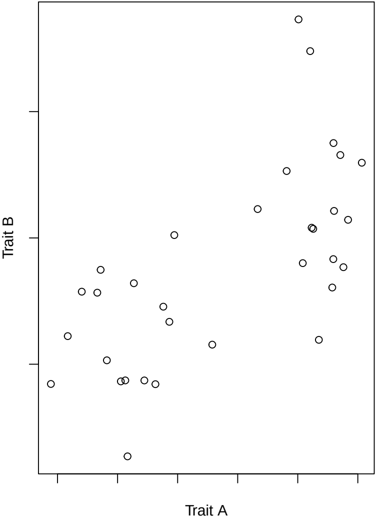
Continuous Trait Evolution

Paul Bastide[1], Julien Clavel[2]

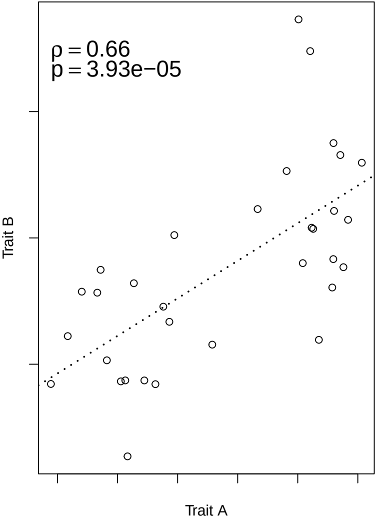
[1] IMAG, Université de Montpellier, CNRS, `paul.bastide@umontpellier.fr`
[2] LEHNA, Université Lyon 1, CNRS, `julien.clavel@univ-lyon1.fr`

December 2022

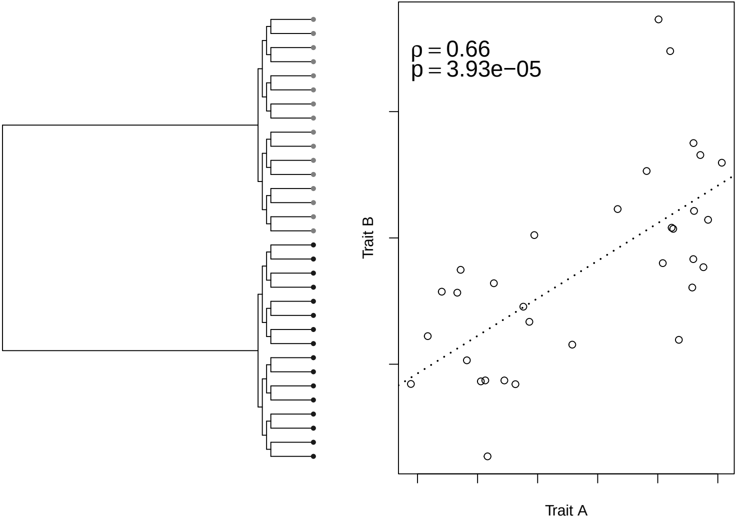
Phylogenetic Comparative Methods



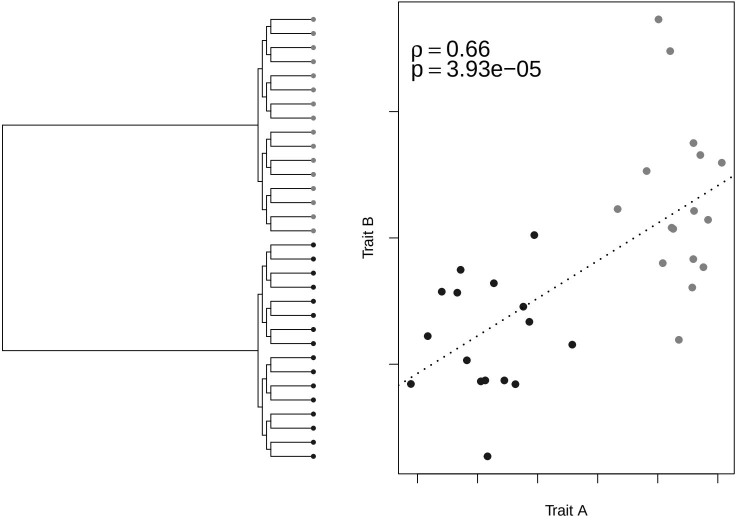
Phylogenetic Comparative Methods



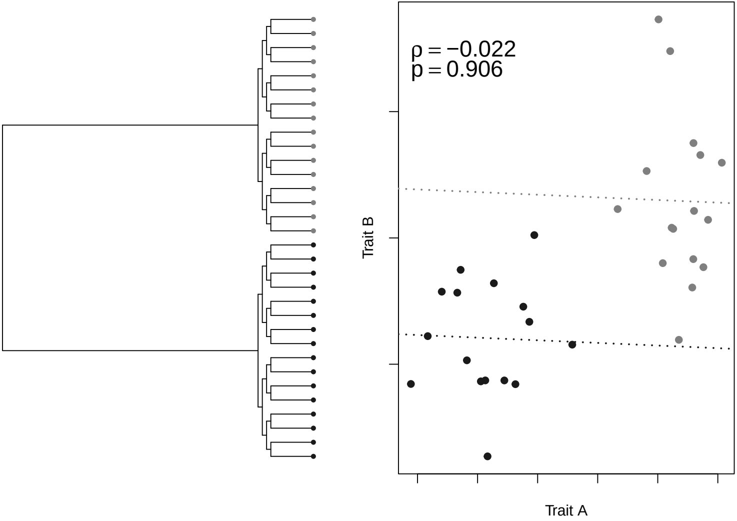
Phylogenetic Comparative Methods



Phylogenetic Comparative Methods



Phylogenetic Comparative Methods



Outline

1. Linear Regression Framework
2. Model Fit and Computations
3. Multidimensional Models

Useful Ressources

- Felsenstein (2004), Chapters 23 to 25.
- Harmon (2019), free online:
  `http://lukejharmon.github.io/pcm/`
- Bastide, Mariadassou, and Robin (2022), in French,
  `https://hal.archives-ouvertes.fr/hal-03762880`
- CRAN Task View: Phylogenetics
  `https://CRAN.R-project.org/view=Phylogenetics`

Outline

1. Linear Regression Framework
    - Phylogenetic Linear Regression
    - Maximum Likelihood Estimation
    - Phylogenetic ANOVA
2. Model Fit and Computations
3. Multidimensional Models

Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

- **Y** observed vector of traits at the tips of the tree (length *n*)

Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{Y}$ observed vector of traits at the tips of the tree (length n)
- $\mathbf{X}$ matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)

Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{Y}$ observed vector of traits at the tips of the tree (length n)
- $\mathbf{X}$ matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)
- $\boldsymbol{\epsilon}$ vector of errors (length n)

Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{Y}$ observed vector of traits at the tips of the tree (length n)
- $\mathbf{X}$ matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)
- $\boldsymbol{\epsilon}$ vector of errors (length n)

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$$

+

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

$$\hat{\beta} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

$$\hat{\beta} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \underset{\mathbf{u} \in \mathcal{M}_{\mathbf{X}}(\mathbb{R}^p)}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{u}\|^2 = \mathbf{P}_{\mathbf{X}}\mathbf{Y}$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$$

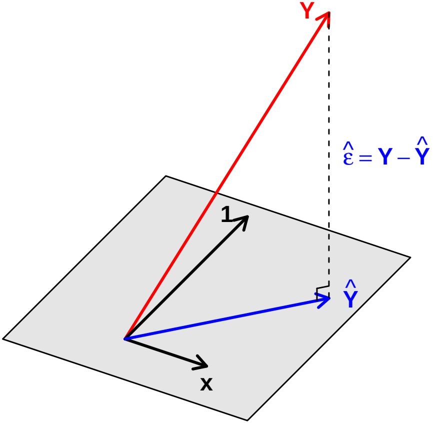
$$\hat{\beta} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \underset{\mathbf{u}\in\mathcal{M}_{\mathbf{X}}(\mathbb{R}^p)}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{u}\|^2 = \mathbf{P}_{\mathbf{X}}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-p}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Linear Regression: Estimation



Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
$$\hat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Distribution:

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{kk}}} \sim$$
$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\boldsymbol{\epsilon}$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
$$\hat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Distribution:

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{kk}}} \sim \mathcal{T}_{n-p}$$
$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim$$

Linear Regression: Estimation

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon$$

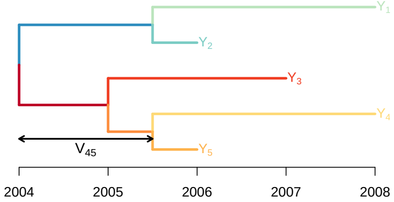
Estimators:

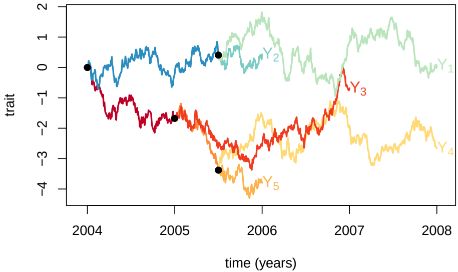
$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
$$\hat{\sigma}^2 = \frac{1}{n-p}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Distribution:

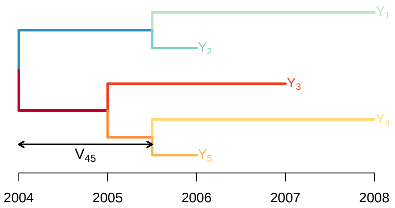
$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{kk}}} \sim \mathcal{T}_{n-p}$$
$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

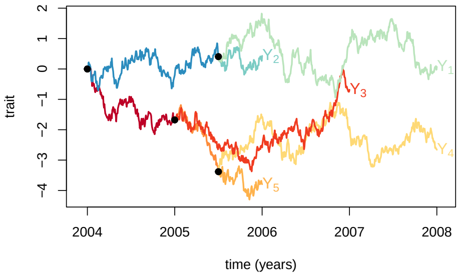
Brownian Motion on a Tree





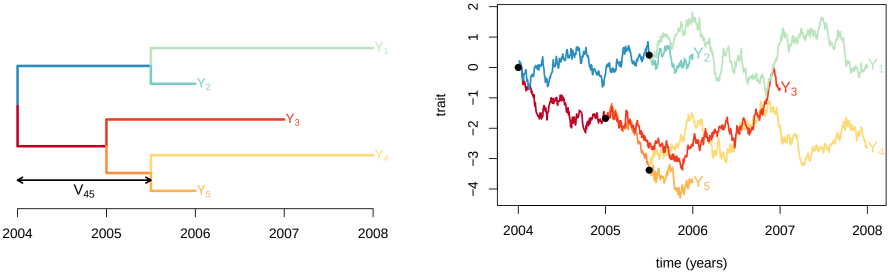
Brownian Motion on a Tree





- Variance: $\mathbb{C}\text{ov}(Y_i, Y_j) = \sigma^2 V_{ij}$
- Expectation: μ

Brownian Motion on a Tree



- Variance: $\mathbb{C}\mathrm{ov}(Y_i, Y_j) = \sigma^2 V_{ij}$
- Expectation: μ
- Distribution: $\mathbf{Y} = \mu\mathbf{1} + \sigma\mathbf{E}$, $\mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$.

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}$$

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}$$

- $\mathbf{Y}$ observed vector of traits at the tips of the tree (length n)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}$$

- **Y** observed vector of traits at the tips of the tree (length n)
- **X** matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}$$

- **Y** observed vector of traits at the tips of the tree (length n)
- **X** matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)
- **E** vector of phylogenetic errors (length n)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}$$

- **Y** observed vector of traits at the tips of the tree (length n)
- **X** matrix of regressors (size $n \times p$)
- β vector of coefficients (length p)
- **E** vector of phylogenetic errors (length n)

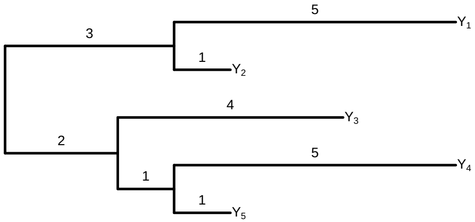
$$\mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

V informed by the tree and the trait model

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

BM on a tree:

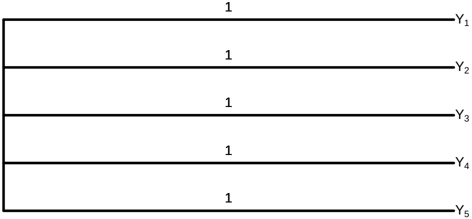


$$\mathbf{V} = \begin{array}{c c} & \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} & \begin{pmatrix} 8 & 3 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 6 & 2 & 2 \\ 0 & 0 & 2 & 8 & 3 \\ 0 & 0 & 2 & 3 & 4 \end{pmatrix} \end{array}$$

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

BM on a star tree:



$$\mathbf{V} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Generalized Least Squares

Assume that the tree and model (BM) is known.

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Generalized Least Squares

Assume that the tree and model (BM) is known.

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

$$\mathbb{V}\mathrm{ar}[\mathbf{L}^{-1}\mathbf{E}] =$$

Generalized Least Squares

Assume that the tree and model (BM) is known.

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

$$\mathbb{V}\text{ar}[\mathbf{L}^{-1}\mathbf{E}] = \mathbf{L}^{-1}\mathbf{V}[\mathbf{L}^{-1}]^T = \mathbf{I}$$

Generalized Least Squares

Assume that the tree and model (BM) is known.

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

$$\mathbb{V}\text{ar}[\mathbf{L}^{-1}\mathbf{E}] = \mathbf{L}^{-1}\mathbf{V}[\mathbf{L}^{-1}]^T = \mathbf{I}$$

De-correlated Regression:

$$\mathbf{L}^{-1}\mathbf{Y} =$$

Generalized Least Squares

Assume that the tree and model (BM) is known.

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

$$\mathbb{V}\mathrm{ar}[\mathbf{L}^{-1}\mathbf{E}] = \mathbf{L}^{-1}\mathbf{V}[\mathbf{L}^{-1}]^T = \mathbf{I}$$

De-correlated Regression:

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Estimators:

$\hat{\beta} =$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Estimators:

$$\begin{aligned}\hat{\beta} &= ((\mathbf{L}^{-1}\mathbf{X})^T(\mathbf{L}^{-1}\mathbf{X}))^{-1}(\mathbf{L}^{-1}\mathbf{X})^T\mathbf{L}^{-1}\mathbf{Y} \\ &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}\end{aligned}$$

$$\hat{\sigma}^2 =$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Estimators:

$$\begin{aligned}\hat{\beta} &= ((\mathbf{L}^{-1}\mathbf{X})^T(\mathbf{L}^{-1}\mathbf{X}))^{-1}(\mathbf{L}^{-1}\mathbf{X})^T\mathbf{L}^{-1}\mathbf{Y} \\ &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \\ \hat{\sigma}^2 &= \frac{1}{n-p}(\mathbf{L}^{-1}\mathbf{Y} - (\mathbf{L}^{-1}\mathbf{X})\hat{\beta})^T(\mathbf{L}^{-1}\mathbf{Y} - (\mathbf{L}^{-1}\mathbf{X})\hat{\beta})\end{aligned}$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

$$\mathbf{L}^{-1}\mathbf{Y} = (\mathbf{L}^{-1}\mathbf{X})\beta + \sigma\mathbf{E}', \quad \mathbf{E}' \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Estimators:

$$\begin{aligned}
\hat{\beta} &= ((\mathbf{L}^{-1}\mathbf{X})^T(\mathbf{L}^{-1}\mathbf{X}))^{-1}(\mathbf{L}^{-1}\mathbf{X})^T\mathbf{L}^{-1}\mathbf{Y} \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \\
\hat{\sigma}^2 &= \frac{1}{n-p}(\mathbf{L}^{-1}\mathbf{Y} - (\mathbf{L}^{-1}\mathbf{X})\hat{\beta})^T(\mathbf{L}^{-1}\mathbf{Y} - (\mathbf{L}^{-1}\mathbf{X})\hat{\beta}) \\
&= \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \frac{1}{n-p}\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2_{\mathbf{V}^{-1}}
\end{aligned}$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \frac{1}{n-p}\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2_{\mathbf{V}^{-1}}$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \frac{1}{n-p}\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2_{\mathbf{V}^{-1}}$$

Projection:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \underset{\mathbf{u}\in\mathcal{M}_{\mathbf{X}}(\mathbb{R}^p)}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{u}\|^2_{\mathbf{V}^{-1}} = \mathbf{P}^{\mathbf{V}^{-1}}_{\mathbf{X}}\mathbf{Y}$$

Orthogonal projection w.r.t. inner product $\langle \mathbf{u}, \mathbf{v}\rangle_{\mathbf{V}^{-1}} = \mathbf{u}^T\mathbf{V}^{-1}\mathbf{v}$.

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \frac{1}{n-p}\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2_{\mathbf{V}^{-1}}$$

Projection:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \underset{\mathbf{u}\in\mathcal{M}_{\mathbf{X}}(\mathbb{R}^p)}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{u}\|^2_{\mathbf{V}^{-1}} = \mathbf{P}_{\mathbf{X}}^{\mathbf{V}^{-1}}\mathbf{Y}$$

Orthogonal projection w.r.t. inner product $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{V}^{-1}} = \mathbf{u}^T\mathbf{V}^{-1}\mathbf{v}$.

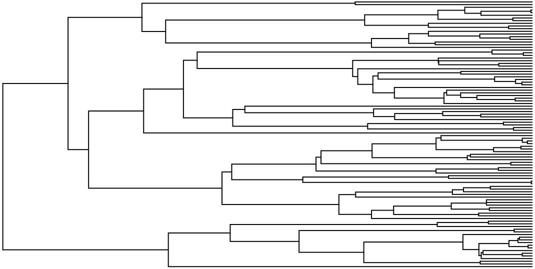
Distribution:

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}]_{kk}}} \sim \mathcal{T}_{n-p} \qquad \frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

Simulated Example - Simple Phylogenetic Regression

```
library(ape); library(phytools); library(phylolm)
```

```
set.seed(12891026)
n <- 100
tree <- rphylo(n, birth = 0.1, death = 0)
plot(tree, show.tip.label = FALSE, no.margin = TRUE)
```

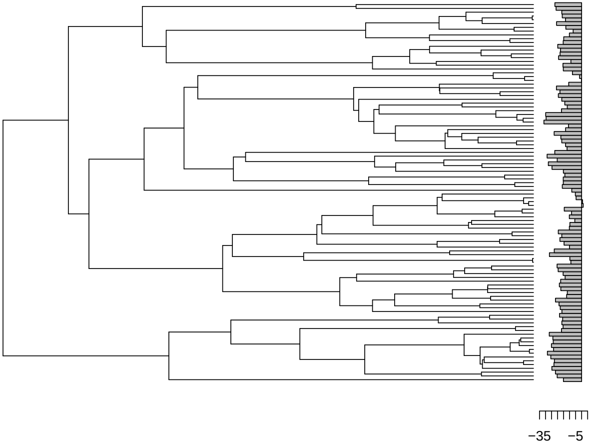


Simulated Example - Simple Phylogenetic Regression

```r
# Regressor
x <- rnorm(n, mean = 0, sd = 4)

# Phylogenetic Noise
eps <- rTrait(n = 1, phy = tree, model = "BM",
              parameters = list(ancestral.state = 0, sigma2 = 2))

# Response variable
beta0 <- -10
beta1 <- 0.5
y <- beta0 + beta1 * x + eps
```

Simulated Example - Simple Phylogenetic Regression

```r
par(mar = c(5, 0, 0, 0) + 0.1)
plot(tree, show.tip.label = FALSE, x.lim = 50)
phydataplot(y, tree, scaling = 0.1, offset = 4)
```



Simulated Example - Simple Phylogenetic Regression

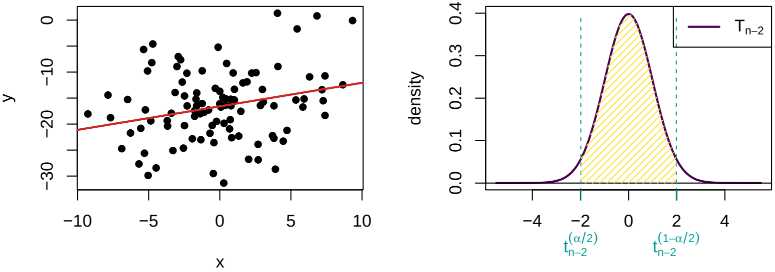
```
# Standard linear regression
fit <- lm(y ~ x)
summary(fit)


##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.0118  -3.8956  -0.1975   3.5091  16.1712
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.4469     0.6578 -25.001   <2e-16 ***
## x             0.3951     0.1661   2.379   0.0193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.577 on 98 degrees of freedom
## Multiple R-squared:  0.0546,Adjusted R-squared:  0.04495
## F-statistic:  5.66 on 1 and 98 DF,  p-value: 0.0193
```

Simulated Example - Simple Phylogenetic Regression

```
# Phylogenetic linear regression
fitphy <- phylolm(y ~ x, phy = tree)
summary(fitphy)


##
## Call:
## phylolm(formula = y ~ x, phy = tree)
##
##    AIC logLik
##  559.0 -276.5
##
## Raw residuals:
##      Min       1Q   Median       3Q      Max
## -14.8836  -3.7381  -0.0423   3.6436  16.0813
##
## Mean tip height: 44.10044
## Parameter estimate(s) using ML:
## sigma2: 1.631192
##
## Coefficients:
##               Estimate     StdErr t.value   p.value
## (Intercept) -16.591222   3.487103 -4.7579 6.743e-06 ***
## x             0.452841   0.047507  9.5320 1.256e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.4811 Adjusted R-squared: 0.4758
```

Simulated Example - Simple Phylogenetic Regression

$$\mathbf{y} = -10 \cdot \mathbf{1} + 0.5 \cdot \mathbf{x} + \boldsymbol{\epsilon}$$



With probability $1 - \alpha$:

$$t_{n-2}(\alpha/2) \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}]_{11}}} \leq t_{n-2}(1 - \alpha/2)$$

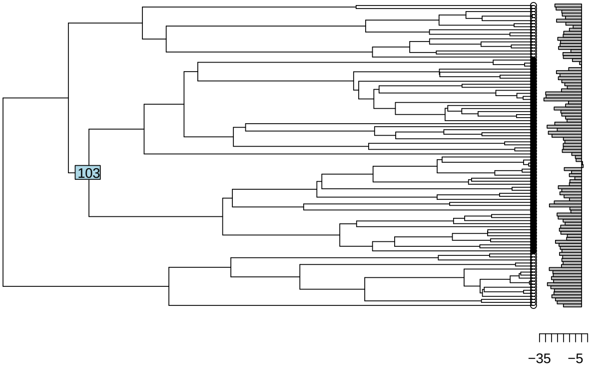
Simulated Example - Simple Phylogenetic Regression

```
# Simple linear regression
confint(fit)

##                    2.5 %      97.5 %
## (Intercept) -17.75232521 -15.1413842
## x             0.06551631   0.7246207
```

```
# Phylogenetic linear regression
confint(fitphy)

##                   2.5 %    97.5 %
## (Intercept) -23.4258188 -9.756624
## x             0.3597283  0.545954
```

Simulated Example - With Meaningless Group

```
# Clade
group <- sapply(1:n,
                function(tip) tip %in% getDescendants(tree, 103))
names(group) <- tree$tip.label
```



Simulated Example - With Meaningless Group

```
# Standard linear regression
fit <- lm(y ~ x + group)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x + group)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4544  -3.7063  -0.0476   3.3292  14.4768
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -18.9628     1.0605 -17.881  < 2e-16 ***
## x             0.4617     0.1614   2.860  0.00519 **
## groupTRUE     3.9399     1.3321   2.958  0.00389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.331 on 97 degrees of freedom
## Multiple R-squared:  0.1328,Adjusted R-squared:  0.1149
## F-statistic: 7.428 on 2 and 97 DF,  p-value: 0.0009968
```

Simulated Example - With Meaningless Group

```
# Phylogenetic linear regression
fitphy <- phylolm(y ~ x + group, phy = tree)
summary(fitphy)


##
## Call:
## phylolm(formula = y ~ x + group, phy = tree)
##
##    AIC logLik
##  560.5 -276.3
##
## Raw residuals:
##      Min       1Q   Median       3Q      Max
## -16.9503  -4.2706  -0.7069   2.5404  14.0122
##
## Mean tip height: 44.10044
## Parameter estimate(s) using ML:
## sigma2: 1.6238
##
## Coefficients:
##               Estimate     StdErr t.value   p.value
## (Intercept) -18.167276   4.225440 -4.2995 4.078e-05 ***
## x             0.453466   0.047653  9.5161 1.488e-15 ***
## groupTRUE     3.642587   5.481397  0.6645    0.5079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.4834 Adjusted R-squared: 0.4728
```

Reminder: One way ANOVA

K groups, each with n_k data points:

$$y_{ki} = \mu_k + \sigma \epsilon_{ik}, \quad \epsilon_{ik} \sim \mathcal{N}(0, 1), \quad 1 \leq k \leq K, \quad 1 \leq i \leq n_k$$

Reminder: One way ANOVA

K groups, each with n_k data points:

$$y_{ki} = \mu_k + \sigma\epsilon_{ik}, \quad \epsilon_{ik} \sim \mathcal{N}(0,1), \quad 1 \leq k \leq K, \quad 1 \leq i \leq n_k$$

One-way ANOVA:

$$\mathcal{H}_0 : \mu_k = \mu_l, \ \forall 1 \leq k, l \leq K \quad vs \quad \mathcal{H}_1 : \exists k, l \mid \mu_k \neq \mu_l$$

Reminder: One way ANOVA

K groups, each with n_k data points:

$$y_{ki} = \mu_k + \sigma\epsilon_{ik}, \quad \epsilon_{ik} \sim \mathcal{N}(0,1), \quad 1 \leq k \leq K, \quad 1 \leq i \leq n_k$$

One-way ANOVA:

$$\mathcal{H}_0 : \mu_k = \mu_l, \ \forall 1 \leq k, l \leq K \quad vs \quad \mathcal{H}_1 : \exists k, l \mid \mu_k \neq \mu_l$$

Group 1 as reference:

$$y_{ki} = \mu_1 + \beta_k + \sigma\epsilon_{ik}, \quad \text{with } \beta_1 = 0$$

Reminder: One way ANOVA

K groups, each with n_k data points:

$$y_{ki} = \mu_k + \sigma\epsilon_{ik}, \quad \epsilon_{ik} \sim \mathcal{N}(0,1), \quad 1 \leq k \leq K, \quad 1 \leq i \leq n_k$$

One-way ANOVA:

$$\mathcal{H}_0 : \mu_k = \mu_l, \ \forall 1 \leq k, l \leq K \quad vs \quad \mathcal{H}_1 : \exists k, l \mid \mu_k \neq \mu_l$$

Group 1 as reference:

$$y_{ki} = \mu_1 + \beta_k + \sigma\epsilon_{ik}, \quad \text{with } \beta_1 = 0$$

One-way ANOVA:

$$\mathcal{H}_0 : \beta_k = 0, \ \forall 2 \leq k \leq K \quad vs \quad \mathcal{H}_1 : \exists k \mid \beta_k \neq 0$$

Reminder: One way ANOVA

Group 1 as reference:

$$y_{ki} = \mu_1 + \beta_k + \sigma\epsilon_{ik}, \quad \text{with } \beta_1 = 0$$

Reminder: One way ANOVA

Group 1 as reference:

$$y_{ki} = \mu_1 + \beta_k + \sigma\epsilon_{ik}, \quad \text{with } \beta_1 = 0$$

Linear Model:

$$\mathbf{y} = \mathbf{X}\beta + \sigma\mathbf{E} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{1}_2, \dots, \mathbf{1}_K) \qquad \boldsymbol{\beta} = (\mu_1, \beta_2, \dots, \beta_K)^T$$

Reminder: One way ANOVA

Group 1 as reference:

$$y_{ki} = \mu_1 + \beta_k + \sigma\epsilon_{ik}, \quad \text{with } \beta_1 = 0$$

Linear Model:

$$\mathbf{y} = \mathbf{X}\beta + \sigma\mathbf{E} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{1}_2, \dots, \mathbf{1}_K) \qquad \boldsymbol{\beta} = (\mu_1, \beta_2, \dots, \beta_K)^T$$

One-way ANOVA is a Fisher global F test

$$\mathcal{H}_0 : \beta_k = 0,\ \forall 2 \leq k \leq K \quad vs \quad \mathcal{H}_1 : \exists k \mid \beta_k \neq 0$$

$$F = \frac{\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2/(K-1)}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2/(n-K)} \underset{\mathcal{H}_0}{\sim} \mathcal{F}^{K-1}_{n-K}.$$

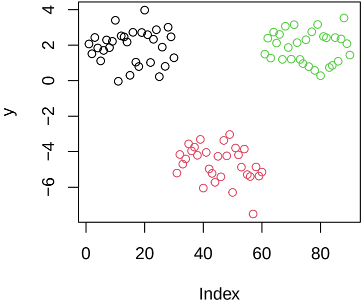
Simulated Example

```r
set.seed(18300718)
n <- 90
# Noise
eps <- rnorm(n, mean = 0, sd = 1)
# Groups
group <- as.factor(rep(c(1, 2, 3), each = n/3))
# Response variable
mu1 <- 2; mu2 <- -5; mu3 <- 2
y <- mu1 * (group == 1) + mu2 * (group == 2) + mu3 * (group == 3)
y <- y + eps
```

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \end{pmatrix}$$

Simulated Example

```r
plot(y, col = group)
```



Simulated Example

```r
fitanova <- lm(y ~ group)
summary(fitanova)
```

```
##
## Call:
## lm(formula = y ~ group)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8389 -0.6950  0.2115  0.6247  2.0544
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.92023    0.17258  11.127   <2e-16 ***
## group2      -6.59378    0.24406 -27.017   <2e-16 ***
## group3      -0.02984    0.24406  -0.122    0.903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9452 on 87 degrees of freedom
## Multiple R-squared:  0.9176,Adjusted R-squared:  0.9157
## F-statistic: 484.4 on 2 and 87 DF,  p-value: < 2.2e-16
```

Phylogenetic One way ANOVA

Linear Model:

$$\mathbf{y} = \mathbf{X}\beta + \sigma\mathbf{E} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{1}_2, \dots, \mathbf{1}_K) \qquad \boldsymbol{\beta} = (\mu_1, \beta_2, \dots, \beta_K)^T$$

Phylogenetic One way ANOVA

Linear Model:

$$\mathbf{y} = \mathbf{X}\beta + \sigma\mathbf{E} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{1}_2, \dots, \mathbf{1}_K) \qquad \boldsymbol{\beta} = (\mu_1, \beta_2, \dots, \beta_K)^T$$

Phylogenetic Errors (known tree and model)

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

Phylogenetic One way ANOVA

Linear Model:

$$\mathbf{y} = \mathbf{X}\beta + \sigma\mathbf{E} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{1}_2, \dots, \mathbf{1}_K) \qquad \boldsymbol{\beta} = (\mu_1, \beta_2, \dots, \beta_K)^T$$

Phylogenetic Errors (known tree and model)

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

One-way ANOVA is a Fisher global F test

$$\mathcal{H}_0 : \beta_k = 0,\ \forall 2 \leq k \leq K \quad vs \quad \mathcal{H}_1 : \exists k \mid \beta_k \neq 0$$

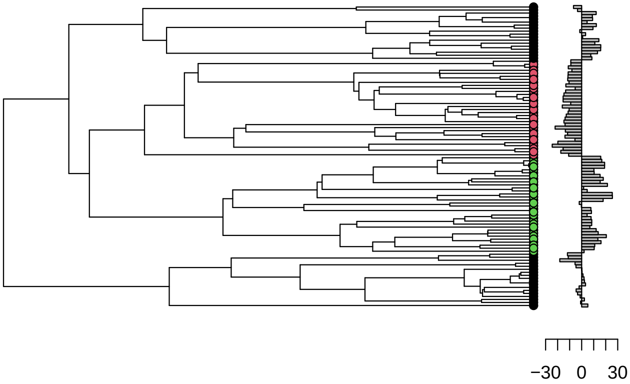
$$F = \frac{\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2_{\mathbf{V}^{-1}}/(K-1)}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2_{\mathbf{V}^{-1}}/(n-K)} \underset{\mathcal{H}_0}{\sim} \mathcal{F}^{K-1}_{n-K}.$$

Simulated Example

```r
set.seed(12891026)
n <- 100
# Tree
tree <- rphylo(n, 0.1, 0)
# Noise
eps <- rTrait(1, tree, "BM",
              parameters = list(ancestral.state = 0, sigma2 = 2))
# Groups
get_group <- function(tip) {
  if (tip %in% getDescendants(tree, 105)) return(2)
  if (tip %in% getDescendants(tree, 110)) return(3)
  return(1)
}
group <- as.factor(sapply(1:n, get_group))
# Response variable
mu1 <- 2; mu2 <- -5; mu3 <- 2
y <- mu1 * (group == 1) + mu2 * (group == 2) + mu3 * (group == 3)
y <- y + eps
```

Simulated Example

```r
par(mar = c(5, 0, 0, 0) + 0.1)
plot(tree, show.tip.label = FALSE, x.lim = 50)
tiplabels(bg = group, pch = 21)
phydataplot(y, tree, scaling = 0.1, offset = 4)
```



Simulated Example

```
fitanova <- lm(y ~ group)
summary(fitanova)


##
## Call:
## lm(formula = y ~ group)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.6482  -3.7608  -0.1398   4.0340  13.5348
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.440      1.121   2.177   0.0319 *
## group2       -15.451      1.634  -9.456 2.01e-15 ***
## group3         9.608      1.634   5.880 5.83e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.726 on 97 degrees of freedom
## Multiple R-squared:  0.7001,Adjusted R-squared:  0.6939
## F-statistic: 113.2 on 2 and 97 DF,  p-value: < 2.2e-16
```

Simulated Example

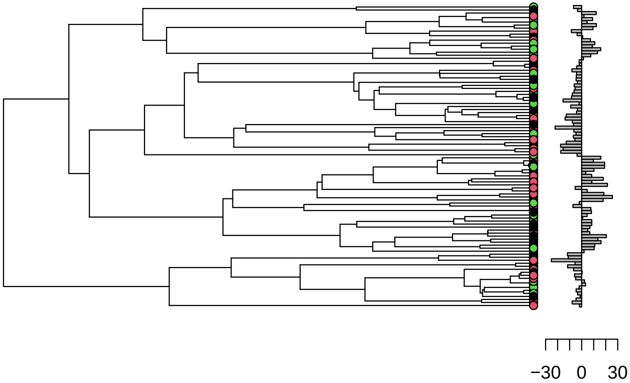
```
fitphyanova <- phylolm(y ~ group, phy = tree)
summary(fitphyanova)


##
## Call:
## phylolm(formula = y ~ group, phy = tree)
##
##    AIC logLik
##  554.9 -273.5
##
## Raw residuals:
##     Min      1Q  Median      3Q     Max
## -18.743  -1.787   1.845   6.918  17.064
##
## Mean tip height: 44.10044
## Parameter estimate(s) using ML:
## sigma2: 1.53545
##
## Coefficients:
##              Estimate   StdErr t.value p.value
## (Intercept)   0.53426  4.10823  0.1300 0.89680
## group2      -14.17576  6.00073 -2.3623 0.02016 *
## group3        7.98443  6.54651  1.2196 0.22556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.112 Adjusted R-squared: 0.09374
```

Simulated Example

```r
# Non-Phylogenetic groups
group <- as.factor(sample(c(1, 2, 3), n, replace = TRUE))
y <- mu1 * (group == 1) + mu2 * (group == 2) + mu3 * (group == 3)
y <- y + eps
```

Simulated Example

```
par(mar = c(5, 0, 0, 0) + 0.1)
plot(tree, show.tip.label = FALSE, x.lim = 50)
tiplabels(bg = group, pch = 21)
phydataplot(y, tree, scaling = 0.1, offset = 4)
```



Simulated Example

```r
fitanova <- lm(y ~ group)
summary(fitanova)
```

```
##
## Call:
## lm(formula = y ~ group)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.291  -7.377  -1.558   7.438  22.496
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0864     1.7317   1.782  0.07783 .
## group2       -7.0041     2.4490  -2.860  0.00519 **
## group3       -0.8096     2.4309  -0.333  0.73982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.948 on 97 degrees of freedom
## Multiple R-squared:  0.09196,Adjusted R-squared:  0.07324
## F-statistic: 4.912 on 2 and 97 DF,  p-value: 0.009292
```

Simulated Example

```r
fitphyanova <- phylolm(y ~ group, phy = tree)
summary(fitphyanova)
```

```
##
## Call:
## phylolm(formula = y ~ group, phy = tree)
##
##    AIC logLik
##  559.6 -275.8
##
## Raw residuals:
##     Min      1Q  Median      3Q     Max
## -18.263  -4.787   1.469  10.331  25.418
##
## Mean tip height: 44.10044
## Parameter estimate(s) using ML:
## sigma2: 1.609363
##
## Coefficients:
##             Estimate   StdErr  t.value p.value
## (Intercept)  0.44856  3.49619   0.1283  0.8982
## group2      -7.39368  0.51250 -14.4267  <2e-16 ***
## group3      -0.33789  0.48929  -0.6906  0.4915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.7699 Adjusted R-squared: 0.7652
```

Outline

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$-2L(\mathbf{Y}|\boldsymbol{\theta}) = n\log(2\pi) + \log\det\mathbf{V} + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \\
\hat{\sigma}^2 &= \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
-2L(\mathbf{Y}|\boldsymbol{\theta}) &= n\log(2\pi) + \log\det\mathbf{V} + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})
\end{aligned}$$

Naive Computation: Need to invert $\mathbf{V}$ → worse than $O(n^2)$.

Generalized Least Squares

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$-2L(\mathbf{Y}|\boldsymbol{\theta}) = n\log(2\pi) + \log\det\mathbf{V} + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Naive Computation: Need to invert $\mathbf{V}$ → worse than $O(n^2)$.

Three point structure:
Estimators computed in one traversal of the tree → $O(n)$.

+

Three Point Structure Algorithm (Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$2L(\mathbf{y}|\boldsymbol{\theta}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Can be computed in linear time.

Three Point Structure Algorithm (Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$2L(\mathbf{y}|\boldsymbol{\theta}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Can be computed in linear time.

- Works for a BM on a tree

Three Point Structure Algorithm (Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \\
\hat{\sigma}^2 &= \frac{1}{n-p}(\mathbf{Y}-\mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\beta}) \\
2L(\mathbf{y}|\boldsymbol{\theta}) &= n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y}-\mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\beta})
\end{aligned}$$

Can be computed in linear time.

- Works for a BM on a tree
- Implemented in phylolm

Three Point Structure Algorithm (Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$2L(\mathbf{y}|\boldsymbol{\theta}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

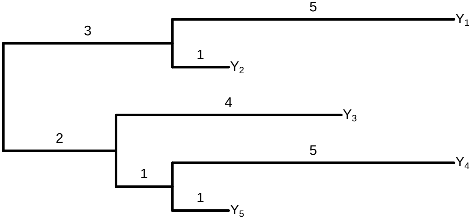
Can be computed in linear time.

- Works for a BM on a tree
- Implemented in phylolm
- What about other models ? $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$.

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

BM on a tree:

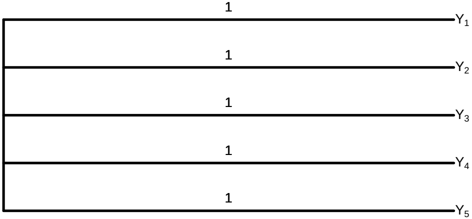


$$\mathbf{V} = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{pmatrix} 8 & 3 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 6 & 2 & 2 \\ 0 & 0 & 2 & 8 & 3 \\ 0 & 0 & 2 & 3 & 4 \end{pmatrix} \end{array}$$

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

BM on a star tree:

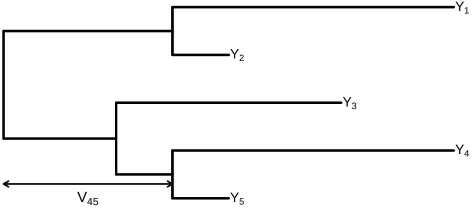


$$\mathbf{V} = \begin{array}{c c} & \begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \end{array} \\ \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\alpha))$$

OU on a tree:



$$\mathbf{V}(\alpha)_{ij} = e^{-\alpha(V_i + V_j)} \frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

Tree Transformation Strategy (Pagel, 1999)

Strategy

- See $\mathbf{V}(\theta)$ as a BM on a modified tree.

Tree Transformation Strategy (Pagel, 1999)

Strategy

- See $\mathbf{V}(\theta)$ as a BM on a modified tree.
- For fixed $\mathbf{V}(\theta)$, use the fast algorithm.

Tree Transformation Strategy (Pagel, 1999)

Strategy

- See $\mathbf{V}(\theta)$ as a BM on a modified tree.
- For fixed $\mathbf{V}(\theta)$, use the fast algorithm.
- Optimize on $\mathbf{V}(\theta)$ numerically.

Tree Transformation Strategy (Pagel, 1999)

Strategy

- See $\mathbf{V}(\theta)$ as a BM on a modified tree.
- For fixed $\mathbf{V}(\theta)$, use the fast algorithm.
- Optimize on $\mathbf{V}(\theta)$ numerically.

Limits

- In general, only works for ultrametric trees.

Tree Transformation Strategy (Pagel, 1999)

Strategy

- See $\mathbf{V}(\boldsymbol{\theta})$ as a BM on a modified tree.
- For fixed $\mathbf{V}(\boldsymbol{\theta})$, use the fast algorithm.
- Optimize on $\mathbf{V}(\boldsymbol{\theta})$ numerically.

Limits

- In general, only works for ultrametric trees.
- Confidence intervals and tests only valid conditionally on $\boldsymbol{\theta}$.

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{V}\mathbf{X})^{-1}]_{kk}}} \sim \mathcal{T}_{n-p}$$

Pagel's Lambda

Relax the BM variance structure:

$$\mathbf{V}(\lambda)_{ii} = \mathbf{V}_{ii}$$
$$\mathbf{V}(\lambda)_{ij} = \lambda \mathbf{V}_{ij}$$

Pagel's Lambda

Relax the BM variance structure:

$$\begin{aligned}\mathbf{V}(\lambda)_{ii} &= \mathbf{V}_{ii}\\ \mathbf{V}(\lambda)_{ij} &= \lambda\mathbf{V}_{ij}\end{aligned}$$

Equivalent to running a BM on a modified tree with:

$$t(\lambda)_i = \begin{cases}\lambda t_i & \text{if } i \text{ internal node}\\ \lambda t_i + (1-\lambda)V_i & \text{if } i \text{ leaf}\end{cases}$$

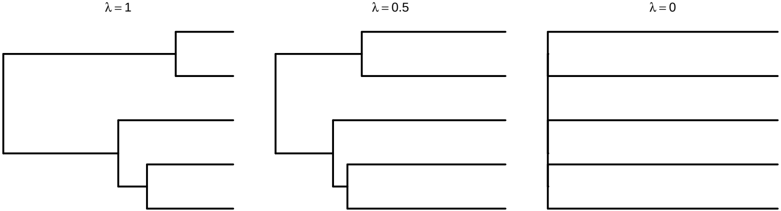
Pagel's Lambda

Relax the BM variance structure:

$$\mathbf{V}(\lambda)_{ii} = \mathbf{V}_{ii}$$
$$\mathbf{V}(\lambda)_{ij} = \lambda \mathbf{V}_{ij}$$

Equivalent to running a BM on a modified tree with:

$$t(\lambda)_i = \begin{cases} \lambda t_i & \text{if } i \text{ internal node} \\ \lambda t_i + (1-\lambda) V_i & \text{if } i \text{ leaf} \end{cases}$$



Pagel's Lambda

- Used a lot in practice.

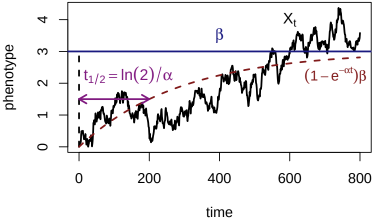
Pagel's Lambda

- Used a lot in practice.
- Also exist Pagel's κ and δ.

Pagel's Lambda

- Used a lot in practice.
- Also exist Pagel's κ and δ.
- λ "phylogenetic heritability" (see below).

Ornstein-Uhlenbeck



$$dX_t = \alpha[\beta - X_t]dt + \sigma dB_t$$

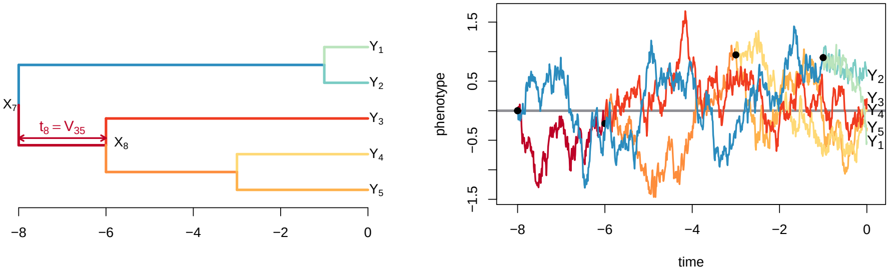
Deterministic part:

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

Stochastic part:

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

Ornstein-Uhlenbeck on a Tree (Hansen, 1997)



- SDE: $dX_t = \alpha[\beta - X_t]dt + \sigma dB_t$
- Covariances: $\mathbb{C}\text{ov}\left[Y_i; Y_j\right] = \frac{\sigma^2}{2\alpha}e^{-\alpha(V_i+V_j)}(e^{2\alpha V_{ij}} - 1)$
- Bounded variance $\gamma^2 = \frac{\sigma^2}{2\alpha}$
- Stationary state, Stabilizing selection

Ornstein-Uhlenbeck

Assuming a ultrametric tree with height T:

$$\mathbf{V}(\alpha)_{ij} = e^{-2\alpha T} \frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

Ornstein-Uhlenbeck

Assuming a ultrametric tree with height T:

$$\mathbf{V}(\alpha)_{ij} = e^{-2\alpha T} \frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

Equivalent to running a BM on a modified tree with:

$$V_i(\alpha) = e^{-2\alpha T} \frac{e^{2\alpha V_i} - 1}{2\alpha}$$

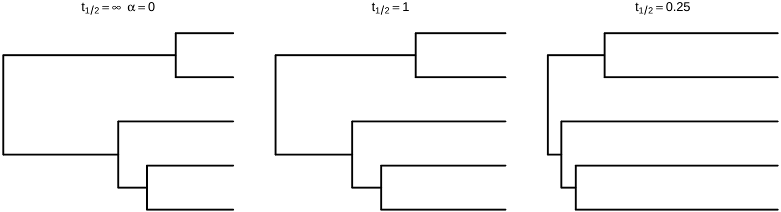
Ornstein-Uhlenbeck

Assuming a ultrametric tree with height T:

$$\mathbf{V}(\alpha)_{ij} = e^{-2\alpha T}\frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

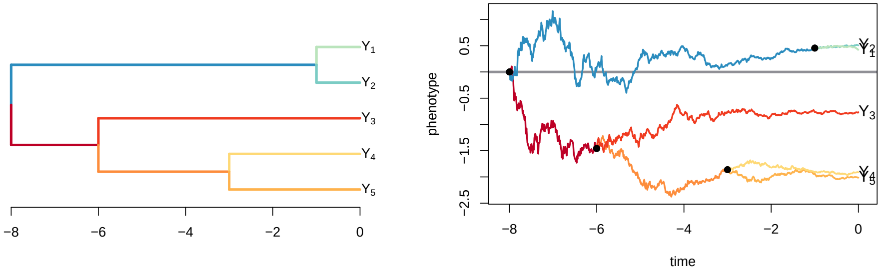
Equivalent to running a BM on a modified tree with:

$$V_i(\alpha) = e^{-2\alpha T}\frac{e^{2\alpha V_i} - 1}{2\alpha}$$



Early Burst - AC/DC

(Harmon et al., 2010)



- SDE: $dX_t = \sigma_0 e^{rt/2} dB_t$
- Covariances: $\mathbb{C}\text{ov}\left[Y_i; Y_j\right] = \sigma_0^2 \frac{e^{rV_{ij}} - 1}{r}$
- $r < 0$: Early Burst (Decelerating)
- $r > 0$: Accelerating

Early Burst - AC/DC

$$\mathbf{V}(r)_{ij} = \frac{e^{rV_{ij}} - 1}{r}$$

Early Burst - AC/DC

$$\mathbf{V}(r)_{ij} = \frac{e^{rV_{ij}} - 1}{r}$$

Equivalent to running a BM on a modified tree with:

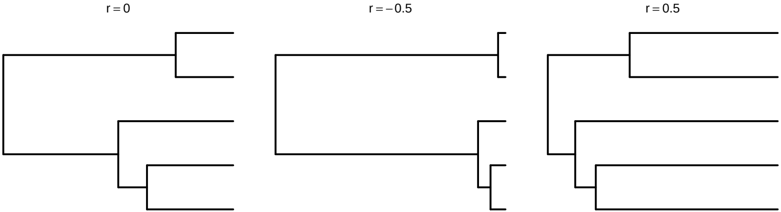
$$V_i(r) = \frac{e^{rV_i} - 1}{r}$$

Early Burst - AC/DC

$$\mathbf{V}(r)_{ij} = \frac{e^{rV_{ij}} - 1}{r}$$

Equivalent to running a BM on a modified tree with:

$$V_i(r) = \frac{e^{rV_i} - 1}{r}$$



Early Burst - AC/DC - Ornstein-Uhlenbeck (Uyeda et al., 2015)

OU:

$$\sigma^2 \mathbf{V}(\alpha)_{ij} = \sigma^2 e^{-2\alpha T} \frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

AC/DC:

$$\sigma_0^2 \mathbf{V}(r)_{ij} = \sigma_0^2 \frac{e^{r V_{ij}} - 1}{r}$$

Early Burst - AC/DC - Ornstein-Uhlenbeck (Uyeda et al., 2015)

OU:

$$\sigma^2 \mathbf{V}(\alpha)_{ij} = \sigma^2 e^{-2\alpha T} \frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

AC/DC:

$$\sigma_0^2 \mathbf{V}(r)_{ij} = \sigma_0^2 \frac{e^{r V_{ij}} - 1}{r}$$

On a ultrametric tree height T, AC/DC is equivalent to OU with:

$$r = 2\alpha$$
$$\sigma_0^2 = \sigma^2 e^{-2\alpha T}$$

Early Burst - AC/DC - Ornstein-Uhlenbeck (Uyeda et al., 2015)

OU:

$$\sigma^2\mathbf{V}(\alpha)_{ij} = \sigma^2 e^{-2\alpha T}\frac{e^{2\alpha V_{ij}} - 1}{2\alpha}$$

AC/DC:

$$\sigma_0^2\mathbf{V}(r)_{ij} = \sigma_0^2\frac{e^{r V_{ij}} - 1}{r}$$

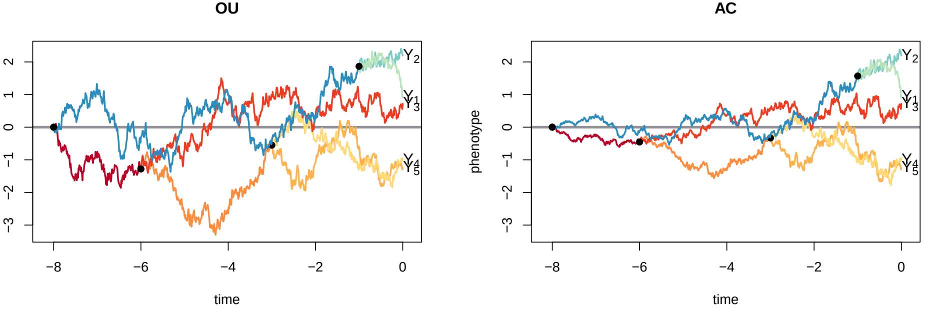
On a ultrametric tree height T, AC/DC is equivalent to OU with:

$$r = 2\alpha$$
$$\sigma_0^2 = \sigma^2 e^{-2\alpha T}$$

Note: case $r < 0$ (EB) is a ”repulsing OU”.

Warning: Identifiability

History is different, extent distribution is the same !

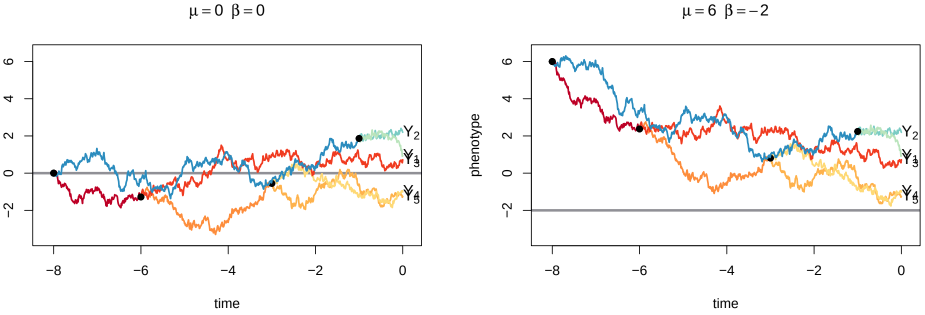


Many identifiability issues on ultrametric trees.

Warning: Identifiability (Ho and Ané, 2014b)

OU Expectation: All tips have the same expectation

$$\mathbb{E}[Y_i] = \mu e^{-\alpha T} + \beta(1 - e^{-\alpha T})$$



Only $\lambda = \mu e^{-\alpha T} + \beta(1 - e^{-\alpha T})$ is identifiable.

Simulated Example - OU

```
set.seed(12891026)
n <- 100
# Tree
tree <- rphylo(n, 0.1, 0)
# Parameters
sigma2 <- 1
alpha <- 0.05
sigma2err <- 1
# Trait
trait <- rTrait(1, tree, "OU",
                parameters = list(ancestral.state = 0,
                                  sigma2 = sigma2,
                                  alpha = alpha,
                                  optimal.value = 0))
# Non phylogenetic noise
trait <- trait + rnorm(n, mean = 0, sqrt(sigma2err))
```

Simulated Example - OU

```
# Phylogenetic Half-Life
log(2) / alpha

## [1] 13.86294

# Height of the tree
vcv(tree)[1, 1]

## [1] 44.10044

# scaled half-life
log(2) / alpha / vcv(tree)[1, 1]

## [1] 0.3143493
```

```
# stationary variance
sigma2 / (2 * alpha)

## [1] 10

# Ratio noise / phylo variance
sigma2err / (sigma2 / (2 * alpha))

## [1] 0.1
```

Simulated Example: True Model

```
phylolm(trait ~ 1,  phy = tree, model = "OUfixedRoot",
        measurement_error = TRUE)

## Call:
## phylolm(formula = trait ~ 1, phy = tree, model = "OUfixedRoot",
##     measurement_error = TRUE)
##
##    AIC logLik
##  472.8 -232.4
##
## Parameter estimate(s) using ML:
## alpha: 0.05781095
## sigma2: 1.124237
## sigma2_error: 0.3055167
##
## Coefficients:
## (Intercept)
##   0.1523982
```

Simulated Example: Likelihood

```
all_fits <- list(
  BM = phylolm(trait ~ 1, phy = tree, model = "BM"),
  BMerr = phylolm(trait ~ 1, phy = tree, model = "BM",
                  measurement_error = TRUE),
  OU = phylolm(trait ~ 1,  phy = tree, model = "OUfixedRoot"),
  OUerr = phylolm(trait ~ 1,  phy = tree, model = "OUfixedRoot",
                  measurement_error = TRUE))
```

```
sapply(all_fits, logLik)

##        BM       BMerr     OU        OUerr
## logLik -243.131 -236.6247 -232.5509 -232.3986
## df     2        3         3         4
```

Simulated Example: Likelihood

```
all_fits <- list(
  BM = phylolm(trait ~ 1, phy = tree, model = "BM"),
  BMerr = phylolm(trait ~ 1, phy = tree, model = "BM",
                  measurement_error = TRUE),
  OU = phylolm(trait ~ 1,  phy = tree, model = "OUfixedRoot"),
  OUerr = phylolm(trait ~ 1,  phy = tree, model = "OUfixedRoot",
                  measurement_error = TRUE))
```

```
sapply(all_fits, logLik)

##         BM       BMerr     OU        OUerr
## logLik -243.131 -236.6247 -232.5509 -232.3986
## df     2        3         3         4
```

→ Likelihood is always increases when df increases.

Simulated Example: AIC

$$AIC = -2logLik + 2df$$

```
sapply(all_fits, AIC)

##       BM    BMerr       OU    OUerr
## 490.2619 479.2493 471.1018 472.7971
```

→ Select model with the smallest AIC.

Regime Painting

Phylogenetic ANOVA:

- Test for a known group structure (habitat, ...).
- Each group has its own expectation.

Regime Painting

Phylogenetic ANOVA:

- Test for a known group structure (habitat, ...).
- Each group has its own expectation.

Regime Painting:

- Test for a known group structure (habitat, ...).
- Each group at the tip has its own process parameter (variance, selection strength, ...)

Regime Painting

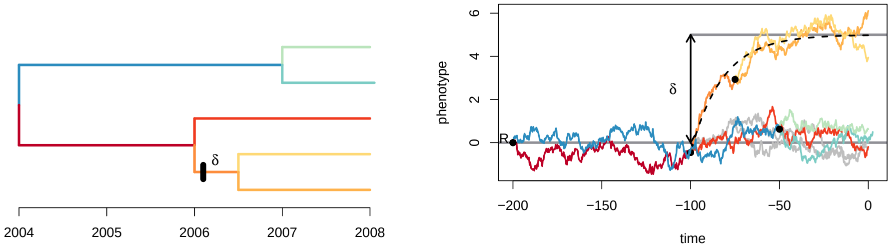
Phylogenetic ANOVA:

- Test for a known group structure (habitat, ...).
- Each group has its own expectation.

Regime Painting:

- Test for a known group structure (habitat, ...).
- Each group at the tip has its own process parameter (variance, selection strength, ...)
- Model selection using LRT, AIC, BIC, ...

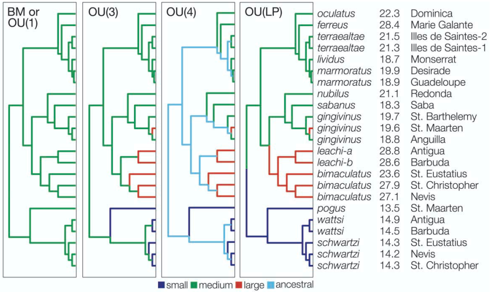
Example: OU optima

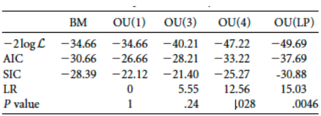


Regime Painting: Can we test that orange tips have a different optimum than others ?

Example: OU optima (Butler and King, 2004)

Test of several evolutionary hypotheses for species of *Anolis bimaculatus*.





	BM	OU(1)	OU(3)	OU(4)	OU(LP)
$-2\log\mathcal{L}$	−34.66	−34.66	−40.21	−47.22	−49.69
AIC	−30.66	−26.66	−28.21	−33.22	−37.69
SIC	−28.39	−22.12	−21.40	−25.27	-30.88
LR		0	5.55	12.56	15.03
P value		1	.24	.028	.0046

Regime Painting

Regime Painting:

- Test for a known group structure (habitat, ...).
- Each group at the tip has its own process parameter (variance, selection strength, ...)
- Model selection using LRT, AIC, BIC, ...

R packages:

- OUwie (univariate) Beaulieu et al. (2012)
- mvMORPH (multivariate) Clavel et al. (2015)

Regime Painting

Regime Painting:

- Test for a known group structure (habitat, ...).
- Each group at the tip has its own process parameter (variance, selection strength, ...)
- Model selection using LRT, AIC, BIC, ...

R packages:

- OUwie (univariate) Beaulieu et al. (2012)
- mvMORPH (multivariate) Clavel et al. (2015)

Problem: Shifts (regimes) must be fixed a priori.

Phylogenetic Mixed Model (Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^p \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\theta))$$

Phylogenetic Mixed Model

(Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}^p \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\theta))$$

Phylogenetic Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}^p + s\mathbf{E}^e \quad \mathbf{E}^e \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Phylogenetic Mixed Model

(Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}^p \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\theta))$$

Phylogenetic Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}^p + s\mathbf{E}^e \quad \mathbf{E}^e \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Distinction between:

- The heritable part of the trait

Phylogenetic Mixed Model

(Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^{p} \quad \mathbf{E}^{p} \sim \mathcal{N}(\mathbf{0}_{n}, \mathbf{V}(\boldsymbol{\theta}))$$

Phylogenetic Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^{p} + s\mathbf{E}^{e} \quad \mathbf{E}^{e} \sim \mathcal{N}(\mathbf{0}_{n}, \mathbf{I})$$

Distinction between:

- The heritable part of the trait
    - Variance $\sigma^2\mathbf{V}(\boldsymbol{\theta})$ given by the trait evolution on the tree

Phylogenetic Mixed Model

(Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^p \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\boldsymbol{\theta}))$$

Phylogenetic Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^p + s\mathbf{E}^e \quad \mathbf{E}^e \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Distinction between:

- The heritable part of the trait
  - Variance $\sigma^2\mathbf{V}(\boldsymbol{\theta})$ given by the trait evolution on the tree
- The environemental part of the trait

Phylogenetic Mixed Model

(Lynch, 1991; Housworth et al., 2004)

Phylogenetic Regression

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^p \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\boldsymbol{\theta}))$$

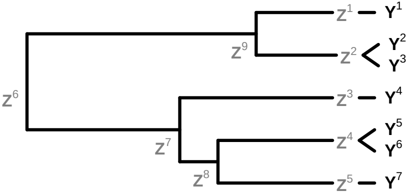
Phylogenetic Mixed Model

$$\mathbf{Y} = \mathbf{X}\beta + \sigma \mathbf{E}^p + s\mathbf{E}^e \quad \mathbf{E}^e \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

Distinction between:

- The heritable part of the trait
  - Variance $\sigma^2\mathbf{V}(\boldsymbol{\theta})$ given by the trait evolution on the tree
- The environemental part of the trait
  - Variance iid s^2 given by the error model

BM on a Ultrametric Tree

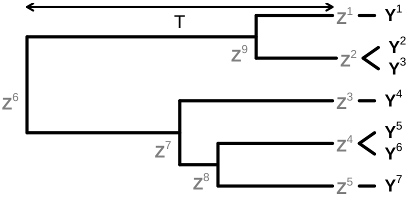


$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \;\middle|\; \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \;\middle|\; \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, s^2\right) \qquad \text{obs}$$

BM on a Ultrametric Tree



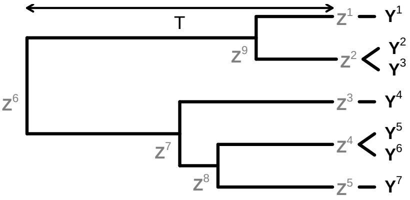
$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, s^2\right) \qquad \text{obs}$$

Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

BM on a Ultrametric Tree



$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, s^2\right) \qquad \text{obs}$$

Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

Phylogenetic Heritability:

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

BM on a Ultrametric Tree

Assume a ultrametric tree with height T.

BM on a Ultrametric Tree

Assume a ultrametric tree with height T.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with} \quad \text{Diag}(\mathbf{V}) = T\mathbf{I}$$

BM on a Ultrametric Tree

Assume a ultrametric tree with height T.

BM PMM total variance is:

$$\sigma^2 \mathbf{V} + s^2 \mathbf{I} \quad \text{with} \quad \text{Diag}(\mathbf{V}) = T\mathbf{I}$$

Phylogenetic Variance: $\sigma^2 T$
Total Variance: $\sigma^2 T + s^2$

BM on a Ultrametric Tree

Assume a ultrametric tree with height T.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with} \quad \text{Diag}(\mathbf{V}) = T\mathbf{I}$$

Phylogenetic Variance: $\sigma^2 T$
Total Variance: $\sigma^2 T + s^2$

The phylogenetic heritability is defined as:

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

Pagel's lambda a Ultrametric Tree

Assume a ultrametric tree with height T.

Pagel's lambda a Ultrametric Tree

Assume a ultrametric tree with height T.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1 - \lambda) T \mathbf{I}\right]$$

Pagel's lambda a Ultrametric Tree

Assume a ultrametric tree with height T.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1 - \lambda) T \mathbf{I}\right]$$

BM PMM total variance is:

$$\sigma^2 \mathbf{V} + s^2 \mathbf{I} = (\sigma^2 + s^2/T) \left[\frac{\sigma^2}{\sigma^2 + s^2/T} \mathbf{V} + \left(1 - \frac{\sigma^2}{\sigma^2 + s^2/T}\right) T \mathbf{I}\right]$$

Pagel's lambda a Ultrametric Tree

Assume a ultrametric tree with height T.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda) T \mathbf{I}\right]$$

BM PMM total variance is:

$$\sigma^2 \mathbf{V} + s^2 \mathbf{I} = (\sigma^2 + s^2/T) \left[\frac{\sigma^2}{\sigma^2 + s^2/T} \mathbf{V} + \left(1 - \frac{\sigma^2}{\sigma^2 + s^2/T}\right) T \mathbf{I}\right]$$

The two models are equivalent with:

$$\lambda = \frac{\sigma^2 T}{\sigma^2 T + s^2} = h^2 \qquad \sigma_\lambda^2 = \sigma^2 + s^2/T$$

Pagel's lambda a Ultrametric Tree

Assume a ultrametric tree with height T.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda) T \mathbf{I}\right]$$

BM PMM total variance is:

$$\sigma^2 \mathbf{V} + s^2 \mathbf{I} = (\sigma^2 + s^2/T) \left[\frac{\sigma^2}{\sigma^2 + s^2/T} \mathbf{V} + \left(1 - \frac{\sigma^2}{\sigma^2 + s^2/T}\right) T \mathbf{I}\right]$$

The two models are equivalent with:

$$\lambda = \frac{\sigma^2 T}{\sigma^2 T + s^2} = h^2 \qquad \sigma_\lambda^2 = \sigma^2 + s^2/T$$

λ is the heritability

Phylogenetic Mixed Model

Phylogenetic Mixed Model

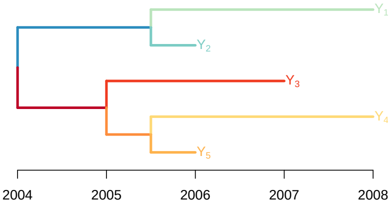
$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E}^p + s\mathbf{E}^e \quad \mathbf{E}^p \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V}(\theta)) \quad \mathbf{E}^e \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I})$$

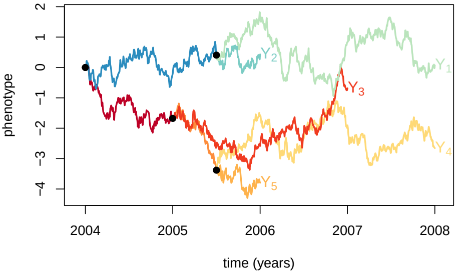
- Works with any phylogenetic model (BM, OU , EB, ...)
- For a ultrametric tree: tree-transformation.
- For a non-ultrametric tree: more subtle.

+

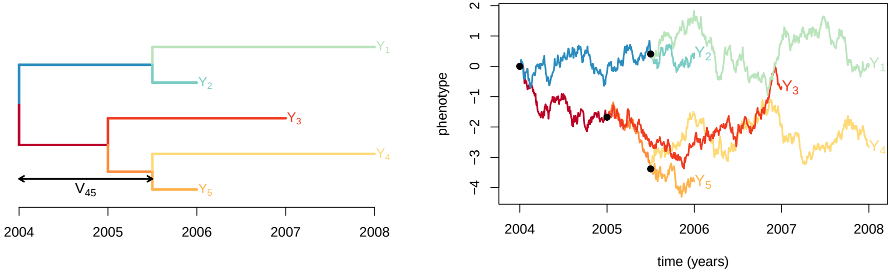
Outline

BM on a Tree


Y_1
Y_2
Y_3
Y_4
Y_5
2004
2005
2006
2007
2008


phenotype
2
1
0
−1
−2
−3
−4
Y_1
Y_2
Y_3
Y_4
Y_5
2004
2005
2006
2007
2008
time (years)

BM on a Tree

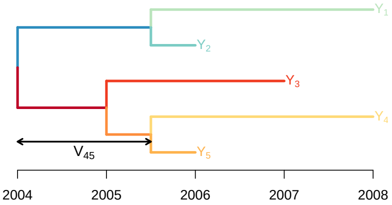


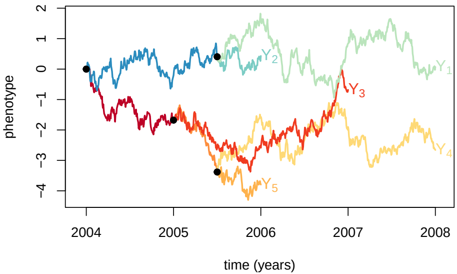
SDE: $\mathrm{d}\,X_t = \sigma\,\mathrm{d}\,B_t$

Variance: $\mathbb{C}\mathrm{ov}\left[Y_4; Y_5\right] = \sigma^2 \times V_{45}$ shared evolution time

Expectation: $\mathbb{E}\left[Y_i\right] = \mu$ ancestral root value

BM on a Tree

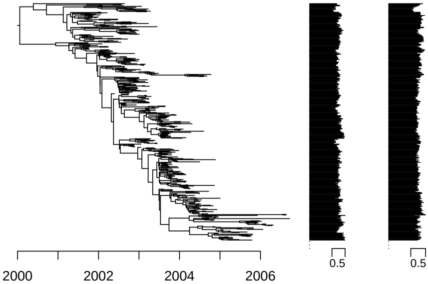




Distribution: Normal

$$\mathbf{Y} \sim \mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathbf{V})$$

Multivariate BM

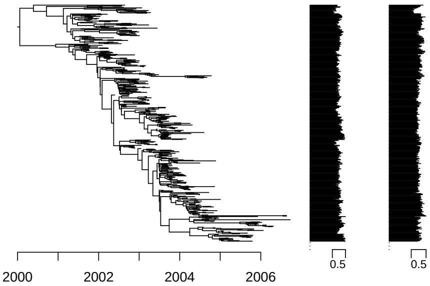


Data: Vectors of p traits

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

Tree: Influenza H3N2 (Lemey et al., 2014)

Multivariate BM



Data: Vectors of p traits

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

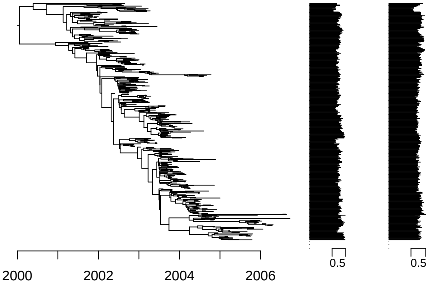
SDE: $\mathrm{d}\,\mathbf{X}_t = \boldsymbol{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$ $\mathbf{R} = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$

Variance: $\mathbb{C}\mathrm{ov}\left[Y_{ik}; Y_{jl}\right] = R_{kl} \times V_{ij}$ shared evolution time

Expectation: $\mathbb{E}\left[\mathbf{Y}_{\cdot k}\right] = \mu_k$ ancestral root value

Tree: Influenza H3N2 (Lemey et al., 2014)

Multivariate BM



Data: Vectors of p traits

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

Distribution: Matrix Normal

$$\mathbf{Y} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R}) \qquad \mathbb{V}\mathrm{ar}\left[\mathrm{vec}(\mathbf{Y})\right] = \mathbf{R} \otimes \mathbf{V}$$

Tree: Influenza H3N2 (Lemey et al., 2014)

Vectorisation: stack columns on top of each other

$$\operatorname{vec}(\mathbf{Y}) = \operatorname{vec}\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{pmatrix} = \begin{pmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{n1} \\ Y_{12} \\ Y_{22} \\ \vdots \\ Y_{n2} \\ \vdots \\ Y_{1p} \\ Y_{2p} \\ \vdots \\ Y_{np} \end{pmatrix} = \begin{pmatrix} \\ \mathbf{Y}_{\cdot 1} \\ \\ \\ \mathbf{Y}_{\cdot 2} \\ \\ \vdots \\ \\ \mathbf{Y}_{\cdot p} \\ \\ \end{pmatrix}$$

Kronecker Product

$$\mathbf{R} \otimes \mathbf{V} = \begin{pmatrix} R_{11}\mathbf{V} & \cdots & R_{1p}\mathbf{V} \\ \vdots & \ddots & \vdots \\ R_{p1}\mathbf{V} & \cdots & R_{pp}\mathbf{V} \end{pmatrix}$$

Kronecker Product

$$\mathbf{R} \otimes \mathbf{V} = \begin{pmatrix} R_{11}\mathbf{V} & \cdots & R_{1p}\mathbf{V} \\ \vdots & \ddots & \vdots \\ R_{p1}\mathbf{V} & \cdots & R_{pp}\mathbf{V} \end{pmatrix}$$

$$\mathbb{C}\mathrm{ov}\left[Y_{ik}; Y_{jl}\right] = R_{kl} \times V_{ij} \qquad \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot k}; \mathbf{Y}_{\cdot l}\right] = R_{kl} \times \mathbf{V}$$

Kronecker Product

$$\mathbf{R} \otimes \mathbf{V} = \begin{pmatrix} R_{11}\mathbf{V} & \cdots & R_{1p}\mathbf{V} \\ \vdots & \ddots & \vdots \\ R_{p1}\mathbf{V} & \cdots & R_{pp}\mathbf{V} \end{pmatrix}$$

$$\mathbb{C}\mathrm{ov}\left[Y_{ik}; Y_{jl}\right] = R_{kl} \times V_{ij} \qquad \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot k}; \mathbf{Y}_{\cdot l}\right] = R_{kl} \times \mathbf{V}$$

$$\mathbb{V}\mathrm{ar}\left[\begin{pmatrix} \mathbf{Y}_{\cdot 1} \\ \vdots \\ \mathbf{Y}_{\cdot p} \end{pmatrix}\right] = \begin{pmatrix} \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot 1}; \mathbf{Y}_{\cdot 1}\right] & \cdots & \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot 1}; \mathbf{Y}_{\cdot p}\right] \\ \vdots & \ddots & \vdots \\ \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot p}; \mathbf{Y}_{\cdot 1}\right] & \cdots & \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot p}; \mathbf{Y}_{\cdot p}\right] \end{pmatrix}$$

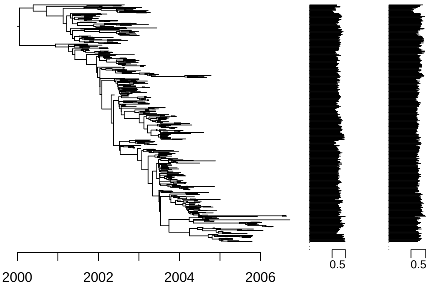
Kronecker Product

$$\mathbf{R} \otimes \mathbf{V} = \begin{pmatrix} R_{11}\mathbf{V} & \cdots & R_{1p}\mathbf{V} \\ \vdots & \ddots & \vdots \\ R_{p1}\mathbf{V} & \cdots & R_{pp}\mathbf{V} \end{pmatrix}$$

$$\mathbb{C}\mathrm{ov}\left[Y_{ik}; Y_{jl}\right] = R_{kl} \times V_{ij} \qquad \mathbb{C}\mathrm{ov}\left[\mathbf{Y}_{\cdot k}; \mathbf{Y}_{\cdot l}\right] = R_{kl} \times \mathbf{V}$$

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

Multivariate BM



Data: Vectors of p traits

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

Distribution: Matrix Normal

$$\mathbf{Y} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R}) \qquad \mathbb{V}\mathrm{ar}\left[\mathrm{vec}(\mathbf{Y})\right] = \mathbf{R} \otimes \mathbf{V}$$

Tree: Influenza H3N2 (Lemey et al., 2014)

Estimators

$$\text{vec}(\mathbf{Y}) \sim \mathcal{MN}(\text{vec}(\mathbf{1}_n\boldsymbol{\mu}^T), \mathbf{R} \otimes \mathbf{V}) \qquad \text{with } \mathbf{V} \text{ known}$$

Estimators

$$\mathrm{vec}(\mathbf{Y}) \sim \mathcal{MN}(\mathrm{vec}(\mathbf{1}_n\boldsymbol{\mu}^T), \mathbf{R} \otimes \mathbf{V}) \qquad \text{with } \mathbf{V} \text{ known}$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

$$\mathrm{vec}(\mathbf{L}^{-1}\mathbf{Y}) \sim \mathcal{MN}(\mathrm{vec}(\mathbf{L}^{-1}\mathbf{1}_n\boldsymbol{\mu}^T), \mathbf{R} \otimes \mathbf{I}_n)$$

Estimators

$$\mathrm{vec}(\mathbf{Y}) \sim \mathcal{MN}(\mathrm{vec}(\mathbf{1}_n\boldsymbol{\mu}^T), \mathbf{R} \otimes \mathbf{V}) \qquad \text{with } \mathbf{V} \text{ known}$$

Cholesky Decomposition:

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$
$$\mathrm{vec}(\mathbf{L}^{-1}\mathbf{Y}) \sim \mathcal{MN}(\mathrm{vec}(\mathbf{L}^{-1}\mathbf{1}_n\boldsymbol{\mu}^T), \mathbf{R} \otimes \mathbf{I}_n)$$

Maximum Likelihood Estimators:

$$\hat{\boldsymbol{\mu}}^T = (\mathbf{1}_n^T\mathbf{V}^{-1}\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{V}^{-1}\mathbf{Y}$$
$$\hat{\mathbf{R}} = \frac{1}{n-1}(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)$$

Phylogenetic Principal Component Analysis (Revell, 2009)

Diagonalisation of the BM estimated variance:

$$\hat{\mathbf{R}} = \hat{\mathbf{W}}\hat{\mathbf{D}}^2\hat{\mathbf{W}}^T$$

- $\hat{\mathbf{W}}_i$: principal axis i (eigenvectors, $\hat{\mathbf{W}}$ is $p \times p$ orthogonal)
- $\hat{\mathbf{D}}_i^2$: variance explained by $\hat{\mathbf{W}}_i$ (eigenvalues, $\hat{\mathbf{D}}^2$ diagonal)

Phylogenetic Principal Component Analysis (Revell, 2009)

Diagonalisation of the BM estimated variance:

$$\hat{\mathbf{R}} = \hat{\mathbf{W}}\hat{\mathbf{D}}^2\hat{\mathbf{W}}^T$$

- $\hat{\mathbf{W}}_i$: principal axis i (eigenvectors, $\hat{\mathbf{W}}$ is $p \times p$ orthogonal)
- $\hat{\mathbf{D}}_i^2$: variance explained by $\hat{\mathbf{W}}_i$ (eigenvalues, $\hat{\mathbf{D}}^2$ diagonal)

Scores: Data coordinates in the new system.

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

Phylogenetic Principal Component Analysis (Revell, 2009)

Diagonalisation of the BM estimated variance:

$$\hat{\mathbf{R}} = \hat{\mathbf{W}}\hat{\mathbf{D}}^2\hat{\mathbf{W}}^T$$

- $\hat{\mathbf{W}}_i$: principal axis i (eigenvectors, $\hat{\mathbf{W}}$ is $p \times p$ orthogonal)
- $\hat{\mathbf{D}}_i^2$: variance explained by $\hat{\mathbf{W}}_i$ (eigenvalues, $\hat{\mathbf{D}}^2$ diagonal)

Scores: Data coordinates in the new system.

$$\hat{\mathbf{S}} = (\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)\hat{\mathbf{W}}$$

Scores are empirically phylogenetically un-correlated:

$$\hat{\mathbf{S}}^T\mathbf{V}^{-1}\hat{\mathbf{S}} =$$

Phylogenetic Principal Component Analysis (Revell, 2009)

Diagonalisation of the BM estimated variance:

$$\hat{\mathbf{R}} = \hat{\mathbf{W}}\hat{\mathbf{D}}^2\hat{\mathbf{W}}^T$$

- $\hat{\mathbf{W}}_i$: principal axis i (eigenvectors, $\hat{\mathbf{W}}$ is $p \times p$ orthogonal)
- $\hat{\mathbf{D}}_i^2$: variance explained by $\hat{\mathbf{W}}_i$ (eigenvalues, $\hat{\mathbf{D}}^2$ diagonal)

Scores: Data coordinates in the new system.

$$\hat{\mathbf{S}} = (\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)\hat{\mathbf{W}}$$

Scores are empirically phylogenetically un-correlated:

$$\begin{aligned}
\hat{\mathbf{S}}^T\mathbf{V}^{-1}\hat{\mathbf{S}} &= \hat{\mathbf{W}}^T(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)\hat{\mathbf{W}} \\
&= \hat{\mathbf{W}}^T[(n-1)\hat{\mathbf{R}}]\hat{\mathbf{W}} \\
&= (n-1)\hat{\mathbf{D}}^2
\end{aligned}$$

Phylogenetic Principal Component Analysis (Revell, 2009)

Diagonalisation of the BM estimated variance:

$$\hat{\mathbf{R}} = \hat{\mathbf{W}}\hat{\mathbf{D}}^2\hat{\mathbf{W}}^T$$

- $\hat{\mathbf{W}}_i$: principal axis i (eigenvectors, $\hat{\mathbf{W}}$ is $p \times p$ orthogonal)
- $\hat{\mathbf{D}}_i^2$: variance explained by $\hat{\mathbf{W}}_i$ (eigenvalues, $\hat{\mathbf{D}}^2$ diagonal)

Scores: Data coordinates in the new system.

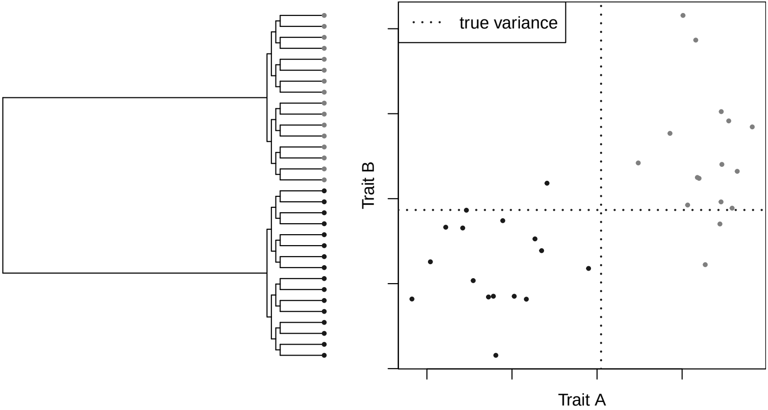
$$\hat{\mathbf{S}} = (\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T)\hat{\mathbf{W}}$$

Scores are empirically phylogenetically un-correlated:

$$\langle \mathbf{S}_i, \mathbf{S}_j \rangle_{\mathbf{V}^{-1}} = 0 \quad \forall\ i \neq j$$

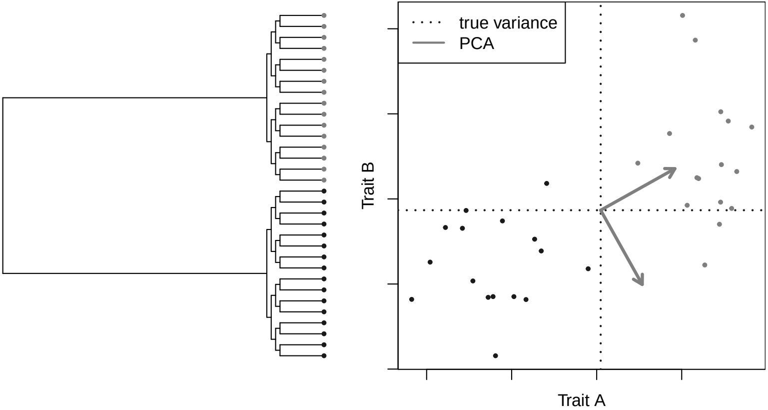
Phylogenetic Principal Component Analysis (Revell, 2009)

Simulation according to an uncorrelated BM.



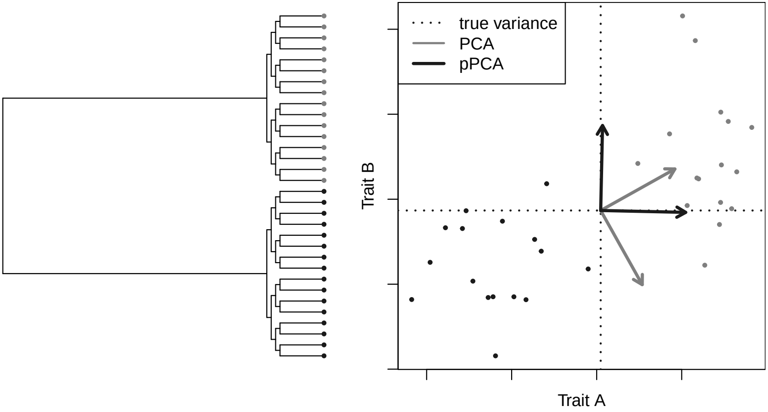
Phylogenetic Principal Component Analysis (Revell, 2009)

Simulation according to an uncorrelated BM.

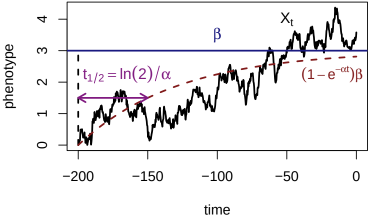


Phylogenetic Principal Component Analysis (Revell, 2009)

Simulation according to an uncorrelated BM.



Ornstein-Uhlenbeck Modeling (Hansen, 1997)



$$d X_t = \alpha[\beta - X_t] \, d t + \sigma \, d B_t$$

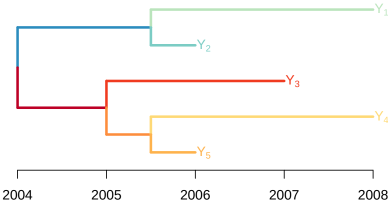
Deterministic part:

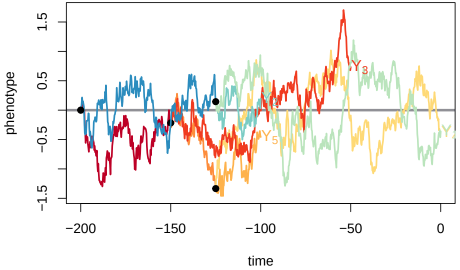
- β: primary optimum (mechanistically defined).
- $\ln(2)/\alpha$: phylogenetic half live.

Stochastic part:

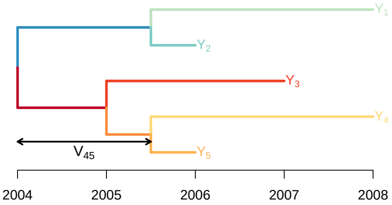
- X_t: trait value (actual optimum).
- $\sigma dB(t)$: Brownian fluctuations.

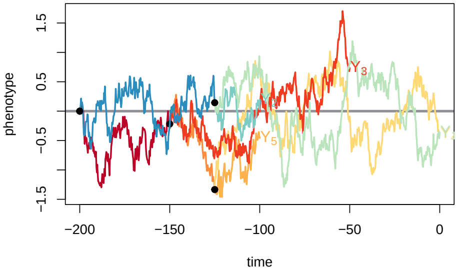
OU on a Tree


Y_1
Y_2
Y_3
Y_4
Y_5
2004
2005
2006
2007
2008


phenotype
1.5
0.5
−0.5
−1.5
Y_3
Y_2
Y_5
−200
−150
−100
−50
0
time

OU on a Tree





SDE: $\mathrm{d}\,X_t = \alpha[\beta - X_t]\,\mathrm{d}\,t + \sigma\,\mathrm{d}\,B_t$

Variance: $\mathbb{C}\mathrm{ov}\left[Y_4; Y_5\right] = \frac{\sigma^2}{2\alpha} e^{-\alpha(V_4+V_5)}(e^{2\alpha V_{45}} - 1)$

Expectation: $\mathbb{E}\left[Y_i\right] = \mu e^{-\alpha V_i} + \beta(1 - e^{-\alpha V_i})$

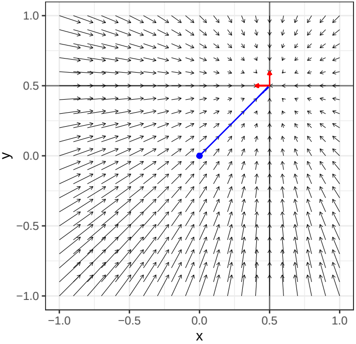
Multivariate OU Modeling

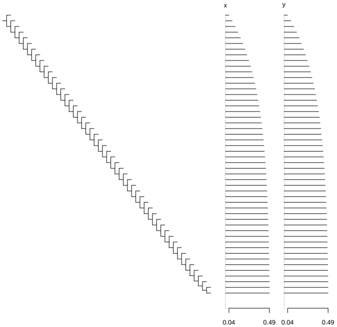
$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Multivariate OU Modeling

$$d\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,d\,t + \mathbf{\Sigma}\,d\,\mathbf{B}_t$$

Scalar: $\mathbf{A} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

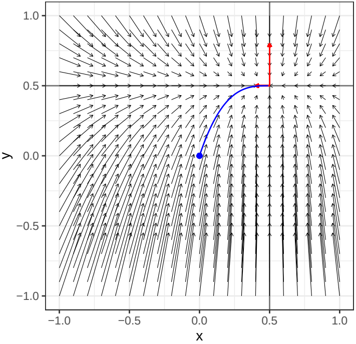


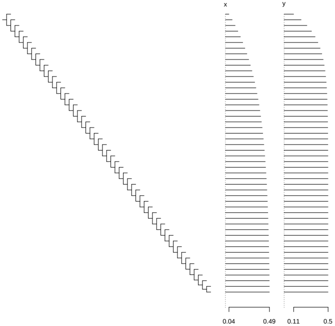


Multivariate OU Modeling

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonal: $\mathbf{A} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.3 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

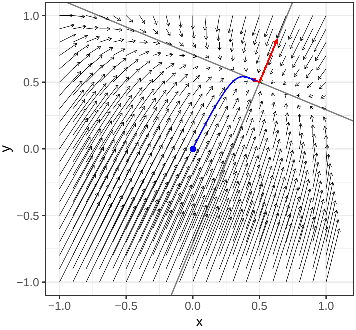


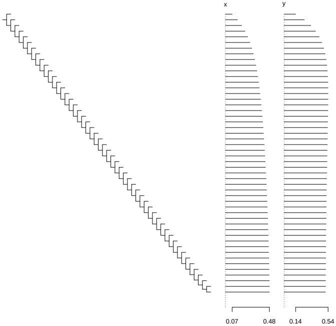


Multivariate OU Modeling

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Symmetric: $\mathbf{A} = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

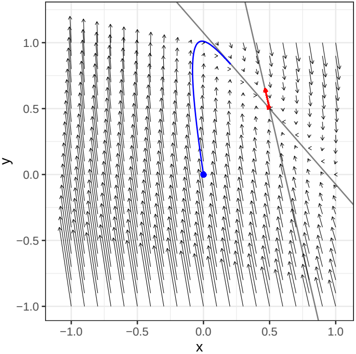


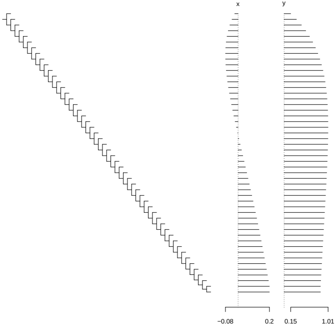


Multivariate OU Modeling

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable in $\mathbb{R}$: $\mathbf{A} = \begin{pmatrix} -0.02 & -0.04 \\ 0.2 & 0.2 \end{pmatrix}$ $\beta = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$





Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}\left[\mathbf{Y}_i\right] = \boldsymbol{\mu} e^{-\mathbf{A} V_i} + \boldsymbol{\beta}(1 - e^{-\mathbf{A} V_i})$

Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}\left[\mathbf{Y}_i\right] = \boldsymbol{\mu} e^{-\mathbf{A}V_i} + \boldsymbol{\beta}(1 - e^{-\mathbf{A}V_i})$

Variance: $\mathbb{C}\mathrm{ov}\left[\mathbf{Y}_i; \mathbf{Y}_j\right] = \mathbf{P}\left[\mathbf{W}_{ij} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right]\mathbf{P}^T$

$$\mathbf{W}_{ij} = \left[\frac{1}{\lambda_q + \lambda_r} e^{-\lambda_q V_i} e^{-\lambda_r V_j}\left(e^{(\lambda_q+\lambda_r)V_{ij}} - 1\right)\right]_{1 \le q,r \le p}$$

Multivariate OU

$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\boldsymbol{\beta} - \mathbf{X}_t]\,\mathrm{d}\,t + \boldsymbol{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Diagonalizable: $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} \quad \lambda_k > 0$

Expectation: $\mathbb{E}[\mathbf{Y}_i] = \boldsymbol{\mu} e^{-\mathbf{A}V_i} + \boldsymbol{\beta}(1 - e^{-\mathbf{A}V_i})$

Variance: $\mathbb{C}\mathrm{ov}[\mathbf{Y}_i; \mathbf{Y}_j] = \mathbf{P}\left[\mathbf{W}_{ij} \odot \mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-T}\right]\mathbf{P}^T$

$$\mathbf{W}_{ij} = \left[\frac{1}{\lambda_q + \lambda_r} e^{-\lambda_q V_i} e^{-\lambda_r V_j}\left(e^{(\lambda_q + \lambda_r)V_{ij}} - 1\right)\right]_{1 \le q,r \le p}$$

Distribution: Still Gaussian.

No nice Kronecker product.

General Model Pruning

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p$$

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p$$

- **Y** observed matrix of traits at the tips of the tree ($n \times p$)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p$$

- **Y** observed matrix of traits at the tips of the tree ($n \times p$)
- **X** matrix of regressors ($n \times k$)
- β matrix of coefficients ($k \times p$)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p$$

- $\mathbf{Y}$ observed matrix of traits at the tips of the tree ($n \times p$)
- $\mathbf{X}$ matrix of regressors ($n \times k$)
- β matrix of coefficients ($k \times p$)
- $\mathbf{E}^p$ matrix of phylogenetic errors ($n \times p$)

Phylogenetic Linear Regression

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p$$

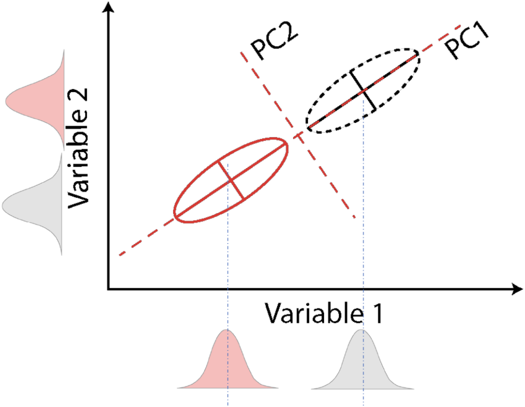
- $\mathbf{Y}$ observed matrix of traits at the tips of the tree $(n \times p)$
- $\mathbf{X}$ matrix of regressors $(n \times k)$
- β matrix of coefficients $(k \times p)$
- $\mathbf{E}^p$ matrix of phylogenetic errors $(n \times p)$
  - For a BM: $\mathbf{E}^p \sim \mathcal{MN}(\mathbf{0}_n, \mathbf{V}, \mathbf{R})$
  - For an OU: more complex.

Phylogenetic Linear Regression

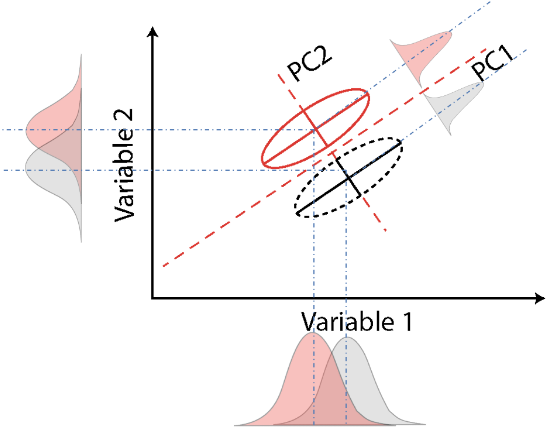
$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}^p + \mathbf{E}^e$$

- $\mathbf{Y}$ observed matrix of traits at the tips of the tree ($n \times p$)
- $\mathbf{X}$ matrix of regressors ($n \times k$)
- β matrix of coefficients ($k \times p$)
- $\mathbf{E}^p$ matrix of phylogenetic errors ($n \times p$)
    - For a BM: $\mathbf{E}^p \sim \mathcal{MN}(\mathbf{0}_n, \mathbf{V}, \mathbf{R})$
    - For an OU: more complex.
- $\mathbf{E}^e$ matrix of independent errors ($n \times p$)
    - $\mathbf{E}^e \sim \mathcal{MN}(\mathbf{0}_n, \mathbf{I}_n, s^2\mathbf{I}_p)$

Phylogenetic Multivariate Linear Regression and MANOVA



Phylogenetic Multivariate Linear Regression and MANOVA



Phylogenetic Multivariate Linear Regression Tests

The multivariate tests (analogous to the F tests seen previously), are all based on the eigenvalues $d_1, \ldots, d_s$ of the p by p matrix:

$$\mathbf{E}^{-1}\mathbf{H}$$

- s is the rank of $\mathbf{E}^{-1}\mathbf{H}$

Phylogenetic Multivariate Linear Regression Tests

The multivariate tests (analogous to the F tests seen previously), are all based on the eigenvalues $d_1, ..., d_s$ of the p by p matrix:

$$\mathbf{E}^{-1}\mathbf{H}$$

- s is the rank of $\mathbf{E}^{-1}\mathbf{H}$
- $\mathbf{E}$ is the "error" or "residual" Sum of Squares and Cross Products (SSCP) matrix, i.e. $\mathbf{E} = (n - q)\hat{\mathbf{R}}$ with q the dimensionality of the design.

Phylogenetic Multivariate Linear Regression Tests

The multivariate tests (analogous to the F tests seen previously), are all based on the eigenvalues $d_1, ..., d_s$ of the p by p matrix:

$$\mathbf{E}^{-1}\mathbf{H}$$

- s is the rank of $\mathbf{E}^{-1}\mathbf{H}$
- $\mathbf{E}$ is the "error" or "residual" Sum of Squares and Cross Products (SSCP) matrix, i.e. $\mathbf{E} = (n-q)\hat{\mathbf{R}}$ with q the dimensionality of the design.
- $\mathbf{H}$ is the "hypothesis" SSCP matrix. It is given by $\mathbf{H} = (\mathbf{X}\hat{\mathbf{B}} - \mathbf{X}_0\hat{\mathbf{B}}_0)^T\mathbf{V}^{-1}(\mathbf{X}\hat{\mathbf{B}} - \mathbf{X}_0\hat{\mathbf{B}}_0)$ where $\mathbf{X}_0$ is the design matrix corresponding to the null hypothesis and $\hat{\mathbf{B}}_0$ is the corresponding matrix of parameter estimates.

Phylogenetic Multivariate Linear Regression Tests (cont.)

Multiple dimensions mean also different ways of computing a statistic. The four most common multivariate statistics are:

- Wilk's Λ statistic for instance is given by $\Lambda = \prod_{i=1}^{s} \frac{1}{1+d_i}$

Phylogenetic Multivariate Linear Regression Tests (cont.)

Multiple dimensions mean also different ways of computing a statistic. The four most common multivariate statistics are:

- Wilk's Λ statistic for instance is given by $\Lambda = \prod_{i=1}^{s} \frac{1}{1+d_i}$
- Pillai's statistic is $V = \prod_{i=1}^{s} \frac{d_i}{1+d_i}$

Phylogenetic Multivariate Linear Regression Tests (cont.)

Multiple dimensions mean also different ways of computing a statistic. The four most common multivariate statistics are:

- Wilk's Λ statistic for instance is given by $\Lambda = \prod_{i=1}^{s} \frac{1}{1+d_i}$
- Pillai's statistic is $V = \prod_{i=1}^{s} \frac{d_i}{1+d_i}$
- Hotelling-Lawley is $U = \sum_{i=1}^{s} d_i$

Phylogenetic Multivariate Linear Regression Tests (cont.)

Multiple dimensions mean also different ways of computing a statistic. The four most common multivariate statistics are:

- Wilk's Λ statistic for instance is given by $\Lambda = \prod_{i=1}^{s} \frac{1}{1+d_i}$
- Pillai's statistic is $V = \prod_{i=1}^{s} \frac{d_i}{1+d_i}$
- Hotelling-Lawley is $U = \sum_{i=1}^{s} d_i$
- Roy's largest root is $\theta = \frac{d_1}{1+d_1}$

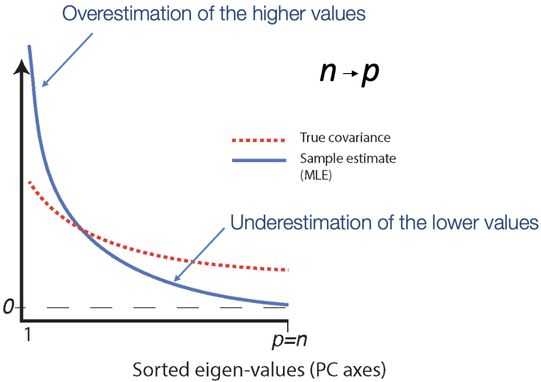
Phylogenetic Multivariate Linear Regression Tests (cont.)

Multiple dimensions mean also different ways of computing a statistic. The four most common multivariate statistics are:

- Wilk's Λ statistic for instance is given by $\Lambda = \prod_{i=1}^{s} \frac{1}{1+d_i}$
- Pillai's statistic is $V = \prod_{i=1}^{s} \frac{d_i}{1+d_i}$
- Hotelling-Lawley is $U = \sum_{i=1}^{s} d_i$
- Roy's largest root is $\theta = \frac{d_1}{1+d_1}$
- They all have the same type I error rate, but different power. Pillai's trace is often considered as a good and robust (e.g., to slight violation of homogeneity of covariance matrices) test. Wilks' Λ is connected to likelihood ratio tests (LRT) and has well known approximations of the F-statistic.

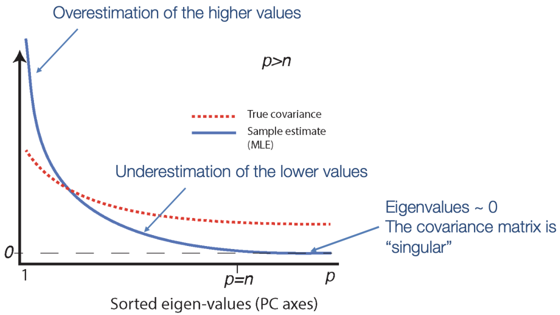
Multivariate models and the curse of dimensionality

Multivariate models estimate a lot of parameters (e.g., for the multivariate BM there's $p(p+1)/2+p$ d.f.). When p is approaching n the estimation of parameters deteriorates.



Multivariate models and the curse of dimensionality (cont.)

When p is equal or larger than n, the covariance matrix $\hat{\mathbf{R}}$ is singular. We can't invert it, nor compute the log determinant.



Multivariate models and the curse of dimensionality (cont.)

Data dimension is often reduced through Principal Component Analysis. But...

- No trivial ways to select the number of relevant PCs (eigenvalues are biased, p approach n problem...)

Multivariate models and the curse of dimensionality (cont.)

Data dimension is often reduced through Principal Component Analysis. But...

- No trivial ways to select the number of relevant PCs (eigenvalues are biased, p approach n problem...)
- Composite traits are sometime difficult to interpret.

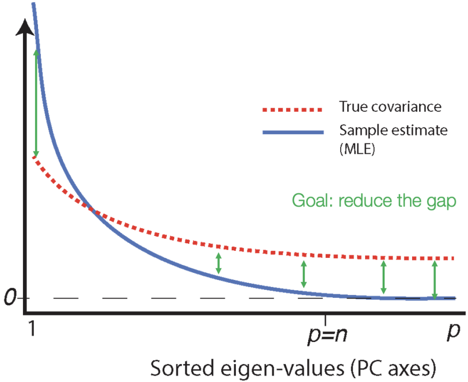
Multivariate models and the curse of dimensionality (cont.)

Data dimension is often reduced through Principal Component Analysis. But...

- No trivial ways to select the number of relevant PCs (eigenvalues are biased, p approach n problem...)
- Composite traits are sometime difficult to interpret.
- Problems with PCMs: model selection, parameter estimation, and statistical tests (Uyeda et al., 2015; Clavel and Morlon, 2020)

Multivariate models and the curse of dimensionality (cont.)

We can use regularization or penalized likelihood approach



Penalized likelihood

γ = regularization parameter (control the amount of penalization)

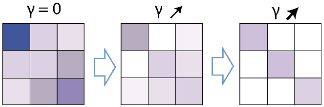
R = Evolutionary covariance matrix

$$\log(\boldsymbol{\mathcal{L}_P}) = \log(\boldsymbol{\mathcal{L}}) - \frac{\gamma}{2}\,\|\boldsymbol{R}^{-1}\|^{q}$$

q = various penalty (1 is LASSO; 2 is Ridge)

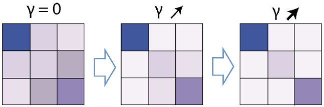
Penalized likelihood

LASSO (e.g., Tibshirani 1996; Friedman et al. 2008)



Sparsity structure

Ridge (e.g., Hoerl & Kennard 1970; Warton 2008)



Regularization

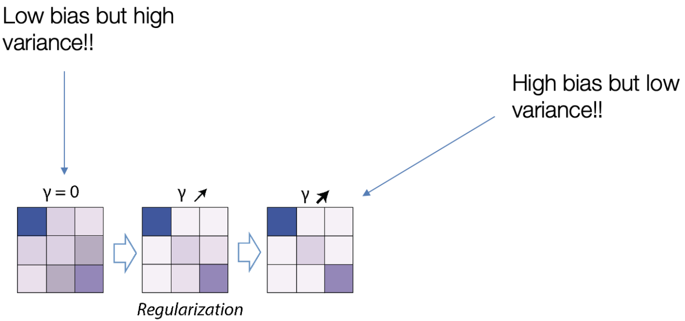
LASSO (least absolute shrinkage and selection operator)

Reduce the variance and remove the non-significant entries (=noise). Produce a sparse and well conditioned estimate

Ridge

Homogenize and reduce the variance (=noise). Produce a well conditioned estimate

Penalized likelihood

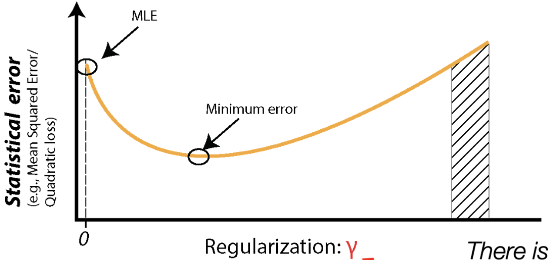


Low bias but high variance!!

High bias but low variance!!

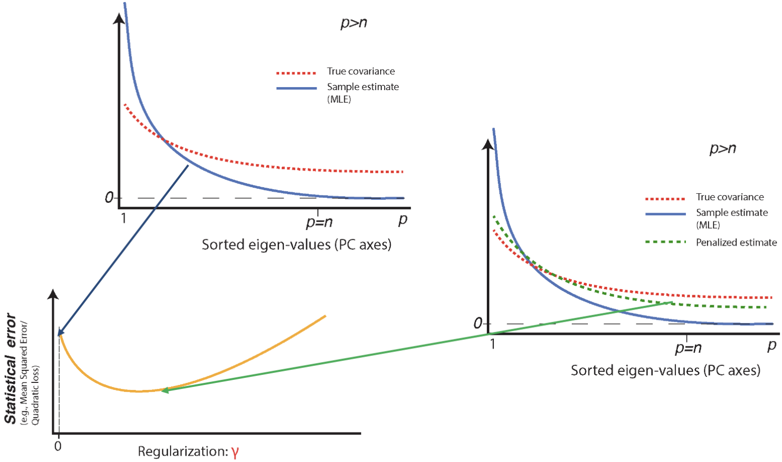
Penalized likelihood

$$\log(\mathcal{L}_P) = \log(\mathcal{L}) - \frac{\gamma}{2} \|\boldsymbol{R}^{-1}\|^q$$



There is an optimal value with less statistical loss than the MLE

Penalized likelihood



Penalized likelihood

The penalized log-likelihood is maximized when the regularization parameter γ is set to "0". This is because the parameters that enter this likelihood are estimated from the same data used to evaluate it.

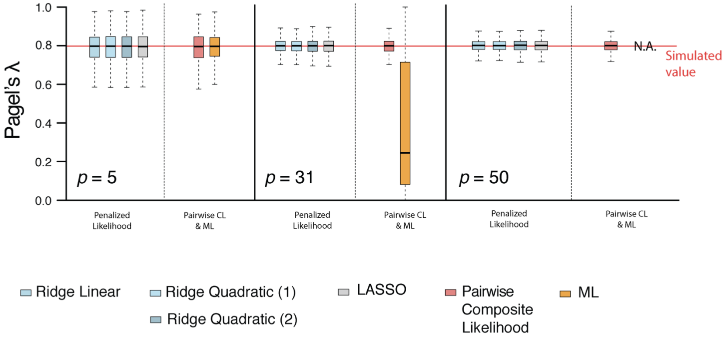
$$log(\mathcal{L}_p) = log(\mathcal{L}) - \frac{\gamma}{2}\left\|\mathbf{R}^{-1}\right\|^q$$

We can use instead a cross-validated likelihood to jointly infer the penalization and model' parameters:

$$log\mathcal{L}(\gamma, x_1, ..., x_n)_{CV} = -\frac{1}{n}\sum_{i=1}^{n} log\mathcal{L}(x_i|\mathbf{R}(\gamma)_{(-i)})$$

Penalized likelihood - simulations

Comparison to Maximum Likelihood (ML) and Pairwise Composite Likelihood (PCL) for $n = 32$ species trees



Model comparison

Penalized likelihood (as well as ML) models fit can be compared using the GIC (Generalized Information Criterion; Konishi and Kitagawa, 1996, 2008):

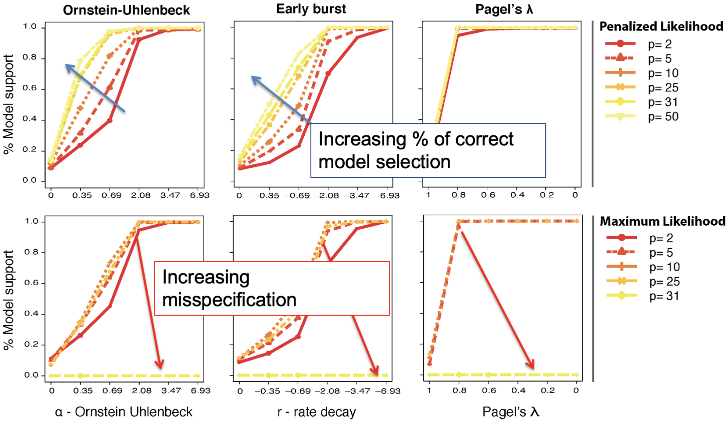
$$GIC = -2 log \mathcal{L}(X|\hat{\theta}) + 2tr(\mathbf{J}^{-1}\mathbf{I})$$

The last term represents the bias/"effective number of d.f."

$$\mathbf{J} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}(x_i|\theta) - \gamma \mathbf{P}(\theta)}{\partial \theta^2}|_{\theta=\hat{\theta}}$$

$$\mathbf{I} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \mathcal{L}(x_i|\theta) - \gamma \mathbf{P}(\theta)}{\partial \theta} \frac{\partial \mathcal{L}(x_i|\theta)}{\partial \theta}|_{\theta=\hat{\theta}}$$

Penalized likelihood - model comparison simulations



Increasing % of correct model selection

Increasing misspecification

High-dimensional multivariate PCMs empirical example

Bibliography I

Bastide, Mariadassou, Robin. 2022. In: Didier, Guindon, editors, Modèles et méthodes pour l'évolution biologique, ISTE Group, pp. 47–85.

Beaulieu, Jhwueng, Boettiger, et al. 2012. *Evolution*. 66:2369–2383.

Butler, King. 2004. *The American Naturalist*. 164:683–695.

Clavel, Escarguel, Merceron. 2015. *Methods in Ecology and Evolution*. 6:1311–1319.

Clavel, Morlon. 2020. *Systematic Biology*. 69:927–943.

Felsenstein. 2004. Sunderland, Massachusetts: Sinauer Associates.

Hansen. 1997. *Evolution*. 51:1341.

Harmon. 2019. Center for Open Science, version 1. edition.

Harmon, Losos, Davies, et al. 2010. *Evolution*. 64:2385–2396.

Ho, Ané. 2014a. *Systematic Biology*. 63:397–408.

Ho, Ané. 2014b. *Methods in Ecology and Evolution*. 5:1133–1146.

Housworth, Martins, Lynch. 2004. *The American Naturalist*. 163:84–96.

Bibliography II

Konishi, Kitagawa. 1996. *Biometrika*. 83:875–890.

Konishi, Kitagawa. 2008. Springer.

Lemey, Rambaut, Bedford, et al. 2014. *PLoS Pathogens*. 10:e1003932.

Leventhal, Bonhoeffer. 2016. *Trends in Microbiology*. 24:687–698.

Lynch. 1991. *Evolution*. 45:1065–1080.

Mitov, Stadler. 2018. *Molecular Biology and Evolution*. 35:756–772.

Pagel. 1999. *Nature*. 401:877–884.

Revell. 2009. *Evolution*. 63:3258–3268.

Uyeda, Caetano, Pennell. 2015. *Systematic Biology*. 64:677–689.

Continuous Trait Evolution

Paul Bastide[1], Julien Clavel[2]

[1] IMAG, Université de Montpellier, CNRS, `paul.bastide@umontpellier.fr`
[2] LEHNA, Université Lyon 1, CNRS, `julien.clavel@univ-lyon1.fr`

December 2022

Appendices

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

- y_i: quantitative response for i
- x_i: quantitative predicting variable for i
- ϵ_i: "error" for i Gaussian random variable

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

- y_i: quantitative response for i
- x_i: quantitative predicting variable for i
- ϵ_i: "error" for i Gaussian random variable
    - Centered: $\mathbb{E}[\epsilon_i] = 0$ for all i
    - Homoscedastic: $\mathbb{V}\text{ar}[\epsilon_i] = \sigma^2$ for all i
    - Independent: $\mathbb{C}\text{ov}[\epsilon_i; \epsilon_j] = 0$ for all $i \neq j$

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

- y_i: quantitative response for i
- x_i: quantitative predicting variable for i
- ϵ_i: "error" for i Gaussian random variable
  - Centered: $\mathbb{E}[\epsilon_i] = 0$ for all i
  - Homoscedastic: $\mathbb{V}\text{ar}[\epsilon_i] = \sigma^2$ for all i
  - Independent: $\mathbb{C}\text{ov}[\epsilon_i ; \epsilon_j] = 0$ for all $i \neq j$
  - iid: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \boldsymbol{\epsilon}$$

Simple Gaussian Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \forall 1 \leq i \leq n$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \boldsymbol{\epsilon}$$

- $\mathbf{Y} = (y_1, \dots, y_n)^T$ **random** vector of responses
- $\mathbf{1}_n = (1, \dots, 1)^T$ vector of ones
- $\mathbf{x} = (x_1, \dots, x_n)^T$ **non random** vector of predictors
- $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ **random iid Gaussian** vector of errors
- β_0, β_1 **non random, unknown** coefficients

Simple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Simple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Simple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Simple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\mathbf{X}$ $(n \times 2)$ matrix of regressors
- $\boldsymbol{\beta}$ (length 2) vector of coefficients

Multiple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + \beta_p \begin{pmatrix} x_{1p} \\ \vdots \\ x_{np} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Multiple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + \beta_p \begin{pmatrix} x_{1p} \\ \vdots \\ x_{np} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Multiple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + \beta_p \begin{pmatrix} x_{1p} \\ \vdots \\ x_{np} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

Multiple Gaussian Regression

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + \beta_p \begin{pmatrix} x_{1p} \\ \vdots \\ x_{np} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

- $\mathbf{X}$ $(n \times p)$ matrix of regressors $(rg(\mathbf{X}) = p)$
- β (length p) vector of coefficients

back

Three Point Structure Algorithm

(Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$2L(\mathbf{y}|\boldsymbol{\theta}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Three Point Structure Algorithm (Ho and Ané, 2014a)

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\mathbf{E} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{V})$$

Estimators:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$2L(\mathbf{y}|\boldsymbol{\theta}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

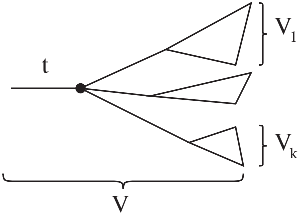
We need to compute:

- $\log|\mathbf{V}|$
- $\mathbf{Q} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$
- $p = \mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}$, $\hat{\mu}_Y = \mathbf{1}^T\mathbf{V}^{-1}\mathbf{Y}/p$ and $\hat{\boldsymbol{\mu}}_X^T = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{1}/p$

Three Point Structure Algorithm (Ho and Ané, 2014a)

Three Points Structure:

$$\mathbf{V} = t\mathbf{1}^T\mathbf{1} + \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_k \end{pmatrix}$$



Ho and Ané (2014a)

Three Point Structure Algorithm (Ho and Ané, 2014a)

Initialization : tree with a single tip of length t

- $\log |\mathbf{V}| = t$
- $\mathbf{Q} = \mathbf{x}^T y / t$
- $p = 1/t$, $\hat{\mu}_Y = y$ and $\hat{\boldsymbol{\mu}}_X^T = \mathbf{x}^T$

Three Point Structure Algorithm (Ho and Ané, 2014a)

Propagation:
Woodbury and Sylvester formulas for $\mathbf{M} = \mathbf{A} + \mathbf{UCV}$

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}$$

$$|\mathbf{M}| = |\mathbf{A}|\,|\mathbf{C}|\,\left|\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right|$$

Three Point Structure Algorithm (Ho and Ané, 2014a)

Propagation:

Woodbury and Sylvester formulas for $\mathbf{M} = \mathbf{A} + \mathbf{UCV}$

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1}$$

$$|\mathbf{M}| = |\mathbf{A}|\,|\mathbf{C}|\,\left|\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right|$$

Applied on $\mathbf{V} = \mathbf{A} + t\mathbf{1}^T\mathbf{1}$, with $\mathbf{A} = \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_k \end{pmatrix}$:

$$\mathbf{V}^{-1} = \mathbf{A}^{-1} - \frac{t}{1 + tp_A}\mathbf{A}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{A}^{-1}, \qquad p_A = \mathbf{1}^T\mathbf{A}^{-1}\mathbf{1} = \sum_{s=1}^{k} p_s$$

Three Point Structure Algorithm (Ho and Ané, 2014a)

Propagation:

Woodbury and Sylvester formulas for $\mathbf{M} = \mathbf{A} + \mathbf{UCV}$

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}$$

$$|\mathbf{M}| = |\mathbf{A}|\,|\mathbf{C}|\,\left|\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right|$$

Gives:

- $\log|\mathbf{V}| = \sum_{s=1}^{k} \log|\mathbf{V}_s| + \log(1 + tp_A)$
- $\mathbf{Q} = \sum_{s=1}^{k} \mathbf{Q}_s - \frac{tp_A^2}{1+tp_A}\hat{\boldsymbol{\mu}}_X^T \hat{\mu}_Y$
- $p = \frac{p_A}{1+tp_A}$, $\hat{\mu}_Y = \sum_{s=1}^{k} \frac{p_s}{p_A}\hat{\mu}_{Y,s}$ and $\hat{\boldsymbol{\mu}}_X^T = \sum_{s=1}^{k} \frac{p_s}{p_A}\hat{\boldsymbol{\mu}}_{X,s}$

Three Point Structure Algorithm

(Ho and Ané, 2014a)

Propagation:

Woodbury and Sylvester formulas for $\mathbf{M} = \mathbf{A} + \mathbf{UCV}$

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}$$

$$|\mathbf{M}| = |\mathbf{A}|\,|\mathbf{C}|\,\left|\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right|$$

Gives:

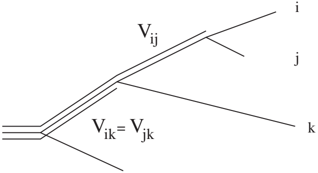
- $\log|\mathbf{V}| = \sum_{s=1}^{k} \log|\mathbf{V}_s| + \log(1 + tp_A)$
- $\mathbf{Q} = \sum_{s=1}^{k} \mathbf{Q}_s - \frac{tp_A^2}{1+tp_A}\hat{\boldsymbol{\mu}}_X^T\hat{\mu}_Y$
- $p = \frac{p_A}{1+tp_A}$, $\hat{\mu}_Y = \sum_{s=1}^{k} \frac{p_s}{p_A}\hat{\mu}_{Y,s}$ and $\hat{\boldsymbol{\mu}}_X^T = \sum_{s=1}^{k} \frac{p_s}{p_A}\hat{\boldsymbol{\mu}}_{X,s}$

Finalization : At the root of the tree.

Three Point Structure (Ho and Ané, 2014a)

Three Points Structure: A matrix $\mathbf{V}$ has the 3-Point Structure if:

- it is symmetric non-negative
- for any i, j, k, the two smallest of V_{ij}, V_{ik} and V_{jk} are equal.



Ho and Ané (2014a)

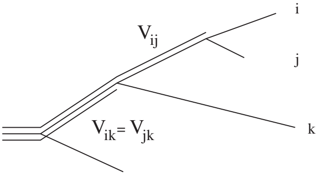
back

Three Point Structure

(Ho and Ané, 2014a)

Three Points Structure: A matrix **V** has the 3-Point Structure if:

- it is symmetric non-negative
- for any i, j, k, the two smallest of V_{ij}, V_{ik} and V_{jk} are equal.

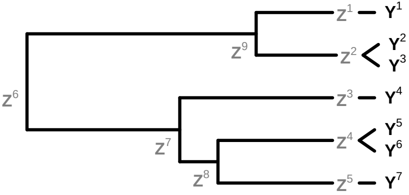


Ho and Ané (2014a)

Theorem: **V** is 3-point structured if and only if it is the covariance matrix of a random variable at the tips of some rooted tree under a BM model.

back

BM on a Non-Ultrametric Tree

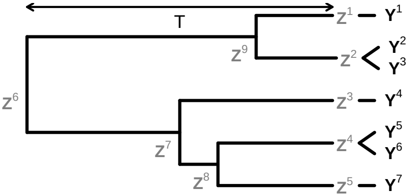


$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, s^2\right) \qquad \text{obs}$$

BM on a Non-Ultrametric Tree



$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

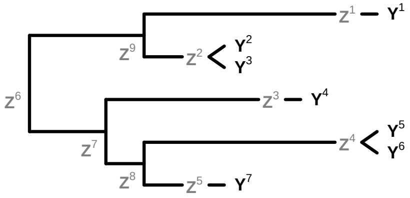
$$\mathbf{Z}^j \;\Big|\; \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \;\Big|\; \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, s^2\right) \qquad \text{obs}$$

Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

BM on a Non-Ultrametric Tree



$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

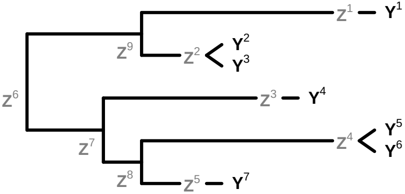
$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, s^2\right) \qquad \text{obs}$$

Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

Non-Ultrametric tree: How to define tip variances ?

BM on a Non-Ultrametric Tree



$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \,\middle|\, \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \,\middle|\, \mathbf{Z}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, s^2\right) \qquad \text{obs}$$

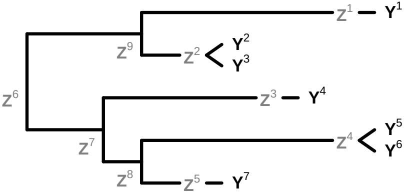
Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

Non-Ultrametric tree: How to define tip variances ?

$$h^2 = \frac{\sigma^2 \bar{T}}{\sigma^2 \bar{T} + s^2} \quad \text{(Mitov and Stadler, 2018)}$$

BM on a Non-Ultrametric Tree



$$\mathbf{Z}^r = \boldsymbol{\mu} \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(j)},\ \sigma^2 t_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, s^2\right) \qquad \text{obs}$$

Ultrametric tree: Tip variance is constant equal to $\sigma^2 T$

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2}$$

Non-Ultrametric tree: How to define tip variances ?

$$h^2 = \frac{\sigma^2 \bar{T}}{\sigma^2 \bar{T} + s^2} \qquad \text{(Mitov and Stadler, 2018)}$$

$$h^2 = \frac{\sigma^2 T_{max}}{\sigma^2 T_{max} + s^2} \qquad \text{(Leventhal and Bonhoeffer, 2016)}$$

BM on a Non-Ultrametric Tree

Assume a non ultrametric tree.

BM on a Non-Ultrametric Tree

Assume a non ultrametric tree.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with}\, \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$$

BM on a Non-Ultrametric Tree

Assume a non ultrametric tree.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with} \, \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$$

Not a unique scale anymore.

BM on a Non-Ultrametric Tree

Assume a non ultrametric tree.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with}\, \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$$

Not a unique scale anymore.

Leventhal and Bonhoeffer (2016) write: $\sigma^2 T_{max}\mathbf{V}/T_{max} + s^2\mathbf{I}$.
The phylogenetic heritability is then: $h^2 = \frac{\sigma^2 T_{max}}{\sigma^2 T_{max} + s^2}$

BM on a Non-Ultrametric Tree

Assume a non ultrametric tree.

BM PMM total variance is:

$$\sigma^2\mathbf{V} + s^2\mathbf{I} \quad \text{with}\,\mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$$

Not a unique scale anymore.

Leventhal and Bonhoeffer (2016) write: $\sigma^2 T_{max}\mathbf{V}/T_{max} + s^2\mathbf{I}$.
The phylogenetic heritability is then: $h^2 = \frac{\sigma^2 T_{max}}{\sigma^2 T_{max} + s^2}$

Mitov and Stadler (2018) use $\bar{T}$ instead.
The phylogenetic heritability is then: $h^2 = \frac{\sigma^2 \bar{T}}{\sigma^2 \bar{T} + s^2}$

Pagel's Lambda on a Non-Ultrametric Tree

Assume a non ultrametric tree.

Pagel's Lambda on a Non-Ultrametric Tree

Assume a non ultrametric tree.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda)\mathbf{V}^*\right]$$

with $\mathbf{V}^* = \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$.

Pagel's Lambda on a Non-Ultrametric Tree

Assume a non ultrametric tree.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda)\mathbf{V}^*\right]$$

with $\mathbf{V}^* = \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$.

Can λ be used as a measure of heritability in this case ?

Pagel's Lambda on a Non-Ultrametric Tree

Assume a non ultrametric tree.

Pagel's lambda variance is:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1 - \lambda) \mathbf{V}^* \right]$$

with $\mathbf{V}^* = \mathrm{Diag}(\mathbf{V}) = \mathrm{Diag}(T_1, \cdots, T_n)$.

Can λ be used as a measure of heritability in this case ?

BM PMM with scaled variance is:

$$\sigma^2 \mathbf{V} + s_s^2 \mathbf{V}^*$$

Error accumulate in time (Leventhal and Bonhoeffer, 2016)
a tip further from the root has a larger error.

Pagel's Lambda on a Non-Ultrametric Tree

Pagel's lambda variance:

$$\sigma_{\lambda}^{2}\mathbf{V}(\lambda) = \sigma_{\lambda}^{2}\left[\lambda\mathbf{V} + (1-\lambda)\mathbf{V}^{*}\right]$$

BM PMM with scaled variance:

$$\sigma^{2}\mathbf{V} + s_{s}^{2}\mathbf{V}^{*} = (\sigma^{2} + s_{s}^{2})\left[\frac{\sigma^{2}}{\sigma^{2}+s_{s}^{2}}\mathbf{V} + (1 - \frac{\sigma^{2}}{\sigma^{2}+s_{s}^{2}})\mathbf{V}^{*}\right]$$

Pagel's Lambda on a Non-Ultrametric Tree

Pagel's lambda variance:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda)\mathbf{V}^*\right]$$

BM PMM with scaled variance:

$$\sigma^2 \mathbf{V} + s_s^2 \mathbf{V}^* = (\sigma^2 + s_s^2) \left[\frac{\sigma^2}{\sigma^2 + s_s^2}\mathbf{V} + (1 - \frac{\sigma^2}{\sigma^2 + s_s^2})\mathbf{V}^*\right]$$

The two models are equivalent with:

$$\lambda = \frac{\sigma^2}{\sigma^2 + s_s^2} \qquad \sigma_\lambda^2 = \sigma^2 + s_s^2$$

Pagel's Lambda on a Non-Ultrametric Tree

Pagel's lambda variance:

$$\sigma_\lambda^2 \mathbf{V}(\lambda) = \sigma_\lambda^2 \left[\lambda \mathbf{V} + (1-\lambda)\mathbf{V}^*\right]$$

BM PMM with scaled variance:

$$\sigma^2 \mathbf{V} + s_s^2 \mathbf{V}^* = (\sigma^2 + s_s^2)\left[\frac{\sigma^2}{\sigma^2 + s_s^2}\mathbf{V} + (1 - \frac{\sigma^2}{\sigma^2 + s_s^2})\mathbf{V}^*\right]$$

The two models are equivalent with:

$$\lambda = \frac{\sigma^2}{\sigma^2 + s_s^2} \qquad \sigma_\lambda^2 = \sigma^2 + s_s^2$$

λ is the heritability if the independent errors are scaled.

BM Heritability

Ultrametric Tree

$$h^2 = \frac{\sigma^2 T}{\sigma^2 T + s^2} = \lambda$$

Non Ultrametric Tree

$$h^2 = \frac{\sigma^2 T_{max}}{\sigma^2 T_{max} + s^2}$$

$$h^2 = \frac{\sigma^2 \bar{T}}{\sigma^2 \bar{T} + s^2}$$

$$h^2 = \frac{\sigma^2}{\sigma^2 + s_s^2} = \lambda$$

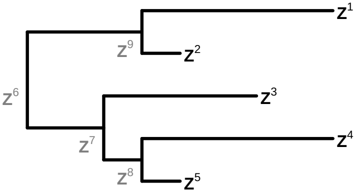
back

General Model

BM, OU: Instance of a general Gaussian propagation model.

General Model

BM, OU: Instance of a general Gaussian propagation model.



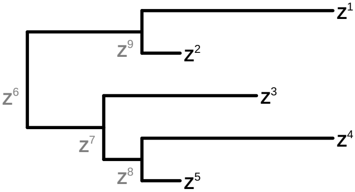
$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j\mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j\mathbf{R}$.

General Model

BM, OU: Instance of a general Gaussian propagation model.



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

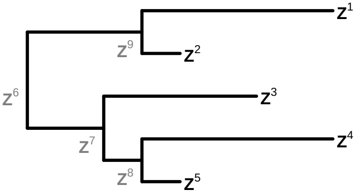
$$\mathbf{Z}^j \;\Big|\; \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \mathbf{\Sigma}_j\right) \qquad \text{nodes}$$

BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\mathbf{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\beta_j$, $\mathbf{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

General Model

BM, OU: Instance of a general Gaussian propagation model.



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

$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

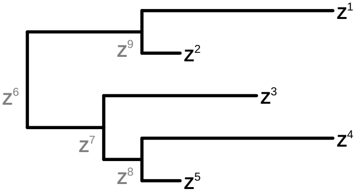
BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\boldsymbol{\beta}_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Drift, shifts, Integrated OU...

General Model

BM, OU: Instance of a general Gaussian propagation model.



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

$$\mathbf{Z}^j \;\Big|\; \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

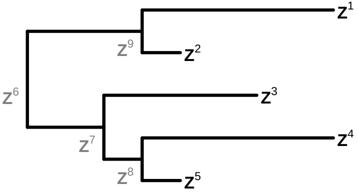
BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\boldsymbol{\beta}_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Drift, shifts, Integrated OU...

Easy computations (Kalman filter) Pruning

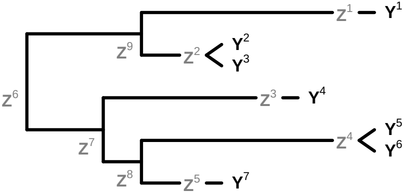
Intra-species variations Model



$$\mathbf{Z}^r \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Gamma}\right) \qquad \text{root}$$

$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

Intra-species variations Model

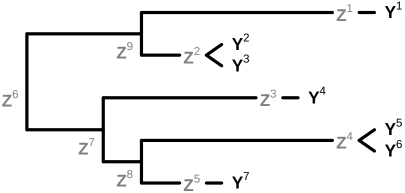


$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \qquad \text{obs}$$

Intra-species variations Model



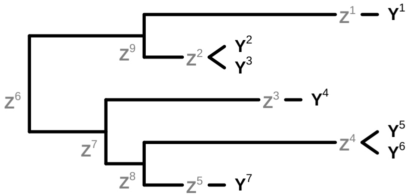
$$\mathbf{Z}^r \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Gamma}\right) \qquad \text{root}$$

$$\mathbf{Z}^j \;\middle|\; \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \;\middle|\; \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \qquad \text{obs}$$

Measurement Errors, Non phylogenetic variations, intra-species variations, ...

Intra-species variations Model



$$\mathbf{Z}^r \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Gamma}\right) \qquad \text{root}$$

$$\mathbf{Z}^j \;\Big|\; \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

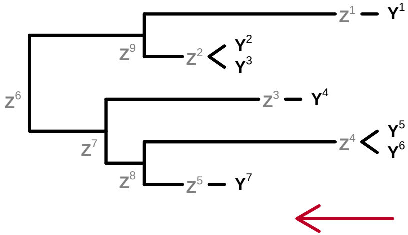
$$\mathbf{Y}^i \;\Big|\; \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \qquad \text{obs}$$

Measurement Errors, Non phylogenetic variations, intra-species variations, ...

"Heritability": $\mathbf{h}^2 = \dfrac{V(\mathbf{Z}_{tips})}{V(\mathbf{Y})}$

back

General Gaussian Model: Likelihood



$$\mathbf{Z}^r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

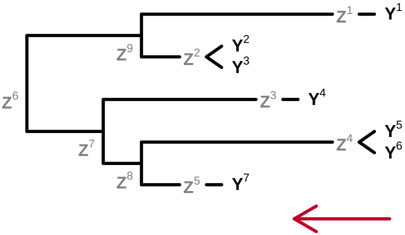
$$\mathbf{Z}^j \mid \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

$$\mathbf{Y}^i \mid \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \quad \text{obs}$$

Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

$\hookrightarrow O(N)$.

General Gaussian Model: Likelihood



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

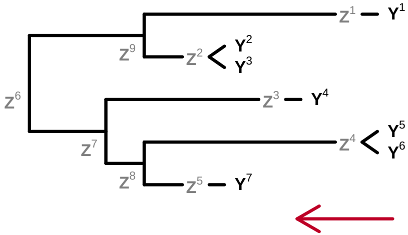
$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \,\Big|\, \mathbf{Z}^{\mathrm{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\mathrm{pa}(i)}, \mathbf{S}_i\right) \qquad \text{obs}$$

Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

- $\hookrightarrow O(N)$.
- $\hookrightarrow$ “Pruning”, “Gaussian elimination”, “Phylogenetic Kalman filter”, ...

General Gaussian Model: Likelihood



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

$$\mathbf{Z}^j \,\Big|\, \mathbf{Z}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{Z}^{\text{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \qquad \text{nodes}$$

$$\mathbf{Y}^i \,\Big|\, \mathbf{Z}^{\text{pa}(i)} \sim \mathcal{N}\left(\mathbf{Z}^{\text{pa}(i)}, \mathbf{S}_i\right) \qquad \text{obs}$$

Likelihood: $\log p(\mathbf{Y})$ in one post-order traversal.

$\hookrightarrow$ $O(N)$.

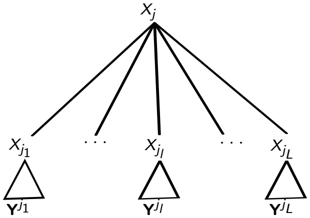
$\hookrightarrow$ “Pruning”, “Gaussian elimination”, “Phylogenetic Kalman filter”, ...

Difficulty: Numerical robustness.

Likelihood Computation

Likelihood function:

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



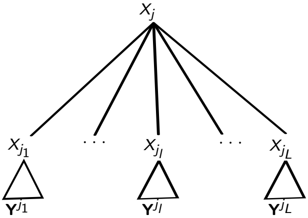
Likelihood Computation

Likelihood function:

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

Initialization: for tips

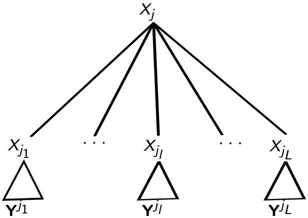
$$f_{Y_i|Y_i}(Y_i; a) = \Phi_{Y_i,0}(a)$$



Likelihood Computation

Likelihood function:

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



Initialization: for tips

$$f_{Y_i|Y_i}(Y_i; a) = \Phi_{Y_i,0}(a)$$

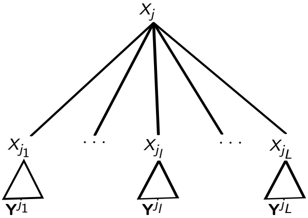
Propagation:

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

Likelihood Computation

Likelihood function:

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



Initialization: for tips

$$f_{Y_i|Y_i}(Y_i; a) = \Phi_{Y_i,0}(a)$$

Propagation:

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

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

back