$, *and:*

$$\Omega' = \sum_{\eta \in \mathcal{M}} (D_\eta + 1) e^{-L_\eta} < +\infty$$

$$\textit{If:} \quad \hat{\eta} = \underset{\eta \in \mathcal{M}}{\mathrm{argmin}} \left\| Y' - \hat{Y}'_\eta \right\|^2 \left(1 + \frac{\mathrm{pen}(\eta)}{N_\eta}\right)$$

$$\textit{with:} \quad \mathrm{pen}(\eta) = \mathrm{pen}_{A,\mathcal{L}}(\eta) = A \frac{N_\eta}{N_\eta - 1} \mathrm{EDkhi}[D_\eta + 1, N_\eta - 1, e^{-L_\eta}] \quad , \quad A > 1$$

$$\textit{Then:}\ \mathbb{E}\left[\frac{\left\|\mathbb{E}\left[Y'\right] - \hat{Y}'_{\hat{\eta}}\right\|^2}{\gamma^2}\right] \leq C(A, \kappa) \left[\inf_{\eta \in \mathcal{M}} \left\{ \frac{\left\|\mathbb{E}\left[Y'\right] - Y'_\eta\right\|^2}{\gamma^2} + \max(L_\eta, D_\eta) \right\} + \Omega' \right]$$

IID Framework ($\alpha = 0$)

$$\text{Assume} \quad K_\eta = D_\eta - 1 \leq p - 1 \leq n - 8, \quad \forall \eta \in \mathcal{M}$$

Then:

$$\begin{aligned}
\Omega' &= \sum_{\eta \in \mathcal{M}} (D_\eta + 1) e^{-L_\eta} = \sum_{\eta \in \mathcal{M}} (K_\eta + 2) e^{-L_\eta} \\
&= \sum_{K=0}^{p-1} \left| \mathcal{S}_K^{PI} \right| (K+2) e^{-L_K} = \sum_{K=0}^{p-1} \left| \mathcal{S}_K^{PI} \right| (K+2) e^{-(\log \left| \mathcal{S}_K^{PI} \right| + 2\log(K+2))} \\
&= \sum_{K=0}^{p-1} \frac{1}{K+2} \leq \log(p) \leq \log(n)
\end{aligned}$$

And:

$$L_K \leq \log \binom{n+m-1}{K} + 2\log(K+2) \leq K \log(n+m-1) + 2(K+1) \leq p(2 + \log(2n-2))$$

Hence, if $p \leq \min\left(\frac{\kappa n}{2+\log(2)+\log(n)}, n-7\right)$, then $\max(L_\eta, D_\eta) \leq \kappa n$ for any $\eta \in \mathcal{M}$.

Non-IID Framework ($\alpha \neq 0$)

Cholesky decomposition: $V = LL^T \quad Y' = L^{-1}Y \quad s' = L^{-1}s \quad E' = L^{-1}E$

$$Y' = \mathbb{E}\left[Y'\right] + \gamma E', \text{ with: } E' \sim \mathcal{N}(0, I_n)$$

$$S'_\eta = L^{-1}S_\eta, \quad \hat{Y}'_\eta = \mathrm{Proj}_{S'_\eta}\, Y' = \underset{a' \in S'_\eta}{\mathrm{argmin}} \left\| Y - La' \right\|_V^2 = L^{-1}\hat{Y}_\eta$$

$$\left\|\mathbb{E}\left[Y\right] - \hat{Y}_{\hat{\eta}}\right\|_V^2 = \left\|\mathbb{E}\left[Y'\right] - \hat{Y}'_{\hat{\eta}}\right\|^2, \quad \left\|Y - \hat{Y}_\eta\right\|_V^2 = \left\|Y' - \hat{Y}'_\eta\right\|^2$$

$$\mathrm{Crit}_{MC}(\eta) = \left\|Y' - \hat{Y}'_\eta\right\|^2 \left(1 + \frac{\mathrm{pen}_{A,\mathcal{L}}(\eta)}{N_\eta}\right) = \left\|Y - \hat{Y}_\eta\right\|_V^2 \left(1 + \frac{\mathrm{pen}_{A,\mathcal{L}}(\eta)}{N_\eta}\right)$$

back

Phylogenetic PCA with shifts

Model $\mathbf{Y}$ size $n \times p$ (n observations, p traits), Brownian

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{E} \quad \mathrm{vec}(\mathbf{E}) \sim \mathcal{N}(\mathbf{0}, \mathbf{R} \otimes \mathbf{C})$$

Empirical Mean and Variance

$$\bar{\mathbf{Y}}^T = \tilde{\mathbf{C}}\mathbf{Y} \qquad \bar{\boldsymbol{\mu}}^T = \mathbb{E}\left[\bar{\mathbf{Y}}^T\right] = \tilde{\mathbf{C}}\boldsymbol{\mu} \quad \text{with} \quad \tilde{\mathbf{C}} = (\mathbf{1}_n^T\mathbf{C}^{-1}\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{C}^{-1}$$

$$\hat{\mathbf{R}} = \frac{1}{n-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}^T)^T\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}^T)$$

Bias on $\hat{\mathbf{R}}$

$$\mathbb{E}\left[\hat{\mathbf{R}}\right] = \mathbf{R} + \frac{1}{n-1}\mathbf{G}^T\mathbf{C}^{-1}\mathbf{G} \quad \text{with} \quad \mathbf{G} = (\boldsymbol{\mu} - \mathbf{1}_n\bar{\boldsymbol{\mu}}^T)$$

Phylogenetic PCA : Scores

Rotation

$$\hat{\mathbf{R}} = \frac{1}{n-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^2\hat{\mathbf{V}}^T$$

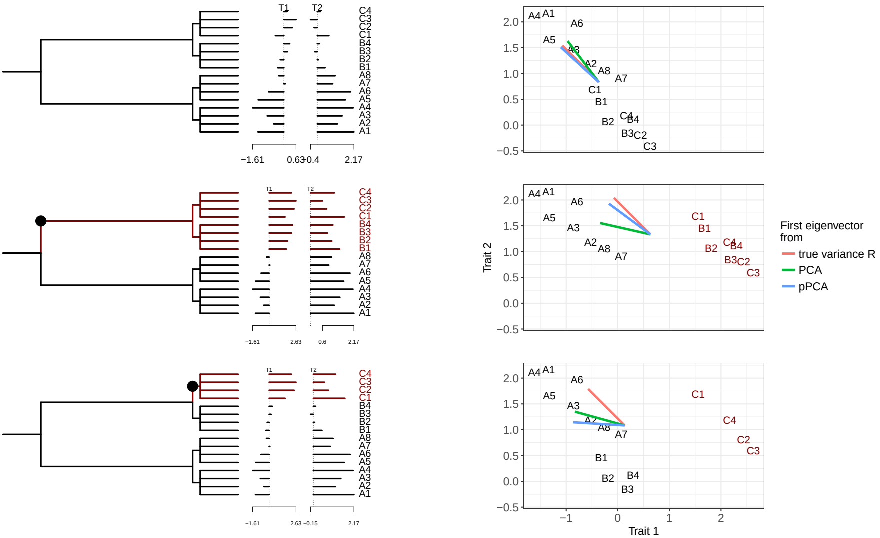
$\mapsto$ If $\hat{\mathbf{R}}$ is biased, then $\hat{\mathbf{V}}$ is the wrong rotation.

Scores

$$\mathbf{S} = (\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}^T)\hat{\mathbf{V}}$$

$\mapsto$ The scores are not decorrelated.

Phylogenetic PCA : Examples



back

OU Model

SDE $\mathbf{A}$ ($p \times p$) *selection strength*

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

OU Model

SDE $\mathbf{A}$ $(p \times p)$ *selection strength*

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\text{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i}\mathbf{\Gamma} e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})}\left(\int_0^{t_{ij}} e^{-\mathbf{A}v}\mathbf{\Sigma}\mathbf{\Sigma}^T e^{-\mathbf{A}^T v} dv\right) e^{-\mathbf{A}^T (t_j - t_{ij})}$$

OU Model

SDE $\mathbf{A}$ $(p \times p)$ *selection strength* $\in \mathcal{S}_n^{++}$

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\begin{aligned}\mathbb{C}\text{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = & \; e^{-\mathbf{A}t_i}\mathbf{\Gamma}e^{-\mathbf{A}^T t_j} \\ & + e^{-\mathbf{A}(t_i - t_{ij})}\left(\int_0^{t_{ij}} e^{-\mathbf{A}v}\mathbf{\Sigma}\mathbf{\Sigma}^T e^{-\mathbf{A}^T v} dv\right) e^{-\mathbf{A}^T(t_j - t_{ij})}\end{aligned}$$

OU Model

SDE $\mathbf{A}$ $(p \times p)$ *selection strength* $\in \mathcal{S}_n^{++}$

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\text{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i}\mathbf{\Gamma}e^{-\mathbf{A}^T t_j} - e^{-\mathbf{A}t_i}\mathbf{S}e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})}\mathbf{S}e^{-\mathbf{A}^T(t_j - t_{ij})}$$

Stationary Variance

$$\mathbf{S} = \mathbf{P}\left(\left[\frac{1}{\lambda_q + \lambda_r}\right]_{1 \leq q,r \leq p} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right)\mathbf{P}^T$$

OU Model

SDE $\mathbf{A}$ $(p \times p)$ *selection strength* $\in \mathcal{S}_n^{++}$

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\mathrm{ov}\left[\mathbf{X}_i ; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i} \mathbf{\Gamma} e^{-\mathbf{A}^T t_j} - e^{-\mathbf{A}t_i} \mathbf{S} e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})} \mathbf{S} e^{-\mathbf{A}^T (t_j - t_{ij})}$$

Stationary Variance

$$\mathbf{S} = \mathbf{P} \left(\left[\frac{1}{\lambda_q + \lambda_r} \right]_{1 \leq q,r \leq p} \odot \mathbf{P}^{-1} \mathbf{R} \mathbf{P}^{-T} \right) \mathbf{P}^T$$

Incomplete Data Representation

$$\mathbf{X}_j \mid \mathbf{X}_{\mathrm{pa}(j)} \sim \mathcal{N}\left(e^{-\mathbf{A}\ell_j} \mathbf{X}_{\mathrm{pa}(j)} + (\mathbf{I}_p - e^{-\mathbf{A}\ell_j}) \beta_j, \mathbf{\Upsilon}_i = \mathbf{S} - e^{-\mathbf{A}\ell_j} \mathbf{S} e^{-\mathbf{A}^T \ell_j} \right)$$

OU Model

SDE $\mathbf{A} = \alpha I_p$ *scalar*

$$d\mathbf{W}(t) = -\mathbf{A}(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\mathrm{ov}\left[\mathbf{X}_i ; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i}\mathbf{\Gamma}e^{-\mathbf{A}^T t_j} - e^{-\mathbf{A}t_i}\mathbf{S}e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})}\mathbf{S}e^{-\mathbf{A}^T(t_j - t_{ij})}$$

Stationary Variance

$$\mathbf{S} = \mathbf{P}\left(\left[\frac{1}{\lambda_q + \lambda_r}\right]_{1 \leq q,r \leq p} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right)\mathbf{P}^T$$

OU Model

SDE $\mathbf{A} = \alpha I_p$ *scalar*

$$d\mathbf{W}(t) = -\alpha(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\mathrm{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i}\mathbf{\Gamma}e^{-\mathbf{A}^T t_j} - e^{-\mathbf{A}t_i}\mathbf{S}e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})}\mathbf{S}e^{-\mathbf{A}^T (t_j - t_{ij})}$$

Stationary Variance

$$\mathbf{S} = \mathbf{P}\left(\left[\frac{1}{\lambda_q + \lambda_r}\right]_{1 \leq q,r \leq p} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right)\mathbf{P}^T$$

OU Model

SDE $\mathbf{A} = \alpha I_p$ *scalar*

$$d\mathbf{W}(t) = -\alpha(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\text{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = e^{-\mathbf{A}t_i}\mathbf{\Gamma}e^{-\mathbf{A}^T t_j} - e^{-\mathbf{A}t_i}\mathbf{S}e^{-\mathbf{A}^T t_j} + e^{-\mathbf{A}(t_i - t_{ij})}\mathbf{S}e^{-\mathbf{A}^T(t_j - t_{ij})}$$

Stationary Variance

$$\mathbf{S} = \frac{1}{2\alpha}\mathbf{R}$$

OU Model

SDE $\mathbf{A} = \alpha I_p$ *scalar*

$$d\mathbf{W}(t) = -\alpha(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\text{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = \frac{1}{2\alpha} e^{-2\alpha h}\left(e^{2\alpha t_{ij}} - 1\right)\mathbf{R}$$

Stationary Variance

$$\mathbf{S} = \frac{1}{2\alpha}\mathbf{R}$$

OU Model

SDE $\mathbf{A} = \alpha I_p$ *scalar*

$$d\mathbf{W}(t) = -\alpha(\mathbf{W}(t) - \beta(t))dt + \mathbf{\Sigma} d\mathbf{B}_t$$

Covariances

$$\mathbb{C}\mathrm{ov}\left[\mathbf{X}_i; \mathbf{X}_j\right] = \frac{1}{2\alpha} e^{-2\alpha h} \left(e^{2\alpha t_{ij}} - 1\right) \mathbf{R}$$

Stationary Variance

$$\mathbf{S} = \frac{1}{2\alpha} \mathbf{R}$$

$\mapsto$ Re-scaling trick.

back

TE: Single Effect

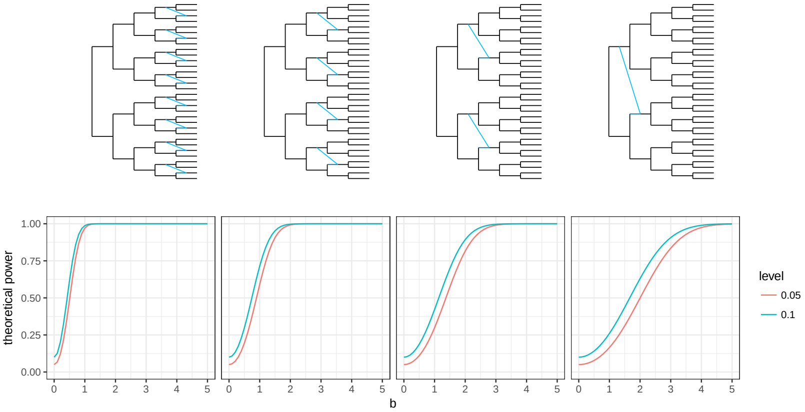
Model: $\mathbf{Y} = \mu \mathbf{1} + b\bar{\mathbf{N}} + \sigma^2 \mathbf{E} \quad , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$

Test: $\mathcal{H}_0 : b = 0$

Stat.:
$$F_{10} = \frac{\left\|\mathbf{Y} - \mathrm{Proj}_{\mathbf{1}}\, \mathbf{Y}\right\|^2_{\mathbf{V}^{-1}} - \left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{1}\ \bar{\mathbf{N}}]}\, \mathbf{Y}\right\|^2_{\mathbf{V}^{-1}}}{\left\|\mathbf{Y} - \mathrm{Proj}_{[\mathbf{1}\ \bar{\mathbf{N}}]}\, \mathbf{Y}\right\|^2_{\mathbf{V}^{-1}}} \frac{n - r_{[\mathbf{1}\ \bar{\mathbf{N}}]}}{r_{[\mathbf{1}\ \bar{\mathbf{N}}]} - r_{\mathbf{1}}}$$

$$\sim \mathcal{F}\left(1, n-2, \frac{b^2}{2\sigma^2} \left\|(\mathbf{I} - \mathrm{Proj}_{\mathbf{1}})\bar{\mathbf{N}}\right\|^2_{\mathbf{V}^{-1}}\right)$$

TE: Single Effect



Detection Power $(\sigma^2 = 1)$

TE: Several Effects

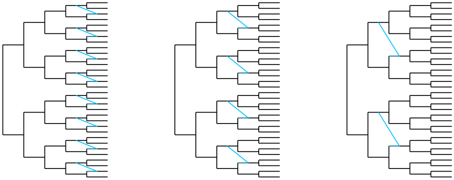
Model: $$\mathbf{Y} = \mu\mathbf{1} + b\bar{\mathbf{N}} + \mathbf{N}\mathbf{d} + \sigma^2\mathbf{E} \quad , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

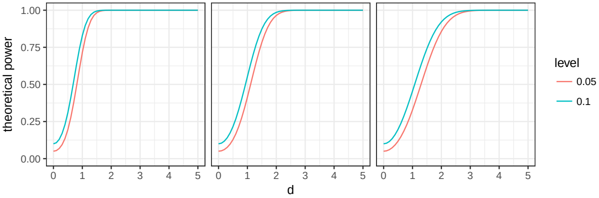
Test: $$\mathcal{H}_1 : d_1 = \cdots = d_h = 0$$

Stat.: $$F_{21} = \frac{\left\| \mathbf{Y} - \mathrm{Proj}_{[\mathbf{1}\ \bar{\mathbf{N}}]}\mathbf{Y} \right\|^2_{\mathbf{V}^{-1}} - \left\| \mathbf{Y} - \mathrm{Proj}_{[\mathbf{1}\ \mathbf{N}]}\mathbf{Y} \right\|^2_{\mathbf{V}^{-1}}}{\left\| \mathbf{Y} - \mathrm{Proj}_{[\mathbf{1}\ \mathbf{N}]}\mathbf{Y} \right\|^2_{\mathbf{V}^{-1}}} \frac{n - r_{[\mathbf{1}\ \mathbf{N}]}}{r_{[\mathbf{1}\ \mathbf{N}]} - r_{[\mathbf{1}\ \bar{\mathbf{N}}]}}$$

$$\sim \mathcal{F}\left(h-1, n-h-1, \frac{1}{2\sigma^2}\left\| (\mathbf{I} - \mathrm{Proj}_{[\mathbf{1}\ \bar{\mathbf{N}}]})\mathbf{N}\mathbf{d} \right\|^2_{\mathbf{V}^{-1}}\right)$$

TE: Several Effects





Detection Power $(\sigma^2 = 1)$

back

Simulations Design

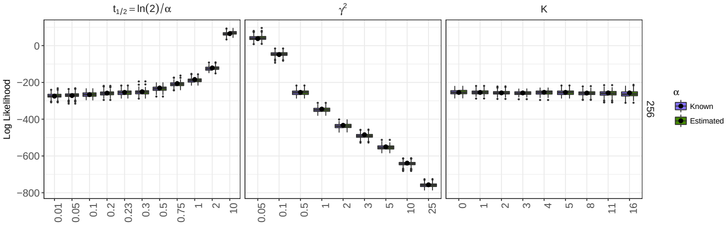
(Uyeda and Harmon, 2014)

- Topology of the tree fixed (unit height, $\lambda = 0.1$, with 64, 128, 256 taxa).
- Initial optimal value fixed: $\beta_0 = 0$
- One "base" scenario $\alpha_b = 3$, $\gamma_b^2 = 0.5$, $K_b = 5$.
- $\alpha \in \log(2)/\{0.01, 0.05, 0.1, 0.2, 0.23, 0.3, 0.5, 0.75, 1, 2, 10\}$.
- $\gamma^2 \in \{0.3, 0.6, 3, 6, 12, 18, 30, 60, 150\}/(2\alpha_b)$.
- $K \in \{0, 1, 2, 3, 4, 5, 8, 11, 16\}$.
- Shifts values $\sim \frac{1}{2}\mathcal{N}(4,1) + \frac{1}{2}\mathcal{N}(-4,1)$
- Shifts randomly placed at regular intervals separated by 0.1 unit length.
- $n = 200$ repetitions: 16200 configurations.

CPU time on cluster MIGALE (Jouy-en-Josas):

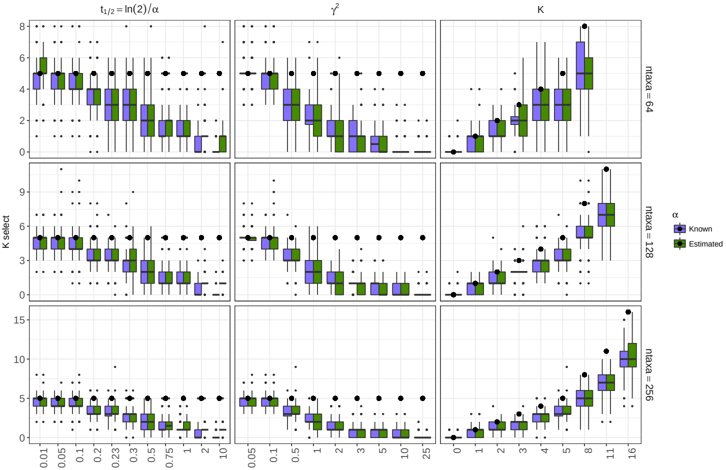
- α known: 6 minutes per estimation (66 days in total).
- α unknown: 52 minutes per estimation (570 days in total).

Log-Likelihood

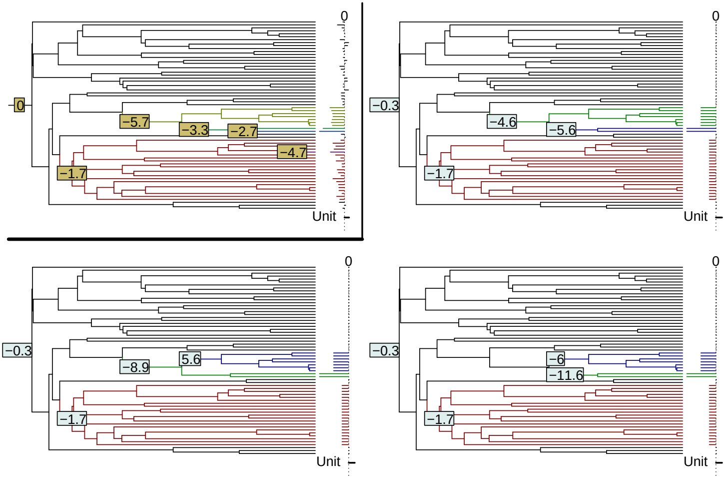


Log likelihood for a tree with 256 tips. Solid black dots are the median of the log likelihood for the true parameters.

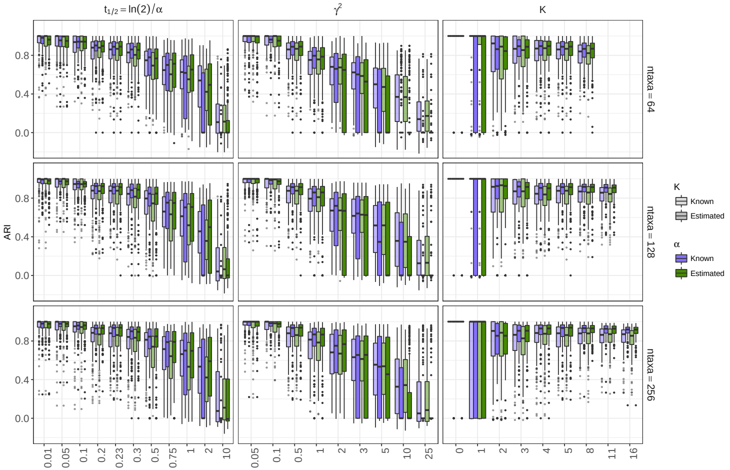
Number of Shifts



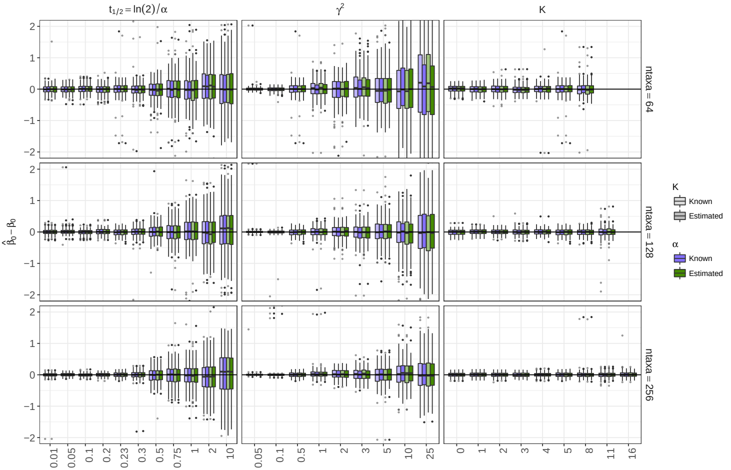
One Example


0
−5.7
−3.3
−2.7
−4.7
−1.7
0
Unit
−0.3
−4.6
−5.6
−1.7
0
Unit
−0.3
−8.9
5.6
−1.7
0
Unit
−0.3
−6
−11.6
−1.7
0
Unit

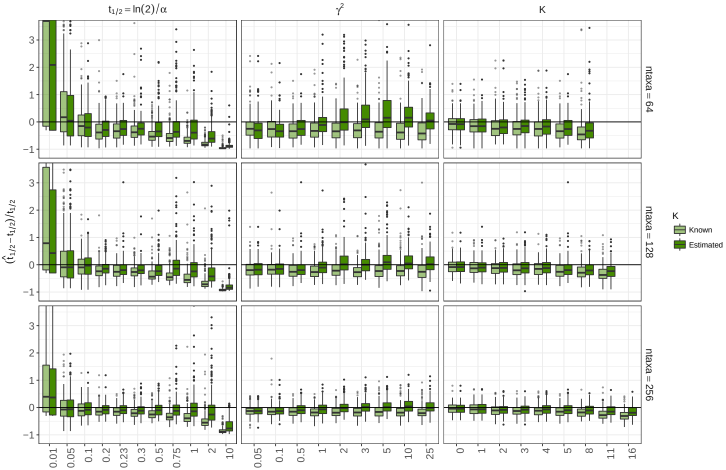
Adjusted Rand Index



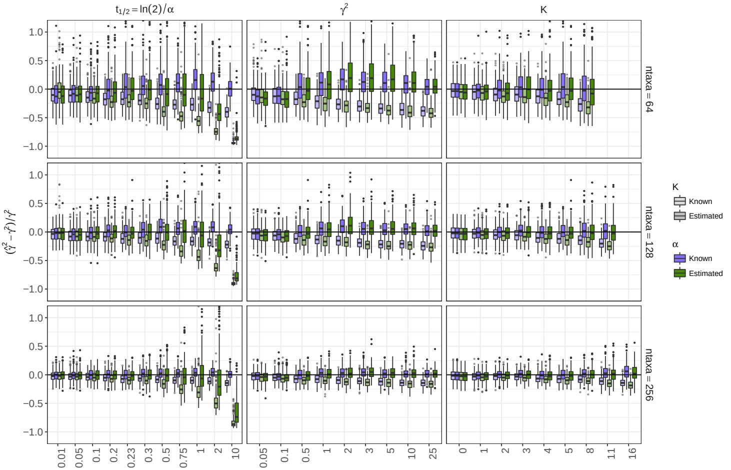
Parameters: β_0



Parameters: α



Parameters: γ^2



Exploration

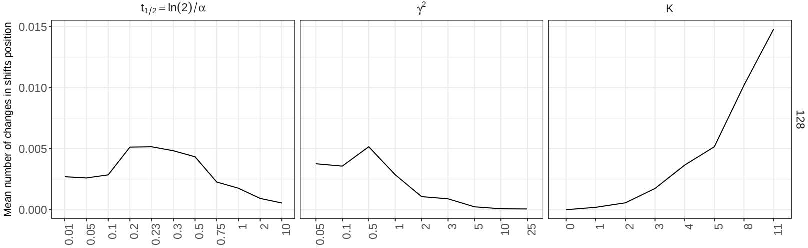
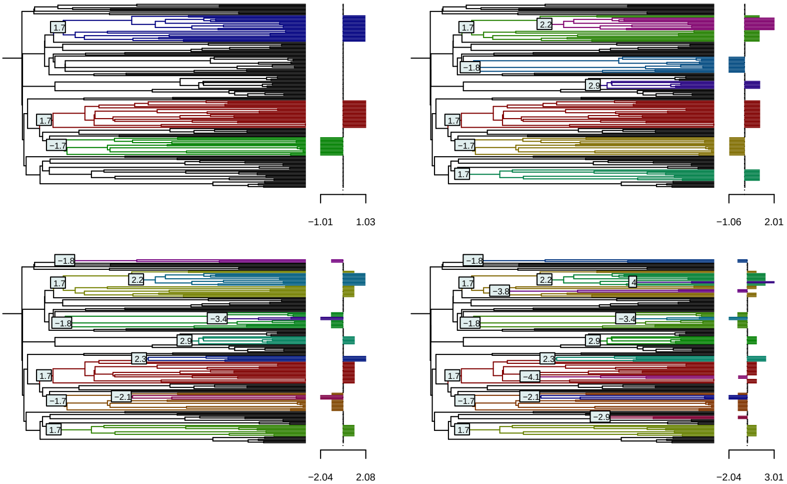


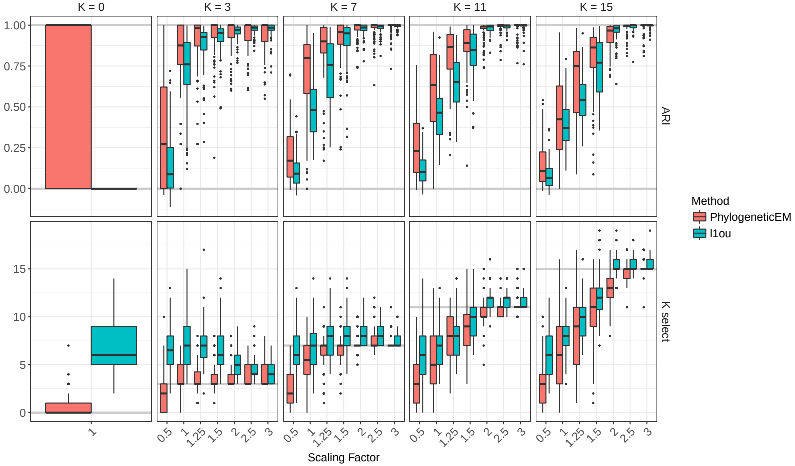
Figure: Mean number changes in the shifts positions during the EM algorithm. Null means that the initial shifts were kept all along.

Simulations: Experimental Design



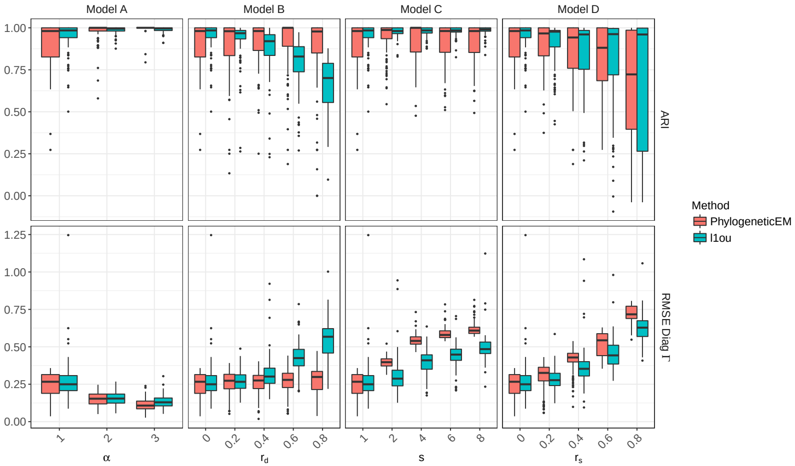
Tree Used

Simulations: Model Selection



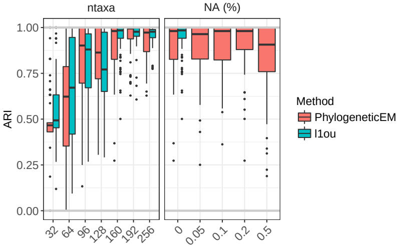
K selected and ARI

Simulations: Model Selection and Estimation

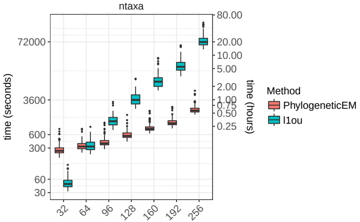


ARI and $\hat{\Gamma}$

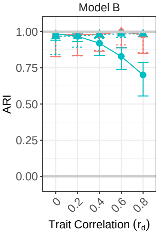
Simulations: Scalability



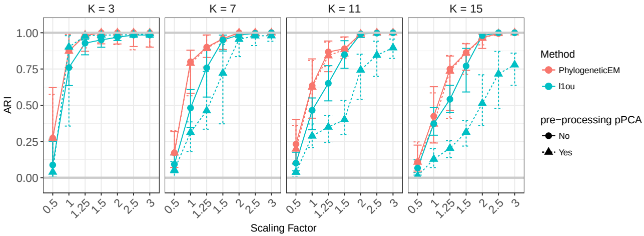
ARI



Timing



(a) Correlation



(b) Number of shifts

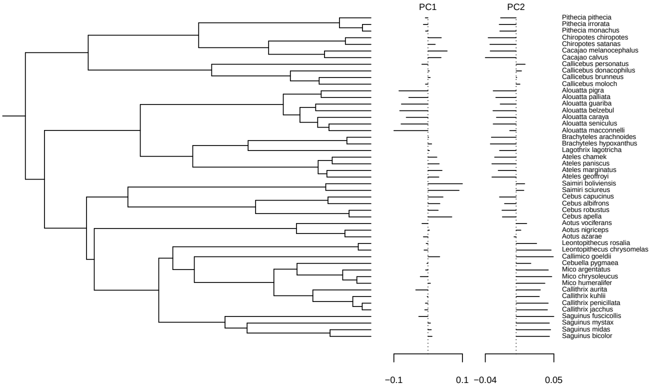
Pre-processing pPCA

back

Monkey Dataset (Aristide et al., 2016)

```r
data(monkeys)

plot(params_BM(p=2), data = monkeys$dat, phylo = monkeys$phy, show.tip.label = TRUE)
```



Analysis

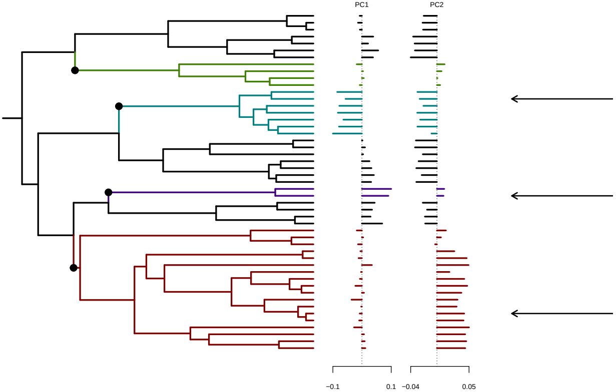
We use function `PhyloEM`:

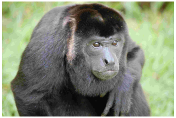
```
res <- PhyloEM(Y_data = monkeys$dat,          ## data
               phylo = monkeys$phy,           ## phylogeny
               process = "scOU",              ## scalar OU
               K_max = 10,                    ## maximal number of shifts
               nbr_alpha = 4,                 ## number of alpha values
               parallel_alpha = TRUE,         ## parallelize on alpha values
               Ncores = 2)
```

Then plot the solution selected by the default method:

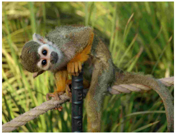
```
plot(res, edge.width = 2)
```

Result

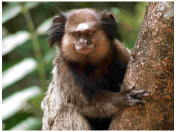




Alouatta palliata



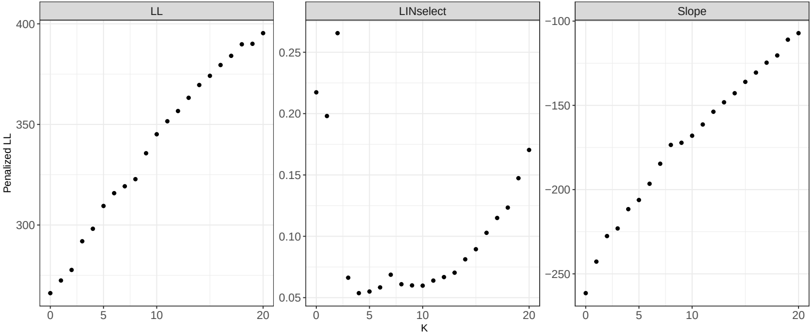
Saimiri sciureus



Callithrix penicillata

Model Selection

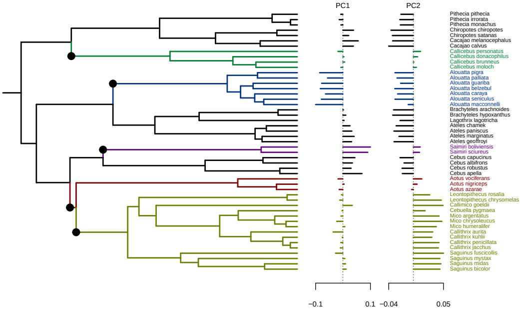
Solution with $K = 5$ seems to be a good solution too.



Solution for $K = 5$

```
plot(res, params = params_process(res, K = 5), edge.width = 2, show.tip.label = TRUE)
```

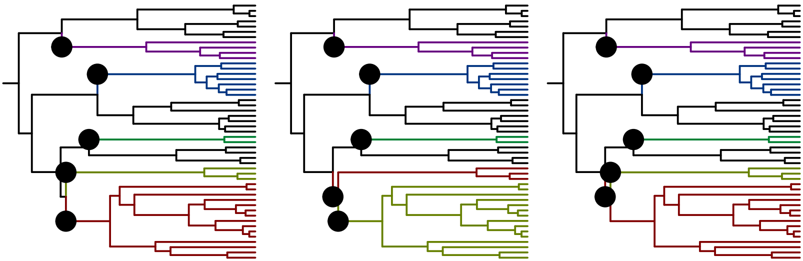
```
## Warning in params_process.PhyloEM(res, K = 5):  There are several equivalent solutions for
this shift position.
```



Solution for $K = 5$

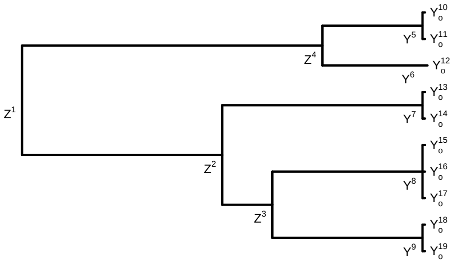
```
params_5 <- params_process(res, K = 5)
eq_shifts <- equivalent_shifts(monkeys$phy, params_5)
```

```
plot(eq_shifts)
```



back

Measurement Error (Felsenstein, 2008)



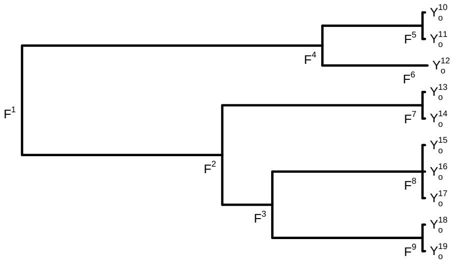
$$\mathbf{X} = \begin{cases} \mathbf{Y}_o & : \text{observed traits} \\ \mathbf{Y} & : \text{latent tips} \\ \mathbf{Z} & : \text{latent nodes} \end{cases}$$

$$\mathbf{X}^1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{X}^j \,\big|\, \mathbf{X}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{X}^{\text{pa}(j)} + \boldsymbol{\Delta}^j, \ell_j \mathbf{R}\right) \qquad \text{nodes } 2 \leq j \leq m+n$$

$$\mathbf{Y}_o^i \,\big|\, \mathbf{Y}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Y}^{\text{pa}(i)}, \mathbf{P}\right) \qquad \text{observations } m+n+1 \leq i \leq m+n+n_o.$$

Factor Analysis (Tolkoff et al., 2017)

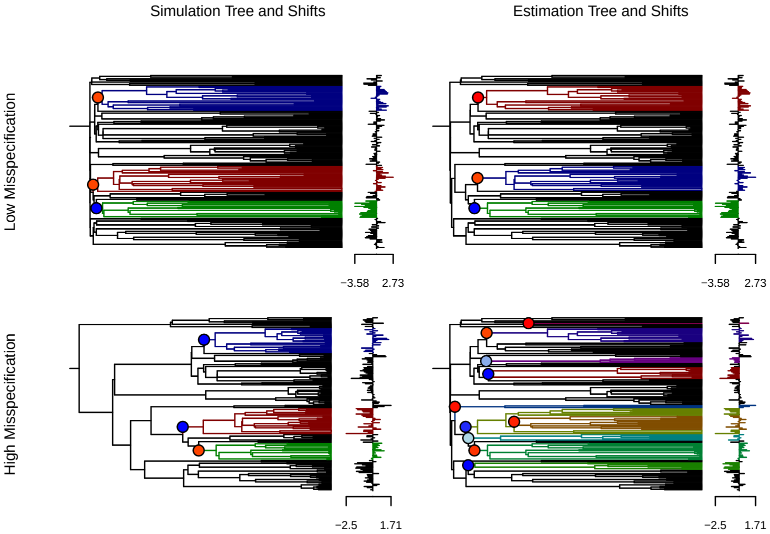


$$\begin{cases} \mathbf{Y}_o : \text{observed traits} & \text{size } p \\ \mathbf{F} : \text{latent } \textbf{features} & \text{size } q < p \end{cases}$$

$$\begin{aligned}
\mathbf{F}^1 &\sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{F}}, \boldsymbol{\Gamma}_{\mathbf{F}}\right) && \text{root} \\
\mathbf{F}^j \mid \mathbf{F}^{\mathrm{pa}(j)} &\sim \mathcal{N}\left(\mathbf{X}^{\mathrm{pa}(j)} + \boldsymbol{\Delta}^j, \ell_j \mathbf{I}_q\right) && \text{nodes } 2 \leq j \leq m+n \\
\mathbf{Y}_o^i \mid \mathbf{F}^{\mathrm{pa}(i)} &\sim \mathcal{N}\left(\mathbf{F}^{\mathrm{pa}(i)}\mathbf{L}, \mathbf{P}\right) && \text{observations } m+n+1 \leq i \leq m+n+n_o.
\end{aligned}$$

back

Tree Misspecification



Identifiability

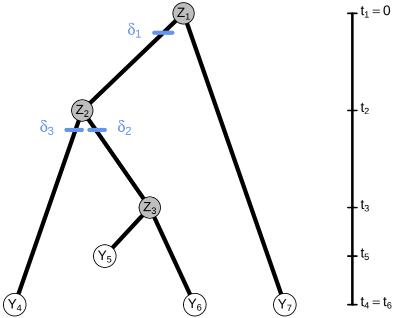


Figure: A non-ultrametric tree, with a "non parsimonious" solution on the left that cannot be reduced to the "parsimonious" one on the right for an OU.

Patterns in Missing Data (Rubin, 1976)

$\mathbf{Y}(n \times p)$	data
$\mathbf{M}(n \times p)$	missing data indicator
$p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,)$	sampling law

Patterns in Missing Data

(Rubin, 1976)

$$\begin{array}{rl} \mathbf{Y}(n \times p) & \text{data} \\ \mathbf{M}(n \times p) & \text{missing data indicator} \\ p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) & \text{sampling law} \end{array}$$

EM:

$$\mathbb{E}\left[\log p_{\boldsymbol{\theta},\psi}(\mathbf{Y}_{\text{obs}}, \mathbf{M}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right]$$
$$= \mathbb{E}\left[\log p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}\,) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] + \mathbb{E}\left[\log p_{\boldsymbol{\theta}}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right]$$

Patterns in Missing Data

(Rubin, 1976)

$$\begin{array}{rl} \mathbf{Y}(n \times p) & \text{data} \\ \mathbf{M}(n \times p) & \text{missing data indicator} \\ p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) & \text{sampling law} \end{array}$$

EM:

$$\begin{aligned} &\mathbb{E}\left[\log p_{\theta,\psi}(\mathbf{Y}_{\text{obs}}, \mathbf{M}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] \\ &\quad = \mathbb{E}\left[\log p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}\,) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] + \mathbb{E}\left[\log p_\theta(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] \end{aligned}$$

MCAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\mathbf{M})$

Patterns in Missing Data

(Rubin, 1976)

$$\begin{array}{rl} \mathbf{Y}(n \times p) & \text{data} \\ \mathbf{M}(n \times p) & \text{missing data indicator} \\ p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) & \text{sampling law} \end{array}$$

EM:

$$\begin{aligned} &\mathbb{E}\left[\log p_{\theta,\psi}(\mathbf{Y}_{\text{obs}}, \mathbf{M}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] \\ &\quad = \mathbb{E}\left[\log p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}\,) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] + \mathbb{E}\left[\log p_\theta(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] \end{aligned}$$

MCAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\mathbf{M})$

MAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}\,)$

Patterns in Missing Data

(Rubin, 1976)

$$\begin{array}{rl} \mathbf{Y}(n \times p) & \text{data} \\ \mathbf{M}(n \times p) & \text{missing data indicator} \\ p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) & \text{sampling law} \end{array}$$

EM:

$$\mathbb{E}\left[\log p_{\theta,\psi}(\mathbf{Y}_{\text{obs}}, \mathbf{M}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right]$$
$$= \mathbb{E}\left[\log p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}\,) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right] + \mathbb{E}\left[\log p_\theta(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}, \mathbf{Z}) \mid \mathbf{Y}_{\text{obs}}, \mathbf{M}\right]$$

MCAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\mathbf{M})$

MAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}\,)$

NMAR: $p_\psi(\,\mathbf{M} \mid \mathbf{Y}\,) = p_\psi(\,\mathbf{M} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}}\,)$

back