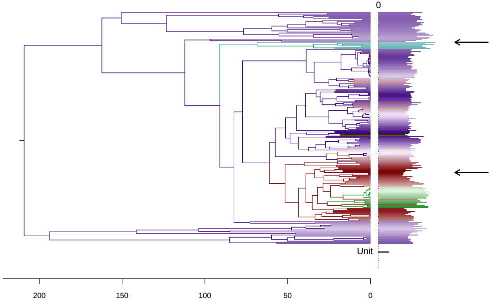
Change-point Detection on a Tree to Study Evolutionary Adaptation from Present-day Species

Cécile Ané[1,2], Paul Bastide[3,4], Mahendra Mariadassou[4], Stéphane Robin[3]

[1] Department of Statistics, University of Wisconsin–Madison, WI, 53706, USA
[2] Department of Botany, University of Wisconsin–Madison, WI, 53706, USA
[3] UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France
[4] MaIAGE, INRA, Université Paris-Saclay, 78352 Jouy-en-Josas, France

28 June 2016

Introduction



Dermochelys Coriacea

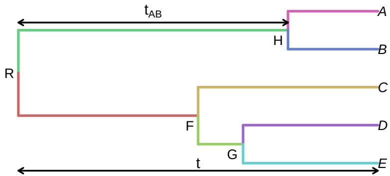
Homopus Areolatus

Turtles phylogenetic tree with habitats.
(Jaffe et al., 2011).

- How can we explain the diversity, while accounting for the phylogenetic correlations ?
- Modelling: a shifted stochastic process on the phylogeny.

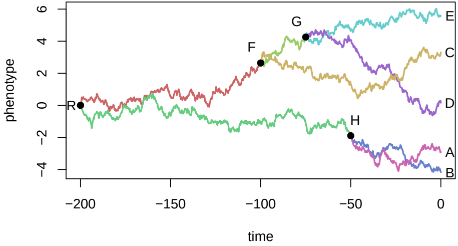
Outline

Stochastic Process on a Tree (Felsenstein, 1985)



The tree is known.

Only *tip* values are observed

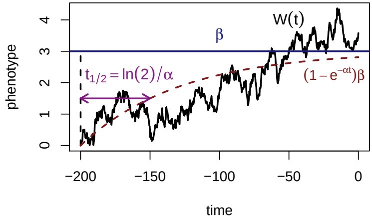


Brownian Motion:

$$\mathbb{V}\mathrm{ar}\left[A \mid R\right] = \sigma^2 t$$

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

Ornstein-Uhlenbeck Modeling (Hansen, 1997)



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

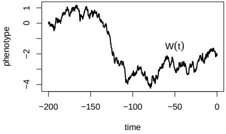
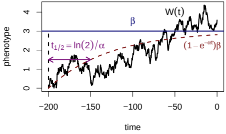
Deterministic part :

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

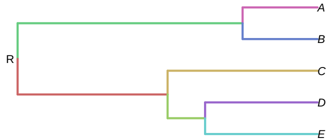
Stochastic part :

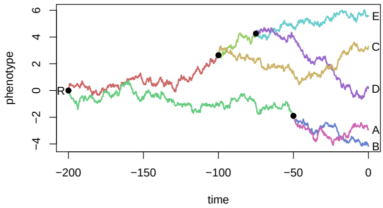
- $W(t)$: trait value (actual optimum).
- $\sigma dB(t)$: Brownian fluctuations.

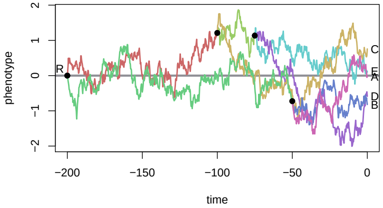
BM vs OU

<table>
<tr><th></th><th>Equation</th><th>Stationary State</th><th>Variance</th></tr>
<tr><td>
</td><td>$dW(t) = \sigma dB(t)$</td><td>None.</td><td>$\sigma_{ij} = \sigma^2 t_{ij}$</td></tr>
<tr><td>
</td><td>$dW(t) = \sigma dB(t)$
$+\alpha[\beta - W(t)]dt$</td><td>$\begin{cases} \mu = \beta \\ \gamma^2 = \dfrac{\sigma^2}{2\alpha} \end{cases}$</td><td>$\sigma_{ij} = \gamma^2 e^{-\alpha(t_i+t_j)}$
$\times (e^{2\alpha t_{ij}} - 1)$</td></tr>
</table>

Shifts







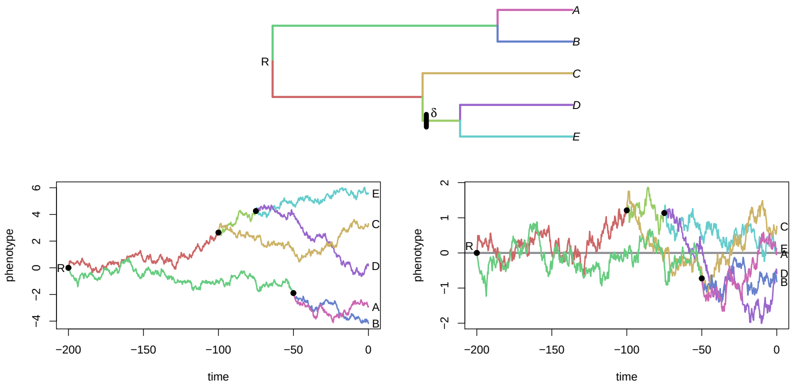
BM Shifts in the **mean**:

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

OU Shifts in the **optimal value**:

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

Shifts



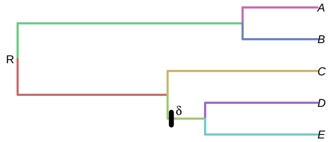
BM Shifts in the **mean**:

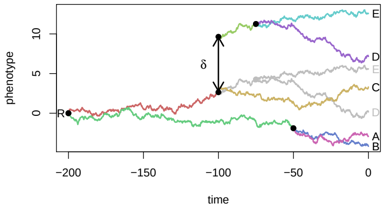
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

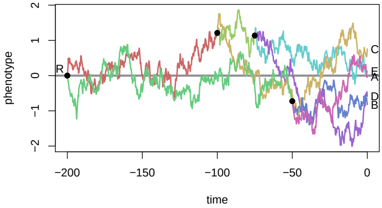
OU Shifts in the **optimal value**:

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

Shifts







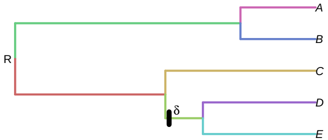
BM Shifts in the **mean**:

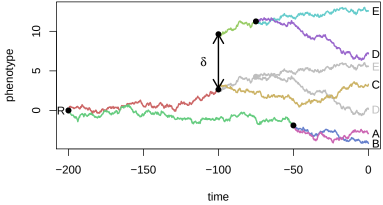
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

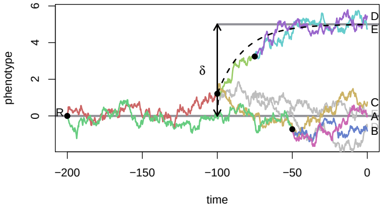
OU Shifts in the **optimal value**:

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

Shifts







BM Shifts in the **mean**:

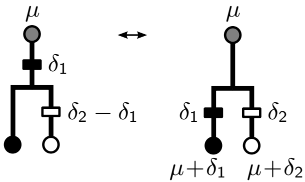
$$m_{\text{child}} = m_{\text{parent}} + \delta$$

OU Shifts in the **optimal value**:

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

Equivalencies

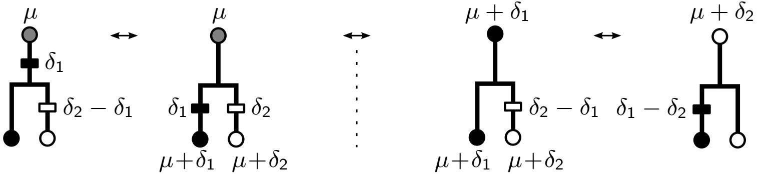
- Number of shifts K fixed, several equivalent solutions.



- Problem of over-parametrization: parsimonious configurations.

Equivalencies

- Number of shifts K fixed, several equivalent solutions.



- Problem of over-parametrization: parsimonious configurations.

Parsimonious Solution : Definition

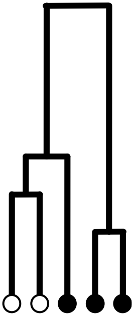
Definition (Parsimonious Allocation)

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)

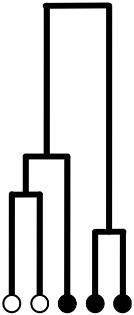
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

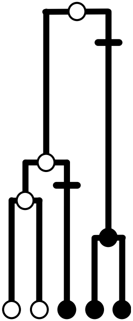


Parsimonious Solution : Definition

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.

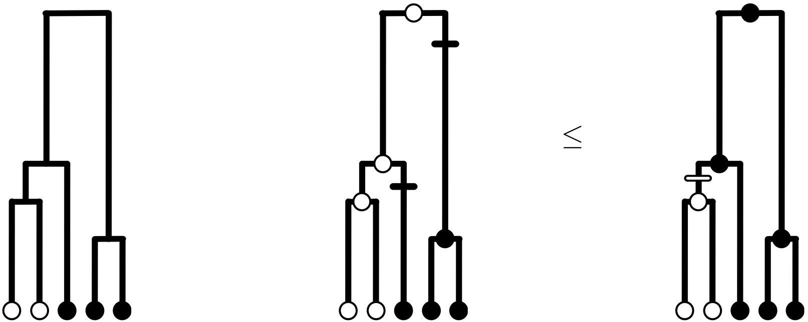




Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

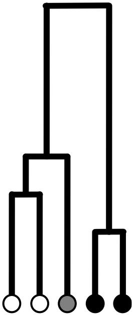
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.



Parsimonious Solution : Definition

Definition (Parsimonious Allocation)

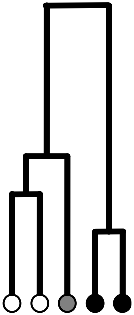
A coloring of the tips being given, a *parsimonious* allocation of the shifts is such that it has a minimum number of shifts.

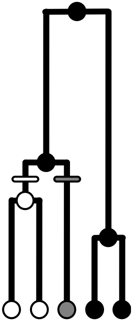


Parsimonious Solution : Definition

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.

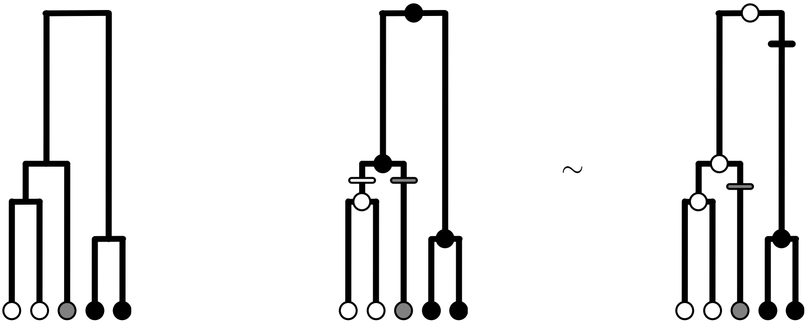




Parsimonious Solution : Definition

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)

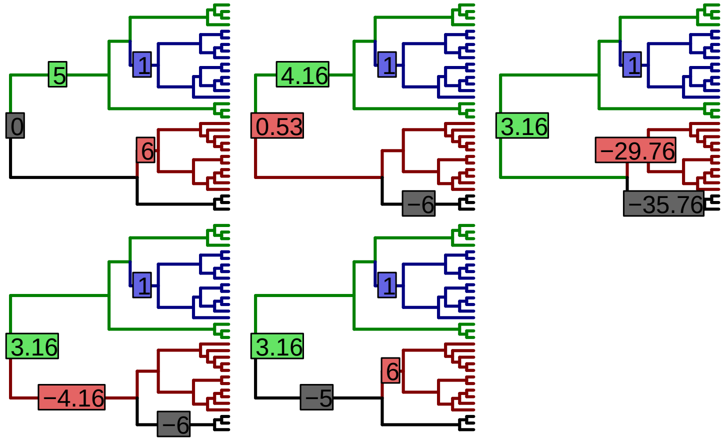
Two allocations are said to be *equivalent* (noted $\sim$) if they are both parsimonious and give the same colors at the tips.

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

Enumerate all solutions New recursive algorithm, adapted from previous ones (and implemented in R).

+

Equivalent Parsimonious Solutions for an OU Model.



Equivalent allocations and values of the shifts - OU.

Collection of Models

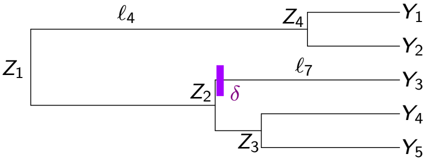
New Problem Number of Equivalence Classes: $\left|\mathcal{S}_K^{PI}\right|$?

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

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

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

EM Algorithm: number of shifts K fixed



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

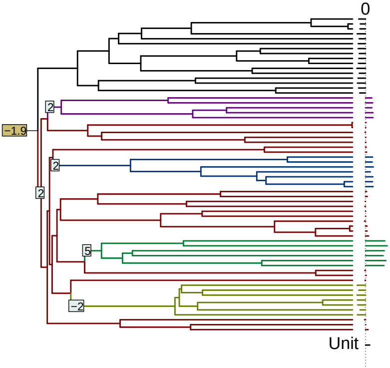
$$p_\theta(Z,Y) = p_\theta(Z_1) \prod_{1<j\leq m} p_\theta(Z_j|Z_{\text{parent}(j)}) \prod_{1\leq i\leq n} p_\theta(Y_i|Z_{\text{parent}(i)})$$

EM Recursive algorithm to find $\hat{\theta}_K = \text{argmax}_{\eta\in\mathcal{S}_K^{PI}}\, p_{\hat{\theta}_\eta}(Y)$:

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

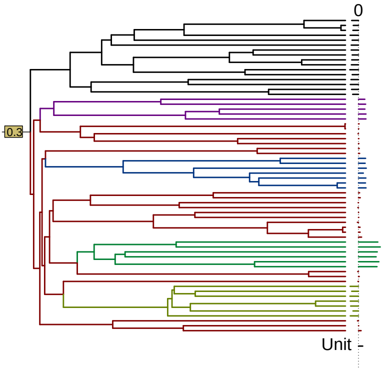
(G)M step θ^{h+1} raises $\mathbb{E}_{\theta^h}[\log p_\theta(Z,Y) \mid Y]$

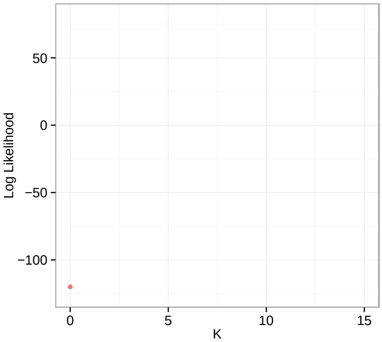
Model Selection on K



Simulated OU (α =3, γ^2 =0.1)

Model Selection on K

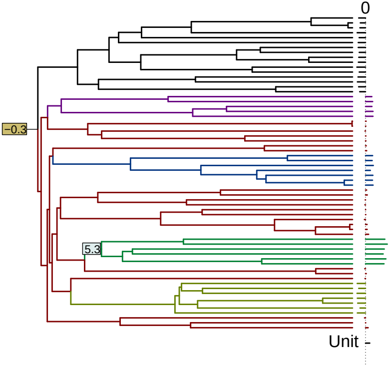


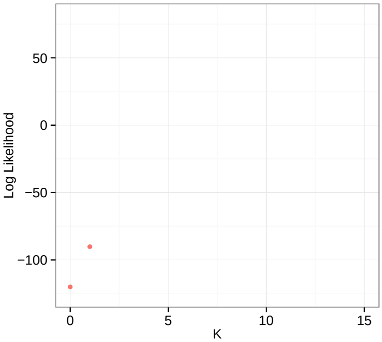


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}} \, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \ p_{\hat{\theta}_K}(Y)$$

Model Selection on K

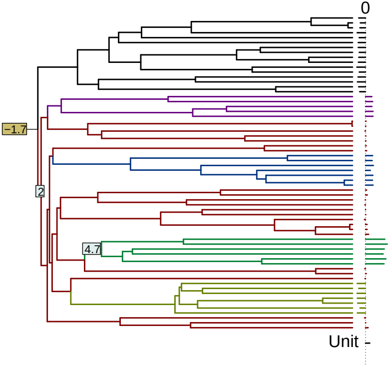


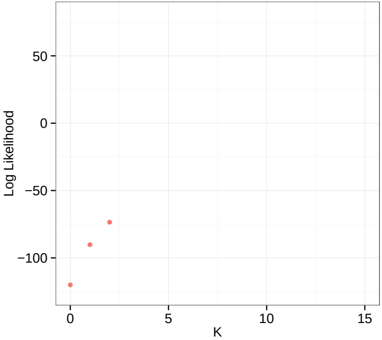


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \ p_{\hat{\theta}_K}(Y)$$

Model Selection on K

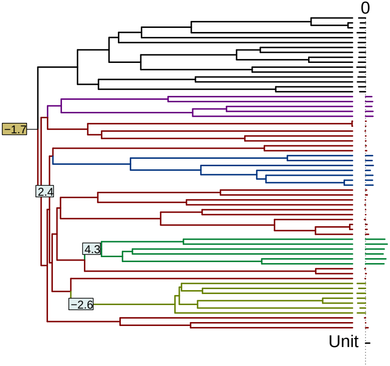


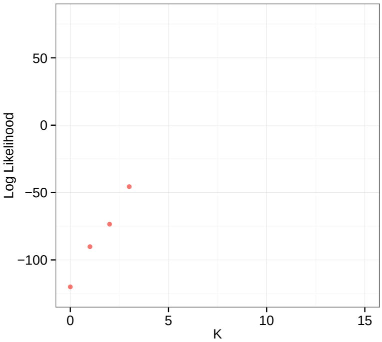


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}} \, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \; p_{\hat{\theta}_K}(Y)$$

Model Selection on K

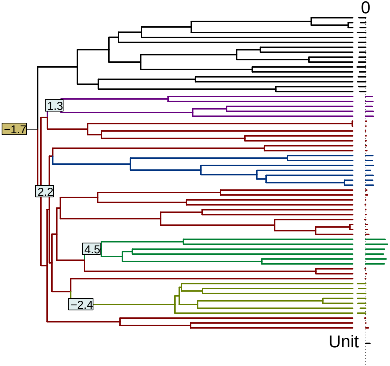


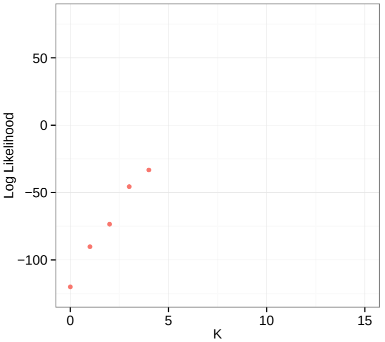


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \; p_{\hat{\theta}_K}(Y)$$

Model Selection on K

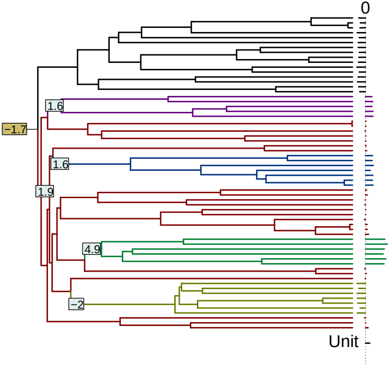


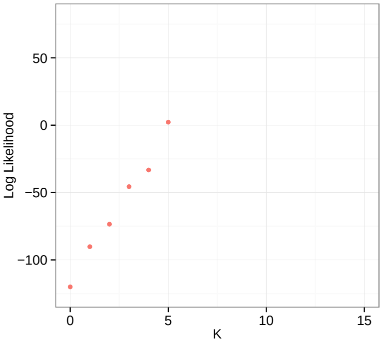


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \; p_{\hat{\theta}_K}(Y)$$

Model Selection on K

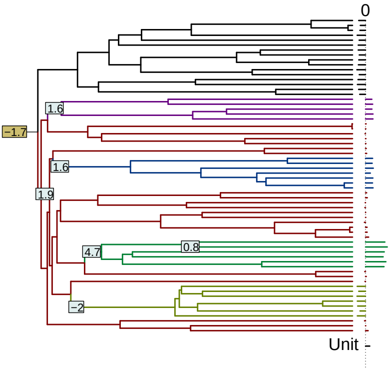


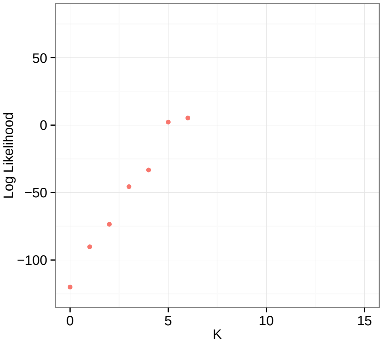


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \; p_{\hat{\theta}_K}(Y)$$

Model Selection on K

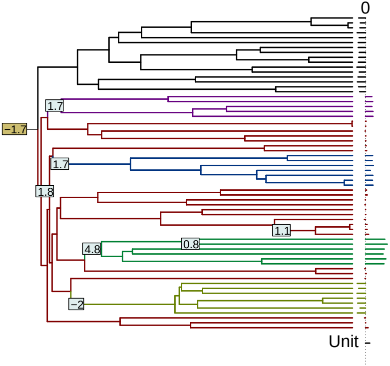


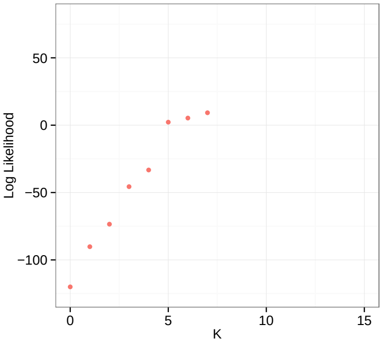


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log\ p_{\hat{\theta}_K}(Y)$$

Model Selection on K

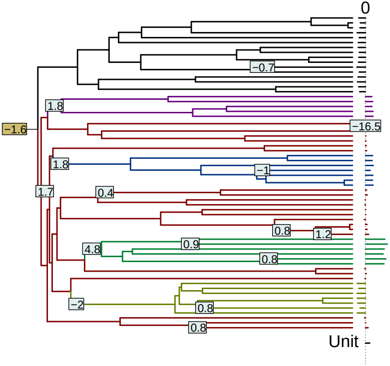


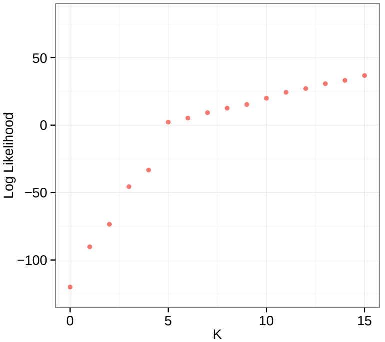


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \ p_{\hat{\theta}_K}(Y)$$

Model Selection on K



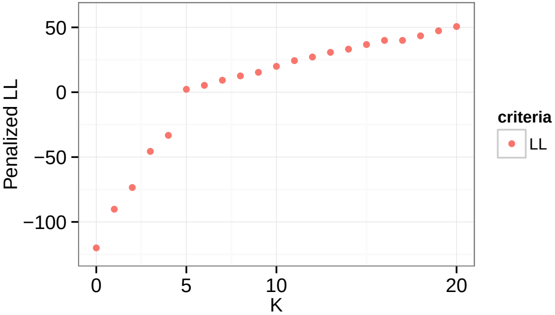


$$\hat{\theta}_K = \underset{\eta \in \mathcal{S}_K^{PI}}{\operatorname{argmax}}\, p_{\hat{\theta}_\eta}(Y)$$

$$LL = \log \ p_{\hat{\theta}_K}(Y)$$

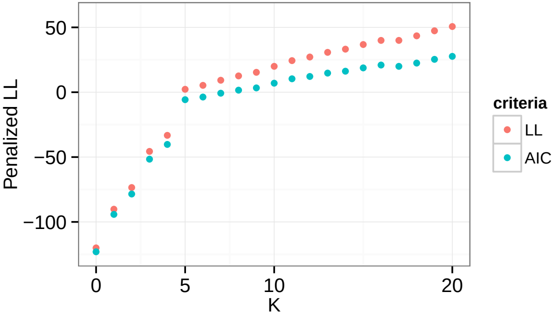
Model Selection: Penalized Likelihood

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



Model Selection: Penalized Likelihood

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

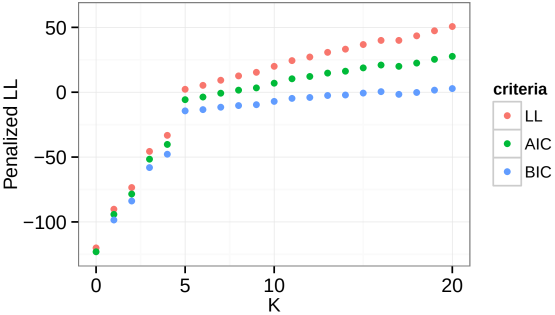


Penalties:

AIC $K + 3$

Model Selection: Penalized Likelihood

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



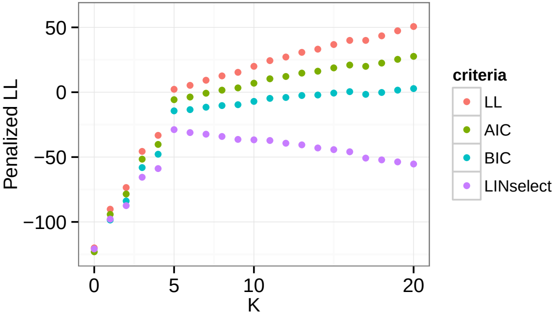
Penalties:

AIC $K + 3$

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

Model Selection: Penalized Likelihood

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



Penalties:

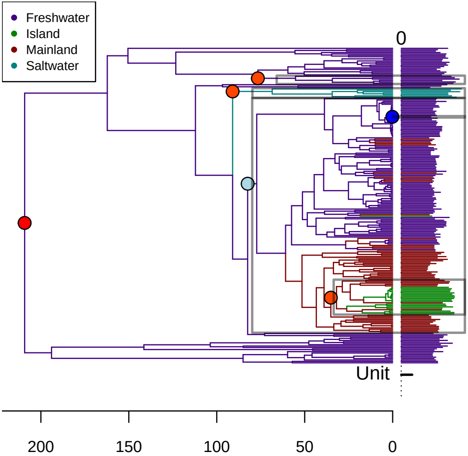
AIC $K + 3$

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

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

Based on Baraud et al. (2009)

Turtles Dataset

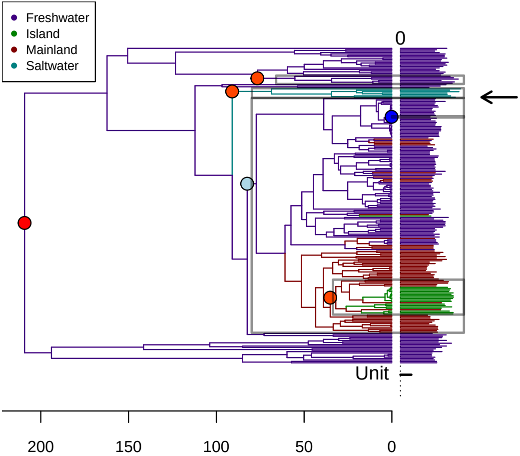


Colors: habitats.
Boxes: selected EM regimes.

	Habitat	EM
No. of shifts	16	5
No. of regimes	4	6
lnL	-133.86	-97.59
$\ln 2/\alpha$ (%)	7.44	5.43
$\sigma^2/2\alpha$	0.33	0.22
CPU t (min)	65.25	134.49

(Jaffe et al., 2011)

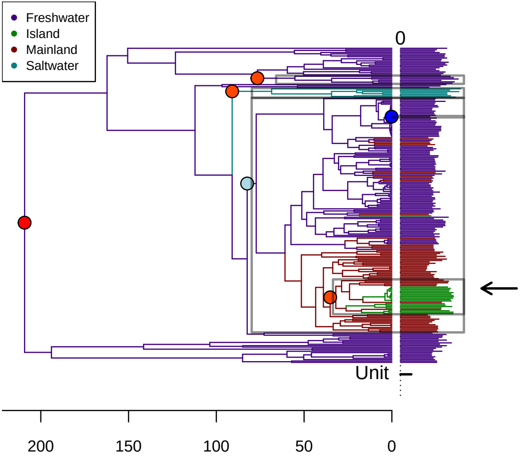
Turtles Dataset



Colors: habitats.
Boxes: selected EM regimes.

Chelonia mydas

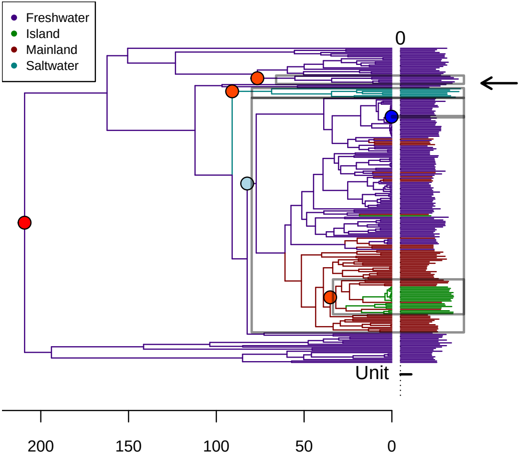
Turtles Dataset



Colors: habitats.
Boxes: selected EM regimes.

Geochelone nigra abingdoni

Turtles Dataset



Colors: habitats.
Boxes: selected EM regimes.

Chitra indica

Conclusion and Perspectives

A general inference framework for trait evolution models.

Conclusions
- Identifiability can be assessed.
- An EM can be written to maximize likelihood.
- Model selection for a non-iid framework.

R codes Available on GitHub:
`https://github.com/pbastide/PhylogeneticEM`

Perspectives
- Multivariate traits.
- Deal with uncertainty (tree, data).
- Phylogenetic networks.

Bibliography

Y. Baraud, C. Giraud, and S. Huet. Gaussian Model Selection with an Unknown Variance. *The Annals of Statistics*, 37(2):630–672, Apr. 2009.

J.-P. Baudry, C. Maugis, and B. Michel. Slope Heuristics: Overview and Implementation. *Statistics and Computing*, 22(2):455–470, March 2012.

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

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

J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Suderland, USA, 2004.

T. F. Hansen. Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution*, 51(5):1341–1351, October 1997.

A. L. Jaffe, G. J. Slater, and M. E. Alfaro. The Evolution of Island Gigantism and Body Size Variation in Tortoises and Turtles. *Biology letters*, 11(11), November 2011.

P. Massart. *Concentration Inequalities and Model Selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 2007.

J. C. Uyeda and L. J. Harmon. A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Systematic Biology*, 63(6):902–918, July 2014.

Thank you for listening

pbastide.github.io

Appendices

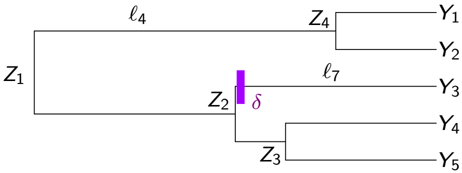
5 Inference
- EM
- Model Selection

6 Identifiability Issues
- Cardinal of Equivalence Classes
- Number of Tree Compatible Clustering

7 Simulations Results

8 Multivariate
- Models
- Inference

E step



Compute the following quantities:

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

- Using Gaussian properties. Need to invert matrices: complexity in $O(n^3)$.
- Using Gaussian properties **and** the tree structure: "Upward-Downward" algorithm. Complexity in $O(n)$.

M Step

Maximize:

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

- μ, γ^2, σ^2: simple maximization
- Discrete location of K shifts
  - $\mapsto$ Exact and fast for the BM
- α: numerical maximization and/or on a grid
  - $\mapsto$ Generalized EM

Initialization

Shifts : Lasso regression.

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \|Y - TW(\alpha)\Delta\|^2_{\Sigma_{YY}^{-1}} + \lambda \|\Delta_{-1}\|_1 \right\}$$

- Initialize $\Sigma_{YY}(\alpha)$, then estimate Δ with a Gauss Lasso procedure, using a Cholesky decomposition.

+

- λ chosen to get K shifts.

The selection strength α : Initialization using couples of tips.

back

Cholesky Decomposition

The problem is:

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \| Y - R\Delta \|^2_{\Sigma_{YY}} + \lambda \left| \Delta_{-1} \right|_1 \right\}$$

Cholesky decomposition of Σ_{YY}:

$$\Sigma_{YY} = LL^T \ , \ L \text{ a lower triangular matrix}$$

Then:

$$\| Y - R\Delta \|^2_{\Sigma_{YY}} = \left\| L^{-1} Y - L^{-1} R\Delta \right\|^2$$

And if $Y' = L^{-1}Y$ and $R' = L^{-1}R$, the problem becomes:

$$\hat{\Delta} = \underset{\Delta}{\operatorname{argmin}} \left\{ \left\| Y' - R'\Delta \right\|^2 + \lambda \left| \Delta_{-1} \right|_1 \right\}$$

Gauss Lasso

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

$$\hat{\Delta}^{\text{Gauss}} = \Pi_{\hat{F}_\lambda}(Y') \text{ with } \hat{F}_\lambda = \text{Span}\{R'_j : j \in \hat{m}_\lambda\}$$

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:

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

Upward

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

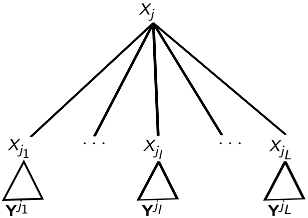
$$f_{\mathbf{Y}^j|X_j}(\mathbf{Y}^j; a) = A_j(\mathbf{Y}^j)\Phi_{M_j(\mathbf{Y}^j), S_j^2(\mathbf{Y}^j)}(a)$$

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

Propagation

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

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



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

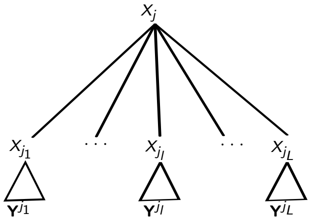
Compute $E_j = \mathbb{E}\left[X_j \mid \mathbf{Y}\right]$, $V_j^2 = \mathbb{V}\text{ar}\left[X_j \mid \mathbf{Y}\right]$, $C_{j,\text{pa}(j)}^2 = \mathbb{C}\text{ov}\left[X_j; X_{\text{pa}(j)} \mid \mathbf{Y}\right]$

Initialization Last step of Upward.

Propagation

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

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



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

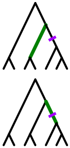
M Step: Segmentation

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

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

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

$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y \right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y \right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\} \delta_k \right)^2$$



Algorithm:

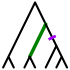
1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

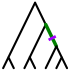
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y \right] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y \right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

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

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



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y \right] - q_j \mathbb{E}\left[X_{\text{pa}(j)} \mid Y \right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$



Algorithm:

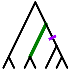
1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

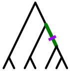
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

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

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



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$



Algorithm:

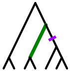
1. Find the K branches $j_1, \dots, j_K$ with largest C_j^0;
2. Allocate one change point in the first K branches;
3. For each of these branches, set $\delta_{j_k}^{(h+1)}$ so that $C_j^1(\tau, \delta) = 0$

M Step: Segmentation

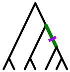
$$C_j(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - r_j - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$

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

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



$$C_j^1(\alpha, \tau, \delta) = \sigma_j^{-2} \left(\mathbb{E}\left[X_j \mid Y\right] - q_j \mathbb{E}\left[X_{\mathrm{pa}(j)} \mid Y\right] - s_j \sum_k \mathbb{I}\{\tau_k = b_j\}\delta_k \right)^2$$



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 $C_j^1(\tau, \delta) = 0$

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.

back

Model Selection on K: LINselect

Goal

$$\hat{K} = \underset{0 \leq K \leq p-1}{\operatorname{argmin}} \left\| Y - \hat{Y}_K \right\|_V^2 \left(1 + \frac{\text{pen}(K)}{n - K - 1} \right)$$

Oracle

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

Definition (Baraud et al. (2009))

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

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

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

Proposition: LINselect Penalty

Proposition (Form of the Penalty and guarantees (α known))

Under our setting: $Y = TW(\alpha)\Delta + \gamma E$ with $E \sim \mathcal{N}(0, V)$, define the penalty:

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

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

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

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

back ?

LINselect Model Selection: Important Points

Based on Baraud, Giraud, and Huet (2009)

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

Novelties

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

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:*

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

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

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

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

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

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

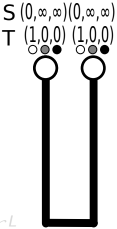
$$S'_\eta = L^{-1}S_\eta, \quad \hat{Y}'_\eta = \text{Proj}_{S'_\eta} Y' = \underset{a' \in S'_\eta}{\text{argmin}} \left\| Y - La' \right\|_V^2 = L^{-1}\hat{Y}_\eta$$

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

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

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

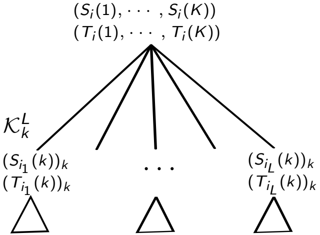
Initialization For tips

Propagation

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

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

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



Termination Sum on the root vector

back

Cardinal of Equivalence Classes

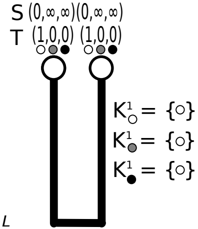
Initialization For tips

Propagation

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

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

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



S(○) = 0 + 0 ; T(○) = 1 x 1
S(●) = 1 + 1 ; T(●) = 1 x 1
S(●) = 1 + 1 ; T(●) = 1 x 1

Termination Sum on the root vector

back

Cardinal of Equivalence Classes

Initialization For tips

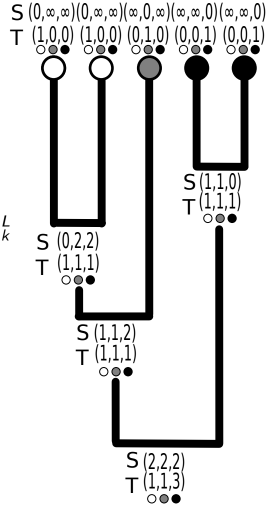
Propagation

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

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

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

Termination Sum on the root vector



back

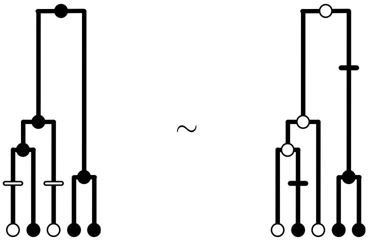
Linking Shifts and Clustering

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

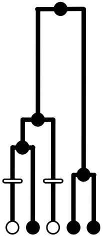


The No Homoplasy hypothesis is not respected.

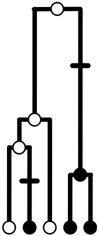
Proposition "K shifts $\iff$ $K+1$ clusters"

Linking Shifts and Clustering

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



$\sim$

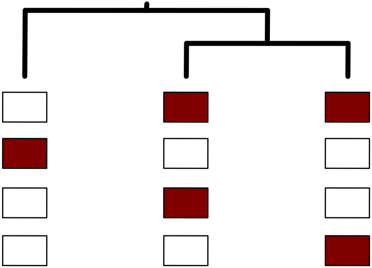


The No Homoplasy hypothesis is not respected.

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

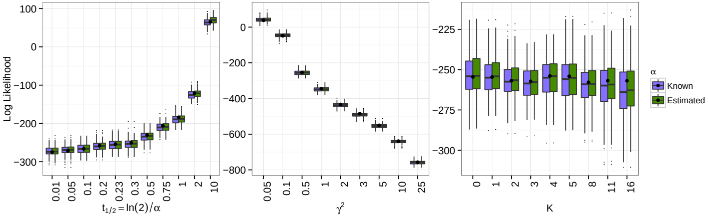
Simulations Design (Uyeda and Harmon, 2014)

- 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):

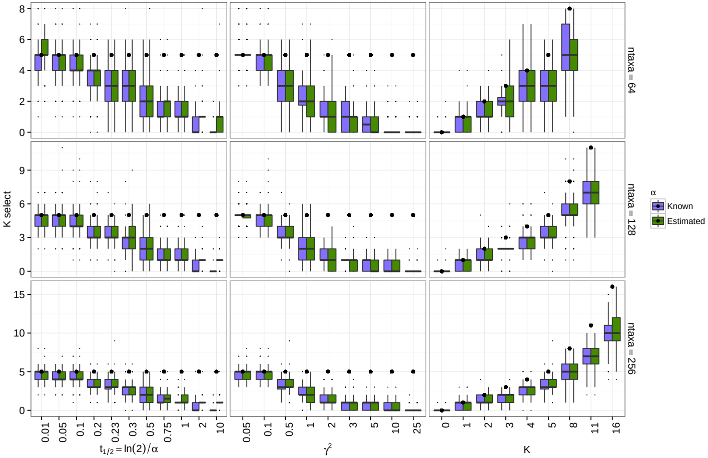
- α known: 6 minutes per estimation (66 days in total).
- α unknown: 52 minutes per estimation (570 days in total).

Log-Likelihood

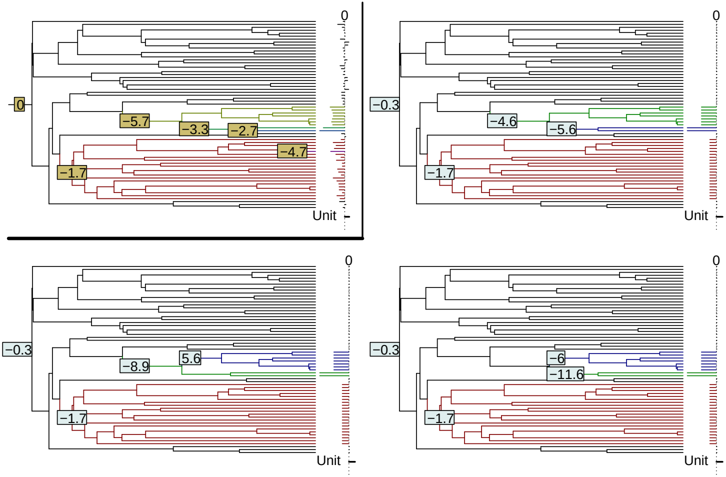


Log likelihood for a tree with 256 *tips. Solid black dots are the median of the log likelihood for the true parameters.*

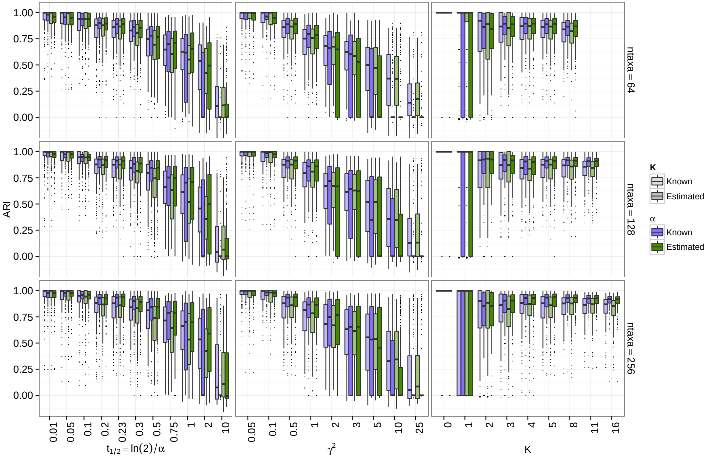
Number of Shifts



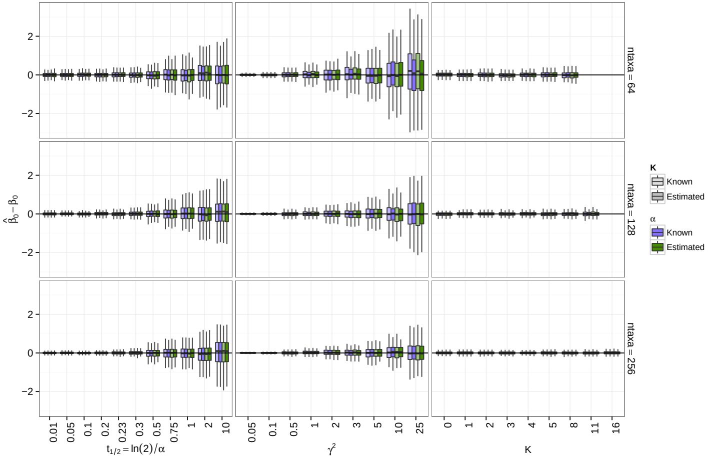
One Example


0
Unit
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

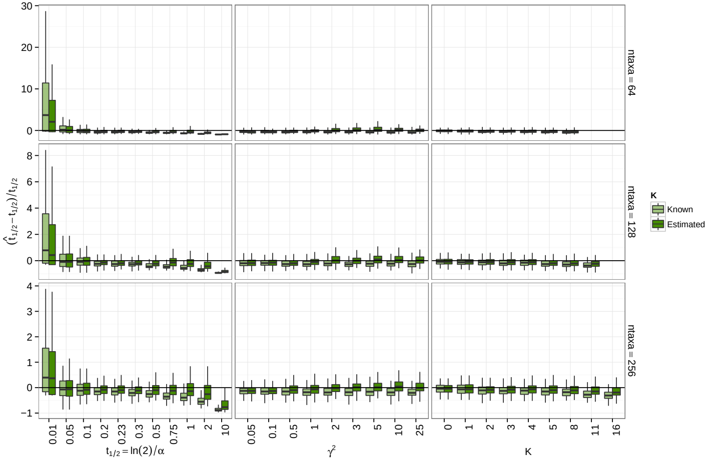
Adjusted Rand Index



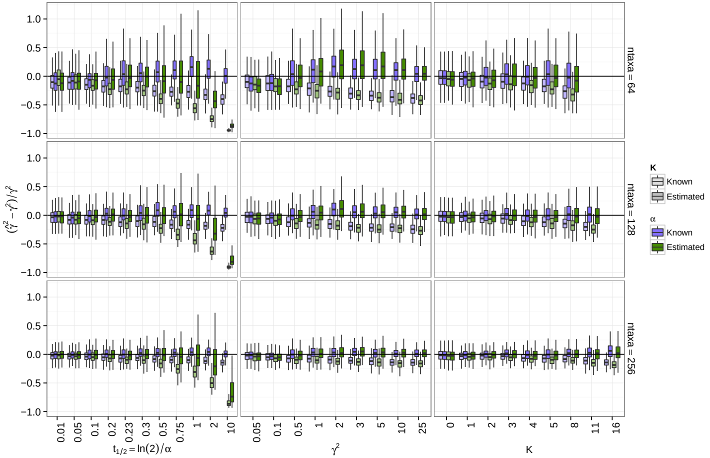
Parameters: β_0



Parameters: α



Parameters: γ^2



BM Model

Data n *vectors* of p traits at the tips: $\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{ip} \end{pmatrix}$

SDE $d\mathbf{W}(t) = \mathbf{\Sigma} d\mathbf{B}_t$, rate matrix $\mathbf{R} = \mathbf{\Sigma}\mathbf{\Sigma}^T$ $(p \times p)$

Covariances $\mathbb{C}\text{ov}\left[Y_{il}; Y_{jq}\right] = t_{ij} R_{lq}$ for i, j tips, and l, q characters

$$\mathbb{V}\text{ar}\left[\text{vec}(\mathbf{Y})\right] = \mathbf{C}_n \otimes \mathbf{R}$$

Shifts K shifts $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_K$ *vectors* size p

$\mapsto$ All the characters shift at the same time

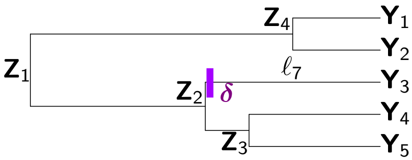
BM Model

Linear Model Representation

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

Incomplete Data Representation

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



OU Model: General Case

Data n *vectors* of p traits at the tips: $\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{ip} \end{pmatrix}$

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

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

Shifts K shifts $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_K$ *vectors* size p

$\mapsto$ On the optimal values

OU Model: **A** scalar

Assumption $\mathbf{A} = \alpha \mathbf{I}_p$ "scalar"

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

Fixed Root For i, j tips and l, q characters:

$$\mathbb{C}\mathrm{ov}\left[Y_{il}; Y_{jq}\right] = \frac{1}{2\alpha} e^{-2\alpha h} \left(e^{2\alpha t_{ij}} - 1\right) R_{lq}$$

$\mapsto$ Can be reduced to a BM on a re-scaled tree

EM algorithm

BM Natural generalization of the univariate case.

OU M step intractable in general.

Incomplete Data Model: Can readily handle missing data.

Model Selection

- Previous criterion cannot be applied
- Solution: "Slope Heuristic"-based method
  - 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: Toy Example

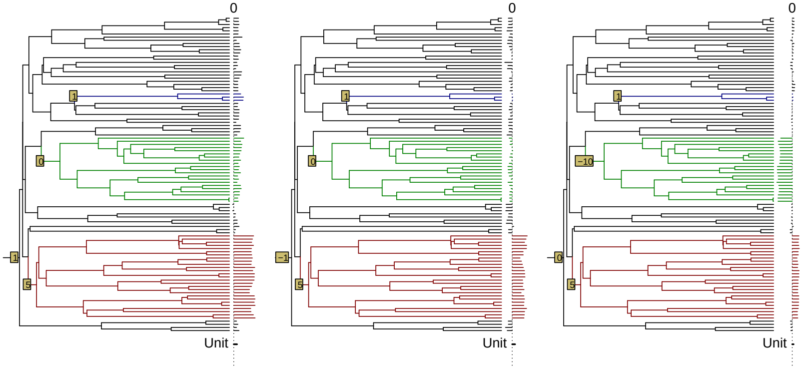


Figure: Simulated Process.

Model Selection: Toy Example

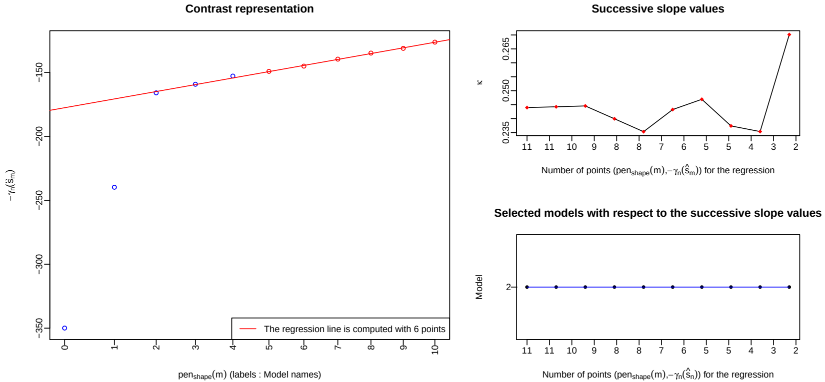


Figure: capushe *output for penalized log-likelihood.*

Model Selection: Toy Example

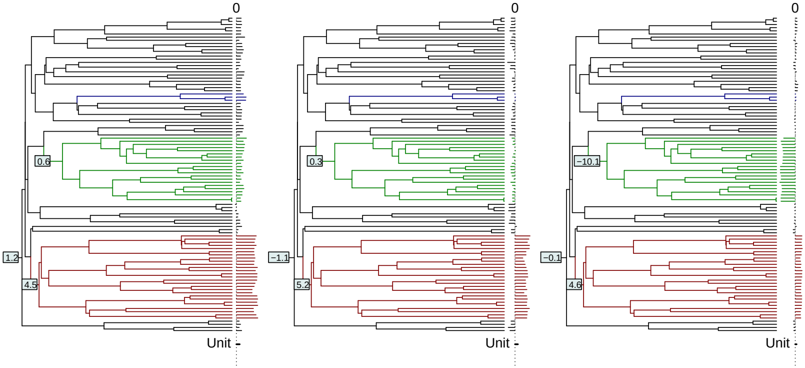


Figure: Reconstructed Process.