

# Bioinformatics — Lecture 5

## Phylogenetic comparative methods (quantitative trait modelling)

Krzysztof Bartoszek

Linköping University  
*krzysztof.bartoszek@liu.se*

26 XI 2024 (R35)

# Today

Introduction

Mathematical models

Discrete traits

Continuous traits

Inference

Discrete traits

Continuous traits

Interactions

Software

## Additional reading

- G C. Gardiner. Stochastic Methods, 2009, Springer.
- E L. C. Evans. An Introduction to Stochastic Differential Equations, 2013, AMS.
- IY S. M. Iacus, N. Yoshida. Simulation and Inference for Stochastic Processes with YUIMA, 2018, Springer.
- K F. Klebaner. Introduction to Stochastic Calculus with Applications, 2005, Imperial College Press.
- P E. Paradis. Analysis of Phylogenetics and Evolution with R, 2006, Springer.

# Input and output

INPUT: A phylogeny relating a group of species

Trait measurements for the species

e.g. body size, eye colour,

Environmental data

OUTPUT: a model describing the evolution of the trait(s) taking into account the (varying) environments

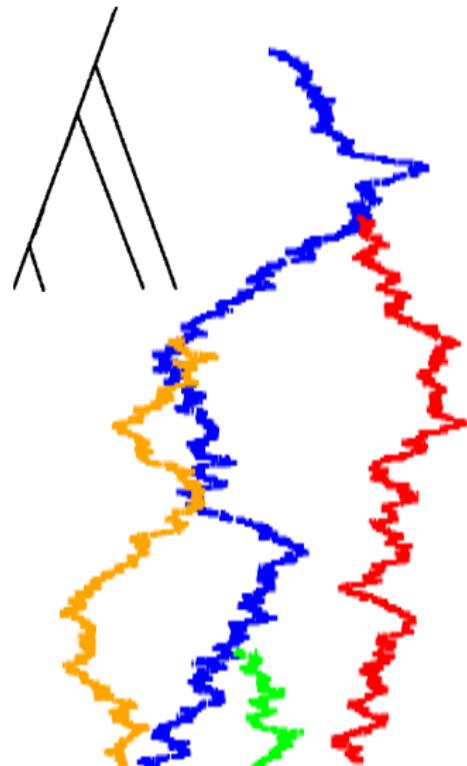
ULTIMATE GOAL: genotype–phenotype map

## Basic setup

Species evolve independently after speciation  
Trait: “average” for species

Introduction:

*Euplectes* spp. (subfam. Ploceinae = weaverbirds)



Graphic courtesy of Maria Prager

# Typical questions

"The *Primary optimum* for a *niche* (given environment) is the average optimum that would be reached for large number of species that were allowed to evolve independently in this niche."

T.F. Hansen, 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51, 1341-1351

- ▶ Does a trait influence the primary optimum of another trait?
- ▶ Does a given environmental factor influence the primary optimum of a trait?
- ▶ Do two (or more) traits evolve in a correlated/dependent fashion
- ▶ Is there a trade-off between traits?
- ▶ Is a trait adapting to another trait or environmental factor?
- ▶ Is there an allometric relationship between traits?
- ▶ What is magnitude of *phylogenetic inertia* (ancestral effects)?

# Allometry



"Allometry is the study of the relationship of body size to shape, anatomy, physiology and finally behaviour"

Relationship between traits:

$$y = kx^a \quad \log y = a \log x + \log k$$

"Two objects of different size, but common shape, will have their dimensions in the same ratio."

*Isometric scaling* proportional relationships are preserved  
*square–cube law* length increase  $r$  times, surface area  $r^2$ , volume  $r^3$   
mass~volume, surfaces~ $k \cdot \text{mass}^{2/3}$ , lengths~ $k \cdot \text{mass}^{1/3}$

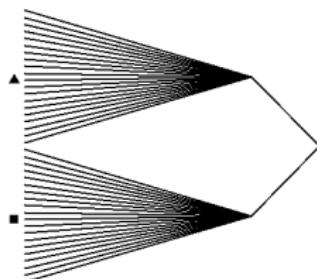
*Allometric scaling* change that deviates from isometry.

e.g. muskellunge fish (freshwater North America)  
weight =  $k \cdot \text{length}^{3.325}$  *positive allometry*

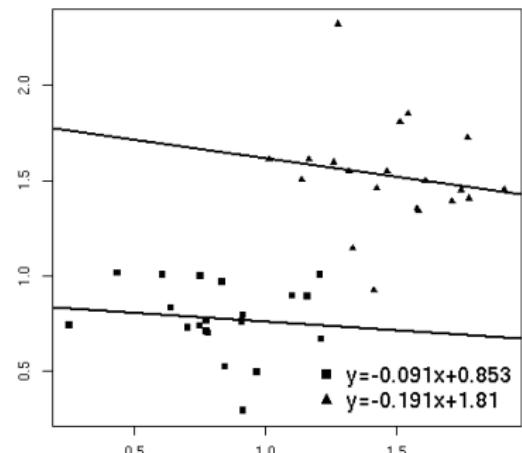
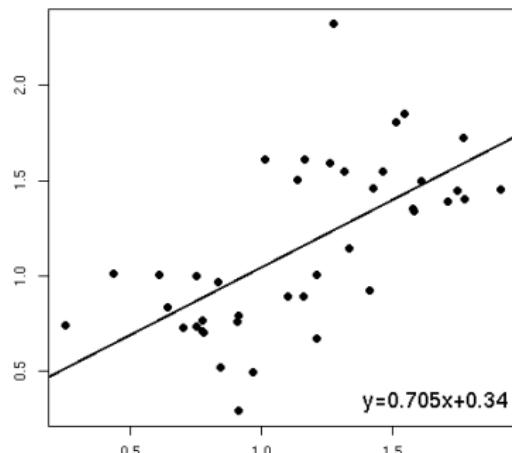
Plot on log scale reveals allometric scaling,  $a$ , factor

# Paradox?

See also J. Felsenstein, 1985. Phylogenies and the comparative method. Am. Nat. 125(1) 1–15

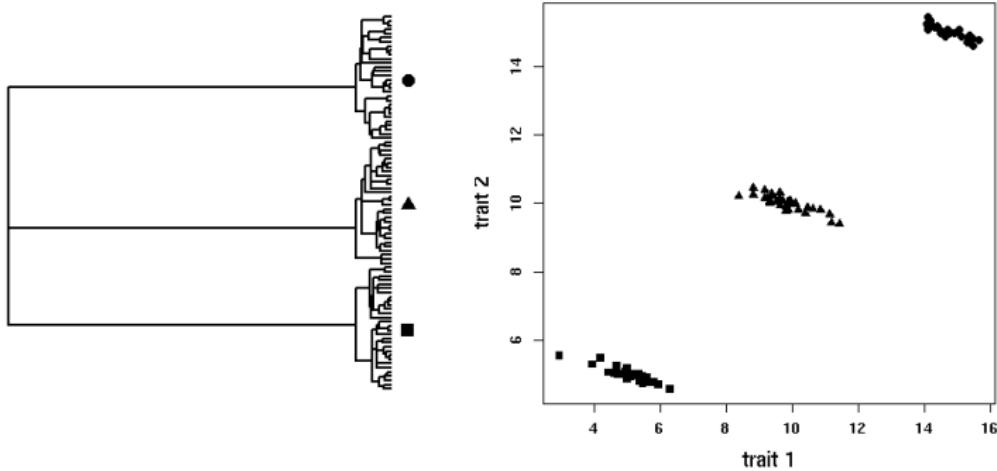


See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide8clustphy1.R



# Simpson's paradox

$$\text{Cov} [Y_i, Y_j] (t)$$



See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide9Simspon.R

# Markov chain

Previous lecture: 4-state MC for nucleotide evolution on tree

Discrete trait on  $k$  levels

Examples:

Nucleotide/amino-acid at a position

Number of chromosomes

Limblessness (binary) in Squamata (lizards, snakes, worm lizards)

sand lizard has, grass snake, slowworm (legless lizard) do not

Pollination strategy: bees/hummingbirds/wind

Diet: grains/insects/plants

Lifestyle: solitary/pair/small group/large herd

# Association between categorical random variables

$(X, Y)$  pair of categorical random variables (cross-tabulated)

$X$ : r levels (row variable)     $Y$ : c levels (column variable)

$X$  and  $Y$  independent:

$$P(X = i, Y = j) = P(X = i)P(Y = j)$$

$$\chi^2 = n \sum_{i=1}^r \sum_{j=1}^c \frac{\left( \frac{n_{ij}}{n} - \frac{n_{i-}}{n} \frac{n_{-j}}{n} \right)^2}{\frac{n_{i-}}{n} \frac{n_{-j}}{n}}$$

$\chi^2$  is unnormalized,  $\chi^2$ -test, tells us if there is a dependency,

# Strength of association between categorical r.vs.

*Cramér's V*: normalized to [0, 1] version of  $\chi^2$ , symmetric  
`lsr::cramersV()`

$$V = \sqrt{(\min(r - 1, c - 1))^{-1} (\chi^2/n)}$$

*Goodman and Kruskal's  $\tau$* : normalized to [0, 1], asymmetric  
 GoodmanKruskal R package

$$\tau_r = \frac{\sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}}{n} \frac{n_{ij}}{n-j} - \sum_{i=1}^r \left( \frac{n_{i-}}{n} \right)^2}{1 - \sum_{i=1}^r \left( \frac{n_{i-}}{n} \right)^2}$$

$$\tau_c = \frac{\sum_{j=1}^c \sum_{i=1}^r \frac{n_{ij}}{n} \frac{n_{ij}}{n_{i-}} - \sum_{j=1}^c \left( \frac{n_{-j}}{n} \right)^2}{1 - \sum_{j=1}^c \left( \frac{n_{-j}}{n} \right)^2}$$

# Ordinary differential equations (ODE)

*Derivative:* defines a functions change

Define a function by its derivative :

$$f'(x) = g(x)$$

then

$$f(x) = \int g(x)dx + C$$

fixing  $f(x_0) = y_0$  gives us  $C$

$g(x)$  can depend on  $f$  !!!!

$$f'(x) = g(x, f(x)), \quad f(x_0) = y_0$$

ODE: equation where the unknown is the function  $f$

# ODE examples

$$f'(x) = c, \quad f(0) = 0, \quad f(x) = cx$$

$$f'(x) = \alpha f(x), \quad f(0) = c_0, \quad f(x) = c_0 e^{\alpha x}$$

$$f'(x) = -\alpha f(x), \quad f(0) = c_0, \quad f(x) = c_0 e^{-\alpha x}$$

$$f'(t) = -\alpha(f(t) - \theta), \quad f(0) = x_0, \quad f(t) = x_0 e^{-\alpha t} + (1 - e^{-\alpha t})\theta$$

# ODE system (multivariate ODE)

Derivatives are taken dimension by dimension.

Examples

$$f : \mathbb{R}^{k_1} \rightarrow \mathbb{R}^{k_2}, f(\vec{0}) = \vec{0}, f'(\vec{x}) = \vec{c}, \quad f(\vec{x}) = [\text{diag}(\vec{c})] \vec{x}$$

$$\begin{aligned} f : \mathbb{R} \rightarrow \mathbb{R}^k, f(0) &= \vec{x}_0, f'(t) = -\mathbf{A}(f(t) - \vec{\theta}), \\ f(t) &= e^{-\mathbf{A}t} \vec{x}_0 + (1 - e^{-\mathbf{A}t})\vec{\theta} \end{aligned}$$

What is  $e^{\mathbf{M}}$  ?

$$e^{\mathbf{M}} = \sum_{i=0}^{\infty} \frac{\mathbf{M}^i}{i!}$$

$M = \mathbf{P}\Lambda\mathbf{P}^{-1}$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$  (eigendecomposition, if exists)

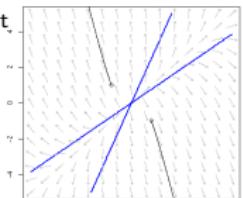
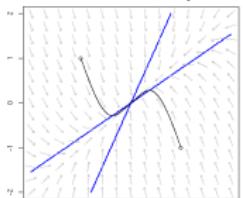
$$e^{\mathbf{M}} = \mathbf{P}e^{\Lambda}\mathbf{P}^{-1}$$

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide16PP.R

$$f'(t) = -\mathbf{A}(f(t)) \cdot 2D$$

 $\lambda_1, \lambda_2 < 0$ 

unstable fixed point

 $\mathbb{R} \ni \lambda_1 \neq \lambda_2 \in \mathbb{R}$  $\lambda_1, \lambda_2 > 0$  $\infty$ stable fixed point $\lambda_1 \lambda_2 < 0$ 

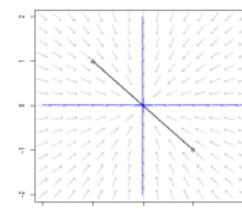
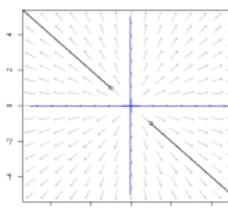
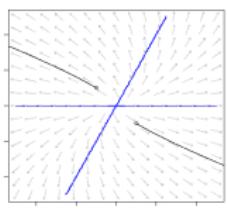
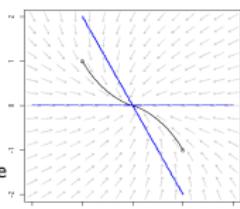
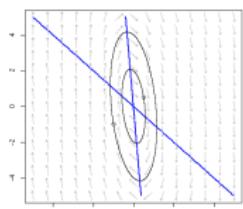
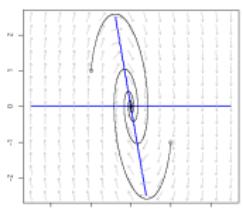
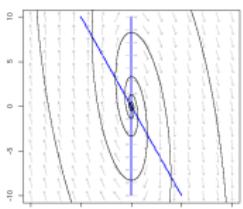
unstable saddle point

1 eigenvector degenerate

2 eigenvectors, star

 $\lambda_1 = \lambda_2 \in \mathbb{R}$  $\lambda < 0$ 

unstable

 $\lambda > 0$  $\infty$ stable $\lambda > 0 \infty$ stable $\lambda < 0$  unstable $\lambda_1, \lambda_2 \in \mathbb{C} \quad \text{Re}(\lambda) < 0$   
unstable spiral $\text{Re}(\lambda) = 0$   
stable centre $\text{Re}(\lambda) > 0 \infty$ stable spiral

# ODE plotting

If ODE solvable, then direct graph

Otherwise:

1. choose small time step  $\Delta$
2.  $f(t_0) = x_0$  initial condition and set  $t = t_0$
3.  $f(t + \Delta) = f(t) + \Delta g(t)$
4.  $t = t + \Delta$
5. Repeat steps 3 and 4 until  $t$  reaches  $t_{\max}$

multivariate ODE system in the exact same way

# Stochastic differential equations (SDE)

Stochastic process:  $X(t)$  “a collection of random variables that together form a random function of time”

Intuitive idea:                    ODE + random noise

But apart from special cases this does not work in this way

Special case: linear SDE (considered here)

Stochastic calculus

Itô formula

Applications:

Finance

Biology

## Brownian motion

*Brownian motion*,  $B(t)$  is a stochastic process with the following properties

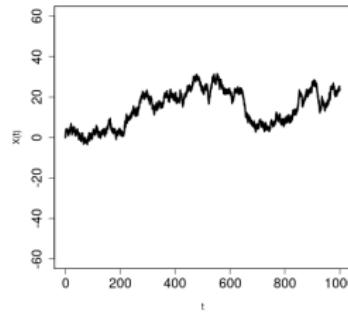
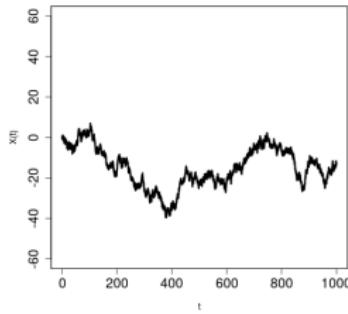
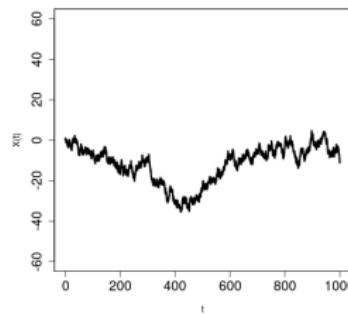
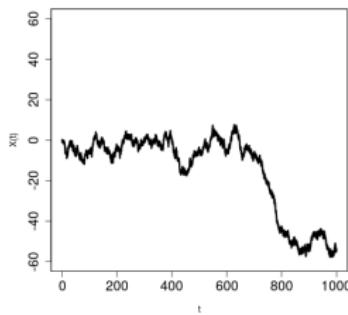
1. (Independence of increments)  $B(t) - B(s)$  for  $t > s$  is independent of the past, that is of  $B(u)$  for  $0 \leq u \leq s$ .
2. (Normal increments)  $B(t) - B(s)$  is normally distributed with mean 0 and variance  $t - s$
3. (Continuity of paths)  $B(t)$  is a continuous function of  $t \geq 0$

$$\begin{aligned} E[B(t)] &= B(0) \\ \text{Var}[B(t)] &= t \end{aligned}$$

# Brownian motion

$$X_0 = 0, \sigma = 1$$

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide20BM.R simulations by yuima R package



# Diffusion type SDE

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t) \quad X(0) = X_0$$

"ODE"                          "noise"

What is this? It is shorthand for:                          "white noise"

$$X(t) = \int_0^t \mu(s, X(s))ds + \int_0^t \sigma(s, X(s))dB(s)$$

The solution to an SDE is a stochastic process that satisfies the above integral equation.

$\int_0^T X(s)dB(s)$  is a "special" integral—Itô integral

$X(t)$  such that  $\int_0^T E[X(s)^2] ds < \infty$ , then  $E\left[\int_0^T X(s)dB(s)\right] = 0$

In some special cases we know what the distribution of  $X(t)$ .

# SDE examples

$$dX(t) = dB(t) \quad X(t) = X_0 + B(t), \quad X(t) \sim \mathcal{N}(X_0, t)$$

$$dX(t) = \sigma dB(t) \quad X(t) = X_0 + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0, \sigma^2 t)$$

$$dX(t) = \mu dt + \sigma dB(t) \quad X(t) = X_0 + \mu t + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0 + \mu t, \sigma^2 t)$$

$$dX(t) = \mu dt + \sigma dB(t) \quad X(t) = X_0 + \mu t + \sigma B(t), \quad X(t) \sim \mathcal{N}(X_0 + \mu t, \sigma^2 t)$$

linear type SDE:

$$dX(t) = (c(t) + d(t)X(t))dt + \sigma(t)dB(t) \quad \text{denote } h(t) = \int_0^t d(s)ds$$

$$X(t) = e^{h(t)}X_0 + e^{h(t)} \int_0^t e^{-h(s)}c(s)ds + e^{h(t)} \int_0^t e^{-h(s)}\sigma(s)dB(s)$$

$$d(t) \equiv d \quad X(t) = e^{Dt}X_0 + \int_0^t e^{d(t-s)}c(s)ds + \int_0^t e^{d(t-s)}\sigma(s)dB(s)$$

# Ornstein–Uhlenbeck (OU) process

$$dX(t) = -\alpha(X(t) - \theta)dt + \sigma dB(t), \quad X(0) = X_0$$

$$E[X(t)] = e^{-\alpha t} X_0 + (1 - e^{-\alpha t})\theta \xrightarrow{\alpha > 0} \theta$$

$$\text{Var}[X(t)] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \xrightarrow{\alpha > 0} \frac{\sigma^2}{2\alpha}$$

$\alpha > 0$ : converges (weakly) to  $\mathcal{N}(\theta, \sigma^2/(2\alpha))$

*Half-life* time to lose half of ancestral

$$\exp(-\alpha t_{0.5}) = 0.5 \text{ then } t_{0.5} = \ln 2 / \alpha$$

T.F. Hansen, 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51, 1341-1351

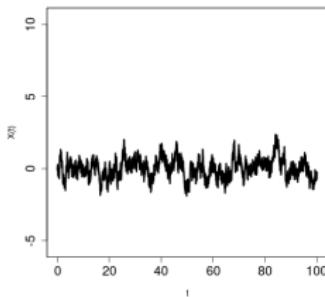
Interest rates model (Vašíček model)

# OU trajectories

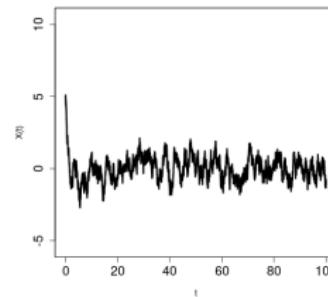
$$\theta = 0, \alpha = 1, \sigma = 1$$

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide24OU.R simulations by yuima R package

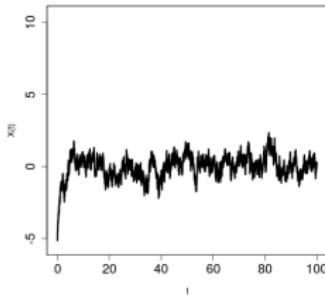
$$X_0 = 0$$



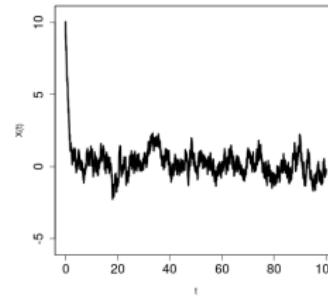
$$X_0 = 5$$



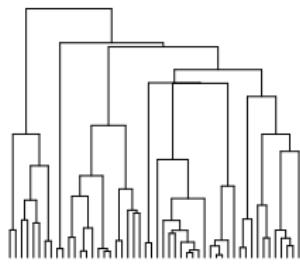
$$X_0 = -5$$



$$X_0 = 10$$



# Phylogenetic BM and OU



$n = 50$



BM



OU

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide25.R  
simulations by TreeSim and mvSLOUCH R package

# Multivariate SDEs

Same as ODEs:

“derivative” /Itô integral taken dimension by dimension.

Diffusion type SDE

$$d\vec{X}(t) = \vec{\mu}(t, \vec{X}(t))dt + \boldsymbol{\Sigma}(t, \vec{X}(t))d\vec{B}(t) \quad \vec{X}(0) = \vec{X}_0$$

$$\vec{X}(t) = \int_0^t \vec{\mu}(s, \vec{X}(s))ds + \int_0^t \boldsymbol{\Sigma}(s, \vec{X}(s))d\vec{B}(s)$$

## Multivariate SDE examples

$$d\vec{X}(t) = d\vec{B}(t) \quad \vec{X}(t) = \vec{X}_0 + \vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0, t)$$

$$d\vec{X}(t) = \boldsymbol{\Sigma} d\vec{B}(t) \quad \vec{X}(t) = \vec{X}_0 + \boldsymbol{\Sigma} \vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0, \boldsymbol{\Sigma} \boldsymbol{\Sigma}^T t)$$

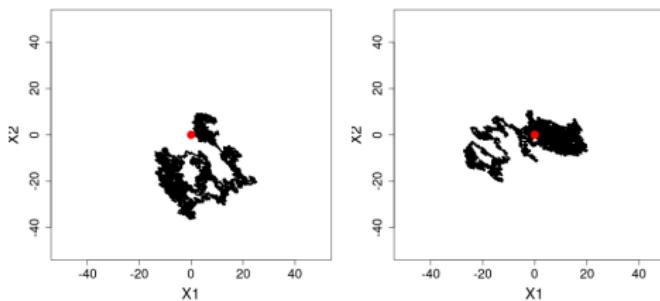
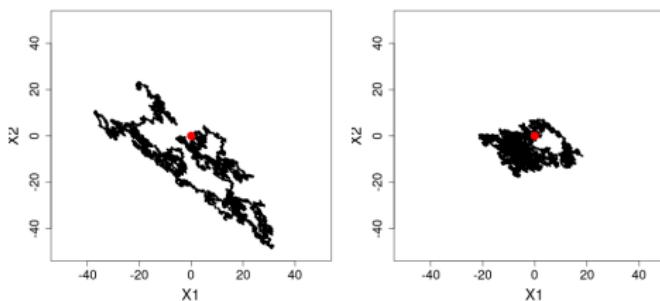
$$\begin{aligned} d\vec{X}(t) &= \vec{\mu} dt + \boldsymbol{\Sigma} d\vec{B}(t) \\ \vec{X}(t) &= \vec{X}_0 + \vec{\mu} t + \boldsymbol{\Sigma} \vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0 + \vec{\mu} t, \boldsymbol{\Sigma} \boldsymbol{\Sigma}^T t) \end{aligned}$$

$$\begin{aligned} d\vec{X}(t) &= \vec{\mu} dt + \boldsymbol{\Sigma} d\vec{B}(t) \\ \vec{X}(t) &= \vec{X}_0 + \mu t + \sigma \vec{B}(t), \quad \vec{X}(t) \sim \mathcal{N}(\vec{X}_0 + \vec{\mu} t, \boldsymbol{\Sigma} \boldsymbol{\Sigma}^T t) \end{aligned}$$

## 2D Brownian motion

$$\vec{X}_0 = (0, 0)^T, \Sigma \Sigma^T = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \Sigma = (\text{chol}(\Sigma \Sigma^T))^T$$

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide28mvBM.R, simulations by yuima R package



# Multivariate linear SDEs

$$d\vec{X}(t) = (\vec{c}(t) + \mathbf{D}(t)\vec{X}(t))dt + \boldsymbol{\Sigma}(t)d\vec{B}(t)$$

$$\vec{X}(t) = \Phi(t)\vec{X}_0 + \Phi(t) \int_0^t \Phi^{-1}(s)\vec{c}(s)ds + \Phi(t) \int_0^t \Phi^{-1}(s)\boldsymbol{\Sigma}(s)d\vec{B}(s)$$

where  $\Phi(t)$  is the solution (fundamental matrix) of

$$\frac{d\Phi(t)}{dt} = \mathbf{D}(t)\Phi(t) \quad \Phi(0) = \mathbf{I}$$

If  $\mathbf{D}(t) \equiv \mathbf{D}$

$$\vec{X}(t) = e^{\mathbf{D}t}\vec{X}_0 + \int_0^t e^{\mathbf{D}(t-s)}\vec{c}(s)ds + \int_0^t e^{\mathbf{D}(t-s)}\boldsymbol{\Sigma}(s)d\vec{B}(s))$$

# Multivariate Ornstein–Uhlenbeck process

$$d\vec{X}(t) = -\mathbf{A}(\vec{X}(t) - \vec{\theta})dt + \boldsymbol{\Sigma}dB(t), \quad \vec{X}(0) = \vec{X}_0$$

$$E[\vec{X}(t)] = e^{-\mathbf{A}t}\vec{X}_0 + (1 - e^{-\mathbf{A}t})\vec{\theta}$$

$$\text{Var}[\vec{X}(t)] = \int_0^t e^{\mathbf{A}s} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^T e^{\mathbf{A}^T s} ds$$

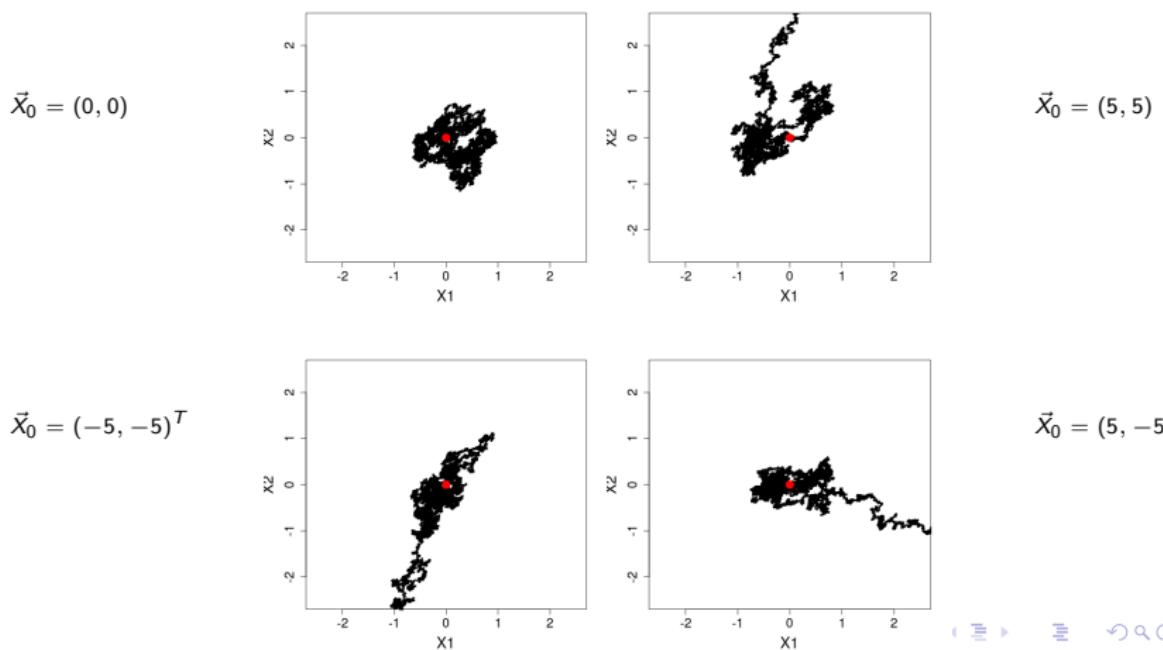
If all eigenvalues of  $\mathbf{A}$  have positive real part then, converges (weakly) to normal with mean  $\vec{\theta}$  and  $\text{Var}[\vec{X}(\infty)]$

*Half-lives* can be calculated in eigenvector space

## 2D OU trajectories

$$\text{OU: } \vec{\theta} = (0, 0)^T, \mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix}, \Sigma\Sigma^T = \begin{bmatrix} 0.4 & 0.15 \\ 0.15 & 0.4 \end{bmatrix}$$

See code: 732A51\_BioinformaticsHT2023\_Lecture05codeSlide31mvOU.R, simulations by yuima R package



# “Random mean” process

Special “singular” case:

$$\begin{aligned} d\vec{Y}(t) &= -\mathbf{D} \left( (\vec{Y})(t) - (\mathbf{B}\vec{X}(t) + \vec{\theta}) \right) + \boldsymbol{\Sigma}_y dB_y(t) \\ d\vec{X}(t) &= \boldsymbol{\Sigma}_x dB_x(t) \end{aligned}$$

No stationary distribution,

$\mathbf{A} = \text{rbind}(\text{cbind}(\mathbf{D}, \mathbf{B}), \mathbf{0})$  has  $\mathbf{0}$  as eigenvalues.

But “residuals process”  $(\vec{Y})(t) - (\mathbf{B}\vec{X}(t) + \vec{\theta})$  stabilizes.

# Simulating SDEs

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t) \quad X(0) = X_0$$

Euler–Maruyama method:

1. choose small time step  $\Delta$

2.  $X(0) = X_0$  initial condition and set  $t = t_0$

If transition density known (e.g. linear—normal),

$X(t)|X_0 \sim f(\cdot, t, X_0)$

3.  $X(t + \Delta) = X(t) + \epsilon, \epsilon \sim f(\cdot, \Delta, X_t)$

Else

2'. Set  $B(0) = 0$

3'. Draw  $\epsilon \sim \mathcal{N}(0, \Delta)$

3''.  $X(t + \Delta) = \mu(\Delta, X(t))\Delta + \sigma(\Delta, X(t))\epsilon$

4.  $t = t + \Delta$

5. Repeat steps 3/(3',3'') and 4 until  $t$  reaches  $t_{\max}$

## (Felsenstein's) Pruning algorithm

$X(t) \in S$ , state space

$i$  inner node,

$L^i(s)$ : likelihood of subtree rooted at  $i$  with  $X$ (at node  $i$ ) =  $s$

$$L^i(s) = \prod_{o: \text{daughter of } i} \left[ \sum_{x \in S} P_{sx}(t_o) L^o(x) \right]$$

and for a leaf  $j$

$$L^j(s) = \begin{cases} 1 & \text{if } s \text{ is leaf } j' \text{'s state} \\ 0 & \text{otherwise} \end{cases}$$

# Estimating ancestral characters

Parismony: Fitch's algorithm (Lec. 4), minimal number of changes

Statistical justification?

Different transition rates between different states?

Change is not rare?

Variation in lineages?

ML: find ancestral state,  $s$ , maximizing  $L^{\text{root}}(s)$

Identifiability?

Model misspecification?

Variation in rates over time, lineages?

Bayesian: posterior distribution

[http://topicspageswiki.plos.org/wiki/Ancestral\\_reconstruction#Methods\\_and\\_algorithms](http://topicspageswiki.plos.org/wiki/Ancestral_reconstruction#Methods_and_algorithms)

# Brownian motion model

$$X(t)|X(0) \sim \mathcal{N}(X(0), \sigma^2 t)$$

Pairs of tip measurements ( $X_1, X_2$ ) are NOT independent

$$\text{Cov}[X_1, X_2] = \sigma^2 t_{12} \quad t_{12} \text{ shared path length}$$

PROOF

Law of total covariance:

$$\text{Cov}[Y, Z] = \mathbb{E}[\text{Cov}[Y, Z|U]] + \text{Cov}[\mathbb{E}[Y|U], \mathbb{E}[Z|U]]$$

$$\begin{aligned} \text{Cov}[X_1, X_2] &= \mathbb{E}[\text{Cov}[X_1, X_2|X_{\text{anc}_{12}}]] + \text{Cov}[\mathbb{E}[X_1|X_{\text{anc}_{12}}], \mathbb{E}[X_2|X_{\text{anc}_{12}}]] \\ &= 0 + \text{Cov}[X_{\text{anc}_{12}}, X_{\text{anc}_{12}}] = \text{Var}[X_{\text{anc}_{12}}] = \sigma^2 t_{12} \end{aligned}$$

$$\vec{X} = (X_1, \dots, X_n)^T \sim \mathcal{N}(\vec{1}X(0), \sigma^2 \mathbf{T})$$

$\mathbf{T}$ : matrix of shared branch lengths

# A regression perspective (BM)

$$\vec{X} = \vec{1}X_0 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{T})$$

GLS estimates:

$$\hat{X}_0 = \left( \vec{1}^T \mathbf{T}^{-1} \vec{1} \right)^{-1} \vec{1}^T \mathbf{T}^{-1} \vec{X}$$

$$\hat{\sigma}^2 = (\vec{X} - \vec{1}X_0) \mathbf{T}^{-1} (\vec{X} - \vec{1}X_0)^T$$

*Independent contrasts:* algorithm to calculate the GLS estimates

J. Felsenstein, 1985. Phylogenies and the comparative method. Am. Nat. 125(1) 1–15

# A regression perspective (General)

$\Theta$ : Model parameters

$$\vec{X} = \mathbf{D}(Tree, \Theta) \vec{\Theta}_{lin} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{V}(Tree, \Theta))$$

Find  $\Theta$

- (iterative) GLS methods
- maximum likelihood
- Bayesian
- hybrid methods

# Maximum likelihood (normal model)

$$\vec{X} \sim \mathcal{N}(\mu(Tree, \Theta), \mathbf{V}(Tree, \Theta))$$

$$(-1)\mathcal{L}(\vec{X} | Tree, \Theta) \propto |\mathbf{V}(Tree, \Theta)| + (\vec{X} - \mu(Tree, \Theta))^T \mathbf{V}(Tree, \Theta)^{-1} (\vec{X} - \mu(Tree, \Theta))$$

Minimize  $(-1)\mathcal{L}$  numerically

R's `optim()` works fine,

but explore `control$parscale` parameter

Obtaining  $\mathbf{V}(Tree, \Theta)$  is a bottleneck

Numerical accuracy?

# OU distribution of phylogenetic sample

$$X(t)|X(0) \sim \mathcal{N}(e^{-\alpha t} X(0) + (1 - e^{-\alpha t})\theta, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}))$$

Pairs of tip measurements ( $X_1, X_2$ ) are NOT independent

$$\text{Cov}[X_1, X_2] = \frac{\sigma^2}{2\alpha}(e^{-2\alpha t_{12}} - e^{-2\alpha t}) \quad t_{12} \text{ shared path length}$$

PROOF

Law of total covariance (same approach as BM)

Easy to generalize for piecewise linear  $\theta$  (notational issues)

# Missing values (normal model)

## Multitrait setting

Often measurements on some traits in some species are missing  
Removing the whole species wastes the observed traits.

Notation:  $\vec{X} = (\vec{X}_1^T, \dots, \vec{X}_n^T)^T$

Normal distribution framework:

- Remove entries of  $E[\vec{X}] = \vec{\mu}(Tree, \Theta)$  corresponding to missing values
- Remove rows and columns of  $\text{Var}[\vec{X}] = \mathbf{V}(Tree, \Theta)$  corresponding to missing values

# Measurement error (normal model)

The tip measurements are usually averages for a species

Hence natural intra-species variability present

Calculate for each species its variance

Add the variances to  $\mathbf{V}(\text{Tree}, \Theta)$ 's (block) diagonal

# Pruning algorithm

Calculating  $\mathbf{V}(\text{Tree}, \Theta)$  is time-consuming

Faster algorithm:

$X(t) \in \mathbb{R}^k$ , state space

$i$  inner node,

$L^i(s)$ : likelihood of subtree rooted at  $i$  with  $X$ (at node  $i$ ) =  $s$

$$L^i(s) = \prod_{o: \text{daughter of } i} \int_{\mathbb{R}^k} f_{s \rightarrow x}(t_o) L^o(x) dx$$

and for a leaf  $j$

$$L^j(s) = \begin{cases} 1 & \text{if } s \text{ is leaf } j's \text{ state} \\ 0 & \text{otherwise} \end{cases}$$

Provided one can calculate  $\int_{\mathbb{R}^k} f_{s \rightarrow x}(t_o) L^o(x) dx$ .

## Summarizing a model

Likelihood in itself does not say much

Plot support surface in some parameters' directions

Report parametric bootstrap confidence intervals

Calculate regression confidence intervals for  $\Theta_{lin}$  conditional on  $\Theta_{non-lin}$

$R^2 = RSS_{model} / RSS_{null\ model}$ , but what is null model?

If one can partition traits: predictors/responses:  
conditional density of responses on predictors

Half-lives (in eigenspace)

# Model selection

Competing models: e.g. BM versus OU

BM: neutral evolution

OU: adaptive evolution

OU:

different levels of  $\theta$  (environments)  
assumptions on **A**

Look at AIC, AIC<sub>c</sub>, BIC

# Estimating ancestral characters (BM, OU)

Difficult to estimate ancestral state if no fossil data

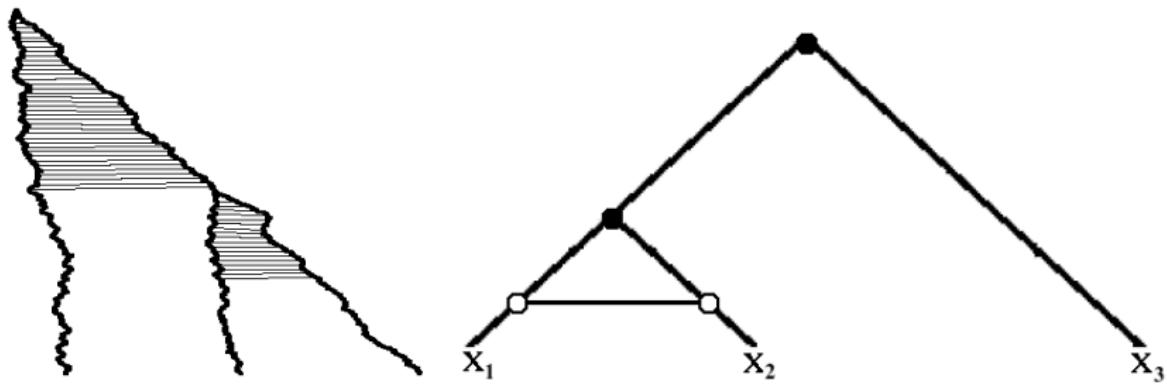
BM:  $\hat{X}_0$  is a linear combination of tips

$$\text{Var} [\hat{X}_0] \not\rightarrow 0, \text{ as BM's variance } \rightarrow \infty$$

OU: estimation is difficult or impossible

wide confidence intervals due to  $e^{-\alpha t} X_0$  in mean  
if  $\theta$  constant, then one cannot distinguish

# Migration and hybridizations



Migration

Hybridization

See code: 732A51\_BioinformaticsHT2023.Lecture05codeSlide47Migr.R

# R packages (a primer)

mvSLOUCH

PCMBase; PCMBaseCpp

PCMFit <https://github.com/venelin/PCMFit>

GLSME

pcmabc

slouch ouch

caper ape

geiger OUwie

mvMORPH bayou

surface rphylopars

PhylogeneticEM diversitree

RPANDA

# Quantitative trait locus

Unrelated to today's topic

BUT

name can be confusing

"A *quantitative trait locus* (QTL) is a section of DNA (the locus) which correlates with variation in a phenotype (the quantitative trait)."

[https://en.wikipedia.org/wiki/Quantitative\\_trait\\_locus](https://en.wikipedia.org/wiki/Quantitative_trait_locus)

# Questions?