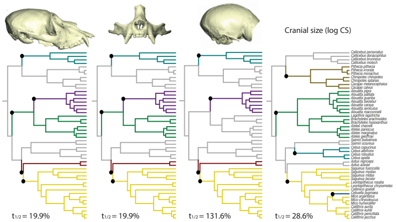
A Flexible Bayesian Framework to Study Viral Trait Evolution

Paul Bastide[1], Marc A. Suchard[2,3,4] and Philippe Lemey[1]

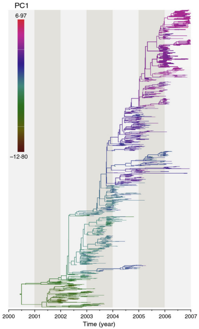
[1] Evolutionary and Computational Virology, Rega Institute, KU Leuven, Belgium
[2] Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, USA
[3] Department of Biomathematics, David Geffen School of Medicine at UCLA, USA
[4] Department of Human Genetics, David Geffen School of Medicine at UCLA, USA

21 January 2019

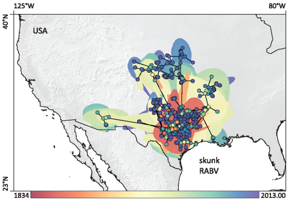
Phylogenetic Comparative Methods



Aristide et al. (2018)



Vrancken et al. (2015)



Dellicour et al. (2017)

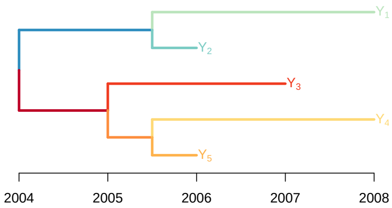
- Various time scales: Myr – decade.
- Various traits: morpho, geo, viral.

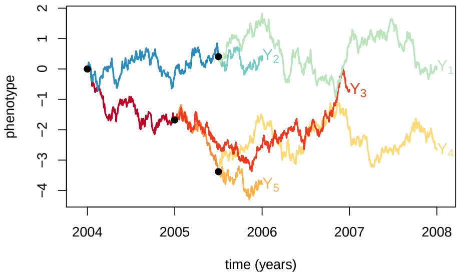
Question: Trait dynamics for an evolving organism ?

Outline

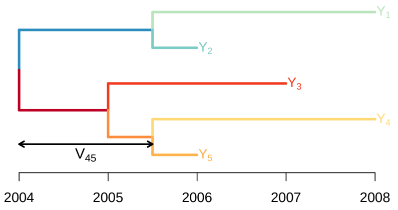
1. Models of Trait Evolution
2. Efficient Bayesian Inference
3. HIV Virulence Heritability Study

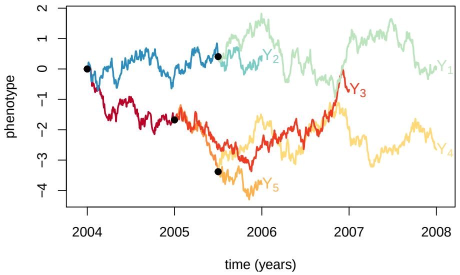
BM on a Tree


Y1
Y2
Y3
Y4
Y5
2004
2005
2006
2007
2008


phenotype
time (years)
Y1
Y2
Y3
Y4
Y5

BM on a Tree



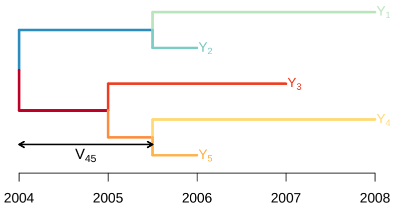


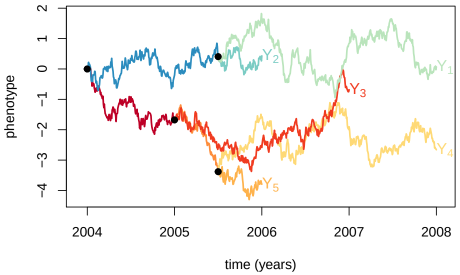
EDS: $\mathrm{d}\, X_t = \sigma \,\mathrm{d}\, B_t$

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

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

BM on a Tree

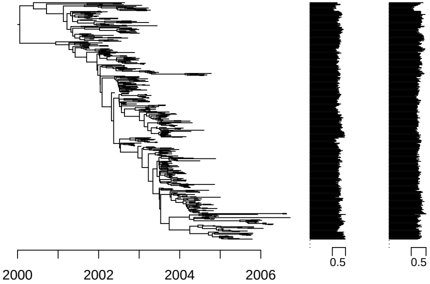




Distribution: Normal

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

Multivariate BM

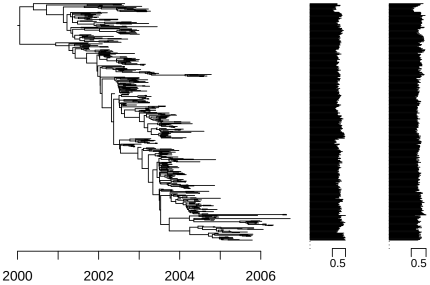


Data: Vectors of p traits

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

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

Multivariate BM



Data: Vectors of p traits

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

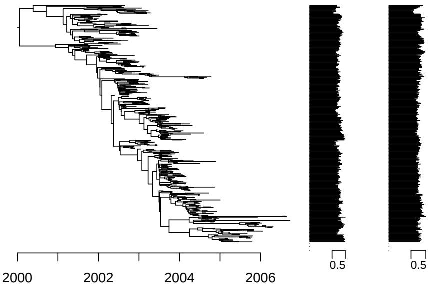
EDS: $\mathrm{d}\,\mathbf{X}_t = \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$ $\quad\mathbf{R} = \mathbf{\Sigma}^T\mathbf{\Sigma}$

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

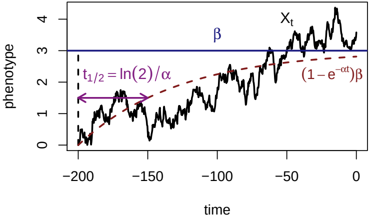
$$\mathbf{Y}_i^T = (Y_{i1}, \dots, Y_{ip})$$

Distribution: Matrix Normal

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

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

Ornstein-Uhlenbeck Modeling (Hansen, 1997)



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

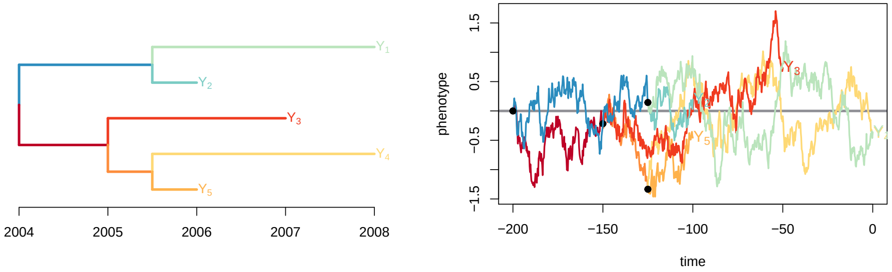
Deterministic part:

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

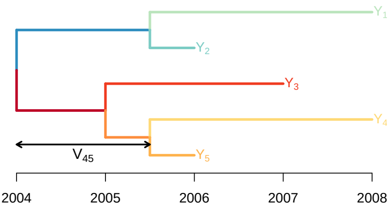
Stochastic part:

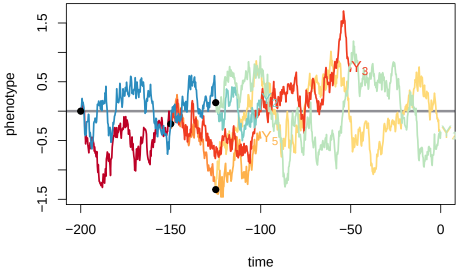
- X_t: trait value (actual optimum).
- $\sigma dB(t)$: Brownian fluctuations.

OU on a Tree


Y1
Y2
Y3
Y4
Y5
2004
2005
2006
2007
2008
phenotype
1.5
0.5
−0.5
−1.5
−200
−150
−100
−50
0
time

OU on a Tree





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

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

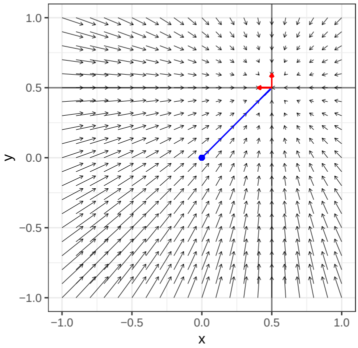
Multivariate OU Modeling

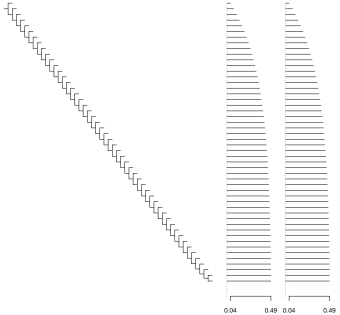
$$\mathrm{d}\,\mathbf{X}_t = \mathbf{A}[\beta - \mathbf{X}_t]\,\mathrm{d}\,t + \mathbf{\Sigma}\,\mathrm{d}\,\mathbf{B}_t$$

Multivariate OU Modeling

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

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

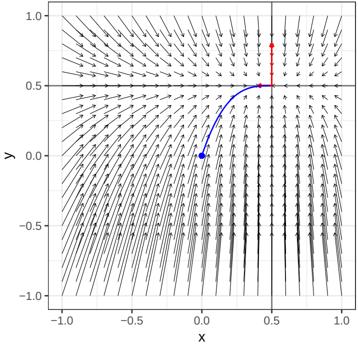


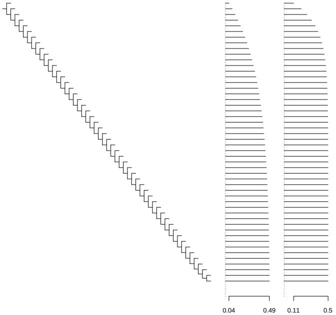


Multivariate OU Modeling

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

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

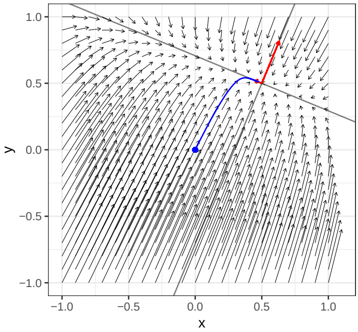


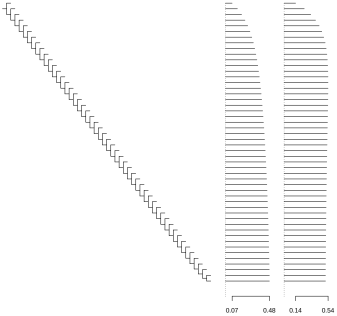


Multivariate OU Modeling

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

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

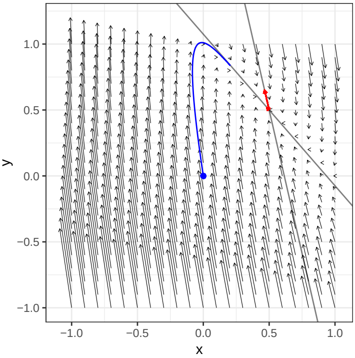


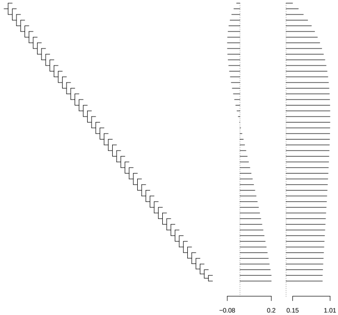


Multivariate OU Modeling

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

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

Multivariate OU

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

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

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

Multivariate OU

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

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

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

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

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

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

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

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

Distribution: Still Gaussian.

No nice Kronecker product.

Bayesian Phylogenetics

Goal:

$$p\left(\boldsymbol{\theta}, \mathcal{T}, \psi \mid \mathbf{Y}, \mathbf{S}\right)$$

Bayesian Phylogenetics

Goal:

$$p\left(\boldsymbol{\theta}, \mathcal{T}, \psi \mid \mathbf{Y}, \mathbf{S}\right) \propto p\left(\mathbf{Y}, \mathbf{S} \mid \boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}\right) p\left(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}\right)$$

Bayesian Phylogenetics

Goal:

$$\begin{aligned} p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi} \mid \mathbf{Y}, \mathbf{S}) &\propto p(\mathbf{Y}, \mathbf{S} \mid \boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi})\, p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}) \\ &\propto p(\mathbf{Y} \mid \boldsymbol{\theta}, \mathcal{T})\, p(\mathbf{S} \mid \mathcal{T}, \boldsymbol{\psi})\, p(\boldsymbol{\theta}, \mathcal{T}, \boldsymbol{\psi}) \end{aligned}$$

Assumption: $\mathbf{Y}$ and $\mathbf{S}$ independent conditionally on $\mathcal{T}$.

Bayesian Phylogenetics

Goal:

$$
\begin{aligned}
p(\boldsymbol{\theta}, \mathcal{T}, \psi \mid \mathbf{Y}, \mathbf{S}) &\propto p(\mathbf{Y}, \mathbf{S} \mid \boldsymbol{\theta}, \mathcal{T}, \psi)\, p(\boldsymbol{\theta}, \mathcal{T}, \psi) \\
&\propto p(\mathbf{Y} \mid \boldsymbol{\theta}, \mathcal{T})\, p(\mathbf{S} \mid \mathcal{T}, \psi)\, p(\boldsymbol{\theta}, \mathcal{T}, \psi) \\
&\propto p(\mathbf{Y} \mid \boldsymbol{\theta}, \mathcal{T})\, p(\boldsymbol{\theta})\, p(\mathbf{S} \mid \mathcal{T}, \psi)\, p(\mathcal{T}, \psi)
\end{aligned}
$$

Assumption: **Y** and **S** independent conditionally on $\mathcal{T}$.

This talk: $\mathcal{T}$ fixed.

MCMC

Goal: Sample from $p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$

MCMC

Goal: Sample from $p\left(\boldsymbol{\theta} \mid \mathbf{Y}\right) \propto p\left(\mathbf{Y} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\theta}^*$ in $q\left(\boldsymbol{\theta} \,\middle|\, \boldsymbol{\theta}^t\right)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ with probability:

$$r_t = \min\left\{1, \frac{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^*\right)}{p\left(\mathbf{Y} \,\middle|\, \boldsymbol{\theta}^t\right)} \frac{p\left(\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^t\right)} \frac{q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}\right\}.$$

MCMC

Goal: Sample from $p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\theta}^*$ in $q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^t\right)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ with probability:

$$r_t = \min\left\{1, \frac{p(\mathbf{Y} \mid \boldsymbol{\theta}^*)}{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^t\right)} \frac{p(\boldsymbol{\theta}^*)}{p\left(\boldsymbol{\theta}^t\right)} \frac{q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})}\right\}.$$

Gibbs:

- Split $\boldsymbol{\theta} = (\boldsymbol{\theta}_{[1]}, \dots, \boldsymbol{\theta}_{[K]})$.
- Draw $\boldsymbol{\theta}^*$ in $p(\boldsymbol{\theta}_{[k]}|\boldsymbol{\theta}^{(t)}_{[-k]}, \mathbf{Y})$ so that $r_t = 1$.

BM: Gibbs with Conjugate Priors

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

BM: Gibbs with Conjugate Priors

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n \boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1}\mathbf{R})$$

BM: Gibbs with Conjugate Priors

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n\boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1}\mathbf{R})$$

Gibbs:

$$\mathbf{R}|\mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

BM: Gibbs with Conjugate Priors

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n\boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1}\mathbf{R})$$

Gibbs:

$$\mathbf{R}|\mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

$\hookrightarrow$ Automatic sampling in the space of variance matrices.

BM: Gibbs with Conjugate Priors

Likelihood:

$$\mathbf{Y}|\mathbf{R}, \boldsymbol{\mu} \sim \mathcal{MN}(\mathbf{1}_n\boldsymbol{\mu}^T, \mathbf{V}, \mathbf{R})$$

Conjugate Priors:

$$\mathbf{R} \sim \mathcal{IW}(\mathbf{R}_0, \nu)$$
$$\boldsymbol{\mu}|\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}_0, \kappa_0^{-1}\mathbf{R})$$

Gibbs:

$$\mathbf{R}|\mathbf{Y}, \boldsymbol{\mu} \sim \mathcal{IW}(\mathbf{R}_n, \nu_n) \quad \text{with} \quad \mathbf{R}_n = f(\mathbf{Y}, \boldsymbol{\mu}, \mathbf{V})$$

$\hookrightarrow$ Automatic sampling in the space of variance matrices. :-)

OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

Gibbs: Not possible.

OU: No Gibbs

Likelihood:

$$\mathbf{Y}|\mathbf{A}, \mathbf{R}, \boldsymbol{\mu} \not\sim \mathcal{MN}$$

Conjugate Priors: ???

Gibbs: Not possible.

Metropolis - Hasting:

Need to sample in constrained spaces (**A**, **R**).

MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

Distribution: For a distribution π:

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \pi_{\boldsymbol{\nu}}(f(\boldsymbol{\theta})) \times |J_f(\boldsymbol{\theta})|$$

MH in constrained space

Transformation:

$$f : \begin{cases} \mathcal{C}_q \to \mathbb{R}^q \\ \boldsymbol{\theta} \mapsto \boldsymbol{\nu} = f(\boldsymbol{\theta}) \end{cases}$$

Distribution: For a distribution π:

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \pi_{\boldsymbol{\nu}}(f(\boldsymbol{\theta})) \times |J_f(\boldsymbol{\theta})|$$

Metropolis - Hasting: Iterate:

- Draw $\boldsymbol{\nu}^*$ in $q\left(\boldsymbol{\nu} \mid \boldsymbol{\nu}^t\right)$.
- Set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^* = f^{-1}(\boldsymbol{\nu}^*)$ with probability

$$r_t = \min\left\{1, \frac{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^*\right)}{p\left(\mathbf{Y} \mid \boldsymbol{\theta}^t\right)} \frac{p\left(\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^t\right)} \frac{q(\boldsymbol{\nu}^{(t)}|\boldsymbol{\nu}^*)}{q(\boldsymbol{\nu}^*|\boldsymbol{\nu}^{(t)})} \frac{|J_f(\boldsymbol{\theta}^{(t)})|}{|J_f(\boldsymbol{\theta}^*)|}\right\}.$$

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Decomposition: Use correlation matrix $\mathbf{C}$

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Decomposition: Use correlation matrix $\mathbf{C}$

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Decomposition: Use correlation matrix **C**

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

C: Correlation matrix :-(

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Decomposition: Use correlation matrix $\mathbf{C}$

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

$\mathbf{C}$: Correlation matrix :-(

LKJ: Transformation on the space of correlation matrices.

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

Decomposition: Use correlation matrix $\mathbf{C}$

$$\mathbf{R} = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix} \begin{pmatrix} 1 & & C_{kl} \\ & \ddots & \\ C_{kl} & & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}$$

$\boldsymbol{\sigma}$: Real positive :-)

$\mathbf{C}$: Correlation matrix :-(

LKJ: Transformation on the space of correlation matrices.
$\hookrightarrow$ Use "vine" theory.
$\hookrightarrow$ Easier and more efficient: Cholesky representation.

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

$\hookrightarrow$ Each column k is in the half euclidian unit sphere $\mathcal{S}_k^{\mathrm{h}}(\mathbb{R})$:

- $W_{1k}^2 + \cdots + W_{kk}^2 = 1$ (correlation)
- $W_{kk} > 0$ (identifiability)

Variance matrix: LKJ transformation (Lewandowski et al., 2009)

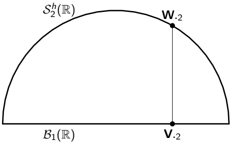
$$\mathbf{C} = \mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}^T \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & W_{pp} \end{pmatrix}$$

$\hookrightarrow$ Each column k is in the half euclidian unit sphere $\mathcal{S}_k^{\mathrm{h}}(\mathbb{R})$:

- $W_{1k}^2 + \cdots + W_{kk}^2 = 1$ (correlation)
- $W_{kk} > 0$ (identifiability)

$\hookrightarrow$ Diffeomorphism to the euclidian (open) ball $\mathcal{B}_{k-1}(\mathbb{R})$:

$$\mathbf{F} : \begin{cases} \mathcal{B}_{k-1}(\mathbb{R}) & \to \mathcal{S}_k^{\mathrm{h}}(\mathbb{R}) \\ \mathbf{V}_{\cdot k} & \mapsto \mathbf{W}_{\cdot k} = \left(\mathbf{V}_{\cdot k},\ \sqrt{1 - \|\mathbf{V}_{\cdot k}\|^2}\right) \end{cases}$$



Precision matrix: LKJ transformation (Lewandowski et al., 2009)

$\hookrightarrow$ Sampling in $\mathcal{B}_{k-1}(\mathbb{R})$?

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

$\hookrightarrow$ Sampling in $\mathcal{B}_{k-1}(\mathbb{R})$? $\rightarrow$ $\mathcal{B}_{k-1}^{\infty}(\mathbb{R})$ is easier.

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

$\hookrightarrow$ Sampling in $\mathcal{B}_{k-1}(\mathbb{R})$? $\rightarrow$ $\mathcal{B}_{k-1}^{\infty}(\mathbb{R})$ is easier.

$\hookrightarrow$ "Fisher Z" transform $\tanh^{-1}$: $\begin{cases}]-1\,,1[\to \mathbb{R} \\ x \mapsto \dfrac{1}{2}\ln\left(\dfrac{1+x}{1-x}\right) \end{cases}$

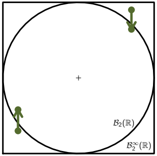
Precision matrix: LKJ transformation (Lewandowski et al., 2009)

$\hookrightarrow$ Sampling in $\mathcal{B}_{k-1}(\mathbb{R})$? $\rightarrow$ $\mathcal{B}^{\infty}_{k-1}(\mathbb{R})$ is easier.

$\hookrightarrow$ "Fisher Z" transform $\tanh^{-1}$: $\begin{cases}]-1\,,1[\rightarrow \mathbb{R} \\ x \mapsto \dfrac{1}{2} \ln\left(\dfrac{1+x}{1-x}\right) \end{cases}$

$\hookrightarrow$ Transformation $\mathcal{B}^{\infty}_{k-1}(\mathbb{R}) \rightarrow \mathcal{B}_{k-1}(\mathbb{R})$?

$$\mathbf{LKJ}_i(\mathbf{z}) = \begin{cases} z_i & \text{if } i = 1 \\ z_i \prod_{k=1}^{i-1}(1 - z_k^2)^{1/2} & \text{if } 1 < i \leq k-1. \end{cases}$$



Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

$$\begin{array}{c} \mathbf{x}_k \in \mathbb{R}^{k-1} \\ \begin{pmatrix} \cdot & x_{12} & \cdots & x_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & x_{p-1p} \\ 0 & & & \cdot \end{pmatrix} \end{array}$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

$$
\begin{array}{ccc}
\mathbf{x}_k \in \mathbb{R}^{k-1} & \rightarrow & \mathbf{z}_k \in \mathcal{B}_{k-1}^{\infty}(\mathbb{R}) \\
\begin{pmatrix}
\cdot & x_{12} & \cdots & x_{1p} \\
 & \ddots & \ddots & \vdots \\
 & & \ddots & x_{p-1p} \\
0 & & & \cdot
\end{pmatrix}
& \rightarrow &
\begin{pmatrix}
\cdot & z_{12} & \cdots & z_{1p} \\
 & \ddots & \ddots & \vdots \\
 & & \ddots & z_{p-1p} \\
0 & & & \cdot
\end{pmatrix}
\end{array}
$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

$$
\begin{array}{ccccc}
\mathbf{x}_k \in \mathbb{R}^{k-1} & \to & \mathbf{z}_k \in \mathcal{B}^{\infty}_{k-1}(\mathbb{R}) & \to & \mathbf{V}_k \in \mathcal{B}_{k-1}(\mathbb{R}) \\
\begin{pmatrix} \cdot & x_{12} & \cdots & x_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & x_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & z_{12} & \cdots & z_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & z_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & V_{12} & \cdots & V_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & V_{p-1p} \\ 0 & & & \cdot \end{pmatrix}
\end{array}
$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

$$\begin{array}{ccccc}
\mathbf{x}_k \in \mathbb{R}^{k-1} & \to & \mathbf{z}_k \in \mathcal{B}^{\infty}_{k-1}(\mathbb{R}) & \to & \mathbf{V}_k \in \mathcal{B}_{k-1}(\mathbb{R}) \\
\begin{pmatrix} \cdot & x_{12} & \cdots & x_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & x_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & z_{12} & \cdots & z_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & z_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & V_{12} & \cdots & V_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & V_{p-1p} \\ 0 & & & \cdot \end{pmatrix}
\end{array}$$

$$\begin{array}{cc}
\to & \mathbf{W}_k \in \mathcal{S}^{\mathrm{h}}_k(\mathbb{R}) \\
\to & \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & \ddots & \vdots \\ & & \ddots & W_{p-1p} \\ 0 & & & W_{pp} \end{pmatrix}
\end{array}$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

$$
\begin{array}{ccccc}
\mathbf{x}_k \in \mathbb{R}^{k-1} & \to & \mathbf{z}_k \in \mathcal{B}^{\infty}_{k-1}(\mathbb{R}) & \to & \mathbf{V}_k \in \mathcal{B}_{k-1}(\mathbb{R}) \\
\begin{pmatrix} \cdot & x_{12} & \cdots & x_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & x_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & z_{12} & \cdots & z_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & z_{p-1p} \\ 0 & & & \cdot \end{pmatrix} & \to & \begin{pmatrix} \cdot & V_{12} & \cdots & V_{1p} \\ & \ddots & \ddots & \vdots \\ & & \ddots & V_{p-1p} \\ 0 & & & \cdot \end{pmatrix}
\end{array}
$$

$$
\begin{array}{cccc}
\to & \mathbf{W}_k \in \mathcal{S}^{h}_k(\mathbb{R}) & \to & \mathbf{C} = \mathbf{W}^T\mathbf{W} \in \{\text{correlation matrices}\} \\
\to & \begin{pmatrix} 1 & W_{12} & \cdots & W_{1p} \\ & W_{22} & \ddots & \vdots \\ & & \ddots & W_{p-1p} \\ 0 & & & W_{pp} \end{pmatrix} & \to & \begin{pmatrix} 1 & C_{12} & \cdots & C_{1p} \\ C_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_{p-1p} \\ C_{1p} & \cdots & C_{p-1p} & 1 \end{pmatrix}
\end{array}
$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

Jacobian: Can be computed.

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

Jacobian: Can be computed.

Associated distribution:

$$\mathrm{LKJ}(\mathbf{C} \mid \eta) = c_p(\eta) \left|\mathbf{C}\right|^{\eta-1}$$

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

Jacobian: Can be computed.

Associated distribution:

$$\text{LKJ}(\mathbf{C} \mid \eta) = c_p(\eta) \, |\mathbf{C}|^{\eta - 1}$$

$\eta = 1$:	Uniform.
$\eta > 1$:	Peak around identity matrix.
$0 < \eta < 1$:	Trough around identity matrix.

Precision matrix: LKJ transformation (Lewandowski et al., 2009)

Result: Smooth transformation from $\mathbb{R}^{p(p-1)/2}$ to the space of correlation matrices.

Jacobian: Can be computed.

Associated distribution:

$$\mathrm{LKJ}(\mathbf{C} \mid \eta) = c_p(\eta)\,|\mathbf{C}|^{\eta-1}$$

$\eta = 1$:	Uniform.
$\eta > 1$:	Peak around identity matrix.
$0 < \eta < 1$:	Trough around identity matrix.

$\hookrightarrow$ Same as taking a "spherical beta" on each of the $\mathbf{V}_{\cdot k}$:

$$\mathrm{SBeta}(\mathbf{V}_{\cdot k}|\beta) \propto (1 - \|\mathbf{V}_{\cdot k}\|^2)^{\beta_k} \quad \text{with} \quad \beta_k = \eta + (p-k)/2$$

Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0$$

Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathsf{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathsf{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$: Decomposition is unique.

Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathrm{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$: Decomposition is unique.

Sampling:

$\mathbf{P}_{\cdot k}$: Same as $\mathbf{W}_{\cdot k}$

Attenuation matrix

Assumptions: $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$

$$\lambda_k \in \mathbb{R} \qquad \lambda_k > 0 \qquad \lambda_1 < \lambda_2 < \cdots < \lambda_p$$

Identifiability: $\mathbf{P}_{\cdot k} \in \mathcal{S}_p^{\mathrm{h}}(\mathbb{R})$

$$\|\mathbf{P}_{\cdot k}\| = 1 \qquad \mathbf{P}_{pk} > 0$$

$\hookrightarrow$: Decomposition is unique.

Sampling:

$\mathbf{P}_{\cdot k}$: Same as $\mathbf{W}_{\cdot k}$

$\boldsymbol{\Lambda}$: Use $\log(\lambda_i) - \log(\lambda_{i-1})$

Summary

We can sample from the posterior $p(\mathbf{P}, \mathbf{\Lambda}, \mathbf{C}, \boldsymbol{\sigma}, \boldsymbol{\mu} | \mathbf{Y})$

Summary

We can sample from the posterior $p(\mathbf{P}, \mathbf{\Lambda}, \mathbf{C}, \boldsymbol{\sigma}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{C}$: LKJ transformation
- $\mathbf{P}$: Unitary eigen-vectors
- $\mathbf{\Lambda}$, $\boldsymbol{\sigma}$: Positive (ordered) vectors

Summary

We can sample from the posterior $p(\mathbf{P}, \mathbf{\Lambda}, \mathbf{C}, \boldsymbol{\sigma}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{C}$: LKJ transformation
- $\mathbf{P}$: Unitary eigen-vectors
- $\mathbf{\Lambda}$, $\boldsymbol{\sigma}$: Positive (ordered) vectors

Priors:

- $\mathbf{C}$: LKJ distribution
- $\mathbf{P}$: Spherical beta
- $\mathbf{\Lambda}$, $\boldsymbol{\sigma}$: half-normal

Summary

We can sample from the posterior $p(\mathbf{P}, \mathbf{\Lambda}, \mathbf{C}, \boldsymbol{\sigma}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- $\mathbf{C}$: LKJ transformation
- $\mathbf{P}$: Unitary eigen-vectors
- $\mathbf{\Lambda}$, $\boldsymbol{\sigma}$: Positive (ordered) vectors

Priors:

- $\mathbf{C}$: LKJ distribution
- $\mathbf{P}$: Spherical beta
- $\mathbf{\Lambda}$, $\boldsymbol{\sigma}$: half-normal

$\hookrightarrow$ We have a running random walk MCMC.

Summary

We can sample from the posterior $p(\mathbf{P}, \mathbf{\Lambda}, \mathbf{C}, \boldsymbol{\sigma}, \boldsymbol{\mu} | \mathbf{Y})$

Transformations:

- **C**: LKJ transformation
- **P**: Unitary eigen-vectors
- **Λ**, $\boldsymbol{\sigma}$: Positive (ordered) vectors

Priors:

- **C**: LKJ distribution
- **P**: Spherical beta
- **Λ**, $\boldsymbol{\sigma}$: half-normal

$\hookrightarrow$ We have a running random walk MCMC.

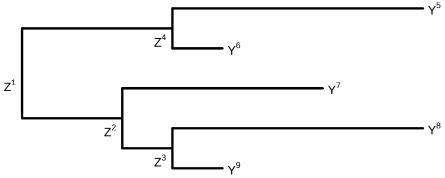
Question: Can we use the tree ?

General Model

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

General Model

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



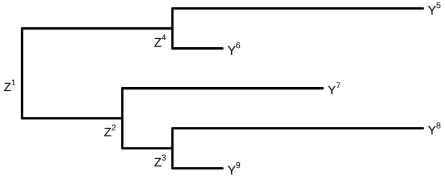
$$\mathbf{X} = \begin{cases} \mathbf{Z} & : \text{latent nodes} \\ \mathbf{Y} & : \text{tips} \end{cases}$$

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

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

General Model

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



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

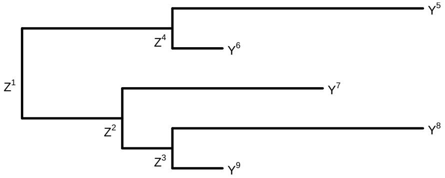
$$\mathbf{X}^1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{X}^j \,\middle|\, \mathbf{X}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\text{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

BM: $\mathbf{q}_i = \mathbf{I}_p$, $\mathbf{r}_i = \mathbf{0}_p$, $\boldsymbol{\Sigma}_i = \ell_i \mathbf{R}$

General Model

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



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

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

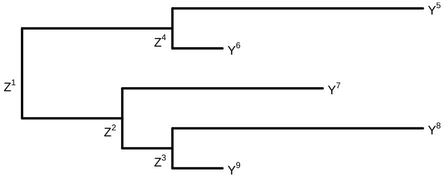
$$\mathbf{X}^j \,\Big|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

BM: $\mathbf{q}_i = \mathbf{I}_p$, $\mathbf{r}_i = \mathbf{0}_p$, $\boldsymbol{\Sigma}_i = \ell_i \mathbf{R}$

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

General Model

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



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

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

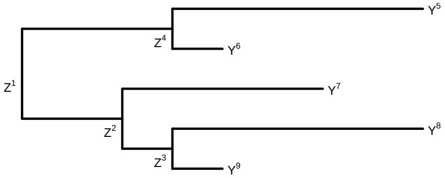
$$\mathbf{X}^j \,\Big|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

BM: $\mathbf{q}_i = \mathbf{I}_p$, $\mathbf{r}_i = \mathbf{0}_p$, $\boldsymbol{\Sigma}_i = \ell_i \mathbf{R}$

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

Drift, shifts, Integrated OU...

Efficient Computations



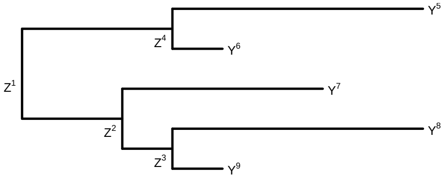
$$\mathbf{X} = \begin{cases} \mathbf{Z} & : \text{latent nodes} \\ \mathbf{Y} & : \text{tips} \end{cases}$$

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

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

Likelihood: $p(\mathbf{X}^1 | \mathbf{Y})$ in one post-order traversal of the tree.

Efficient Computations



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

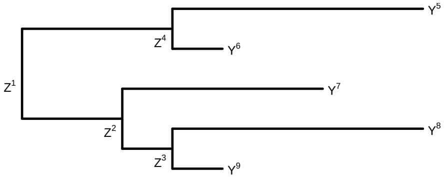
$$\mathbf{X}^1 \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Gamma}\right) \qquad \text{root}$$

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

Likelihood: $p(\mathbf{X}^1 | \mathbf{Y})$ in one post-order traversal of the tree.

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

Efficient Computations



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

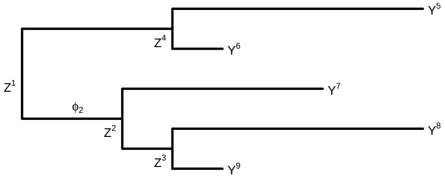
$$\mathbf{X}^1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{X}^j \,\Big|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Likelihood: $p(\mathbf{X}^1 | \mathbf{Y})$ in one post-order traversal of the tree.
$\hookrightarrow$ “Pruning”, “Gaussian elimination”, “Phylogenetic Kalman filter”, ...

Difficulty: Numerical robustness.

Efficient Computations: Gradient



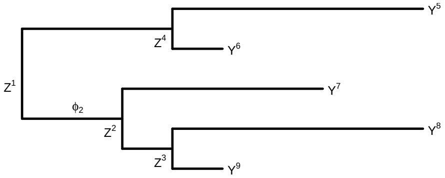
$$\mathbf{X} = \begin{cases} \mathbf{Z} & : \text{latent nodes} \\ \mathbf{Y} & : \text{tips} \end{cases}$$

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

$$\mathbf{X}^j \,\Big|\, \mathbf{X}^{\text{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\text{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Branch-specific Gradient: $\frac{\partial}{\partial \phi_j} [\log p(\mathbf{Y})] = \mathbb{E}\left[\mathbf{F}(\mathbf{X}^j; \phi_j) \,\middle|\, \mathbf{Y}\right]$

Efficient Computations: Gradient



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

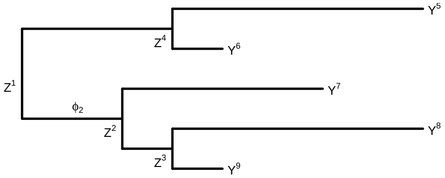
$$\mathbf{X}^1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \qquad \text{root}$$

$$\mathbf{X}^j \mid \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Branch-specific Gradient: $\frac{\partial}{\partial \phi_j}[\log p(\mathbf{Y})] = \mathbb{E}\left[\mathbf{F}(\mathbf{X}^j; \phi_j) \mid \mathbf{Y}\right]$

$\hookrightarrow$ One pre-order traversal.

Efficient Computations: Gradient



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

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

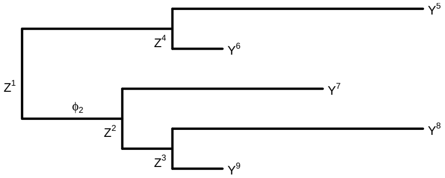
$$\mathbf{X}^j \,\middle|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Branch-specific Gradient: $\frac{\partial}{\partial \phi_j}[\log p(\mathbf{Y})] = \mathbb{E}\left[\mathbf{F}(\mathbf{X}^j; \phi_j) \,\middle|\, \mathbf{Y}\right]$

$\hookrightarrow$ One pre-order traversal.

Chain rule: Get the gradient w.r.t. any parameter.

Efficient Computations: Gradient



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

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

$$\mathbf{X}^j \,\middle|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}^{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Branch-specific Gradient: $\frac{\partial}{\partial \phi_j} \left[\log p(\mathbf{Y})\right] = \mathbb{E}\left[\mathbf{F}(\mathbf{X}^j; \phi_j) \,\middle|\, \mathbf{Y}\right]$

$\hookrightarrow$ One pre-order traversal.

Chain rule: Get the gradient w.r.t. any parameter. $\rightarrow$ HMC

Implementation

(Suchard et al., 2018)

BEAST

Bayesian Evolutionary Analysis Sampling Trees

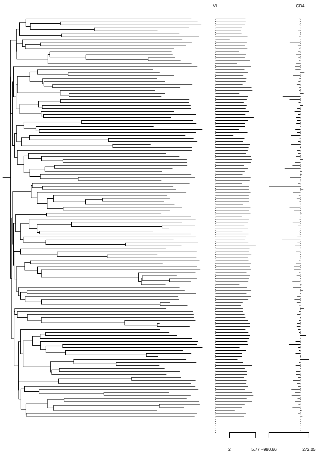
- MCMC for tree estimation
- Comprehensive set of tools:
  - Factor model
  - Marginal Likelihood
  - ...
- Developed in `Java` since 2002.
- This is `BEAST` 1.10.

Don't ask about `BEAST 2`.

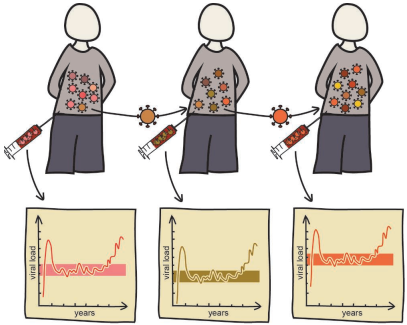
What's new:

- Flexible OU models
- Efficient sampling of variance
- Efficient HMC (in progress)

HIV virulence heritability (Alizon et al., 2010; Vrancken et al., 2015)



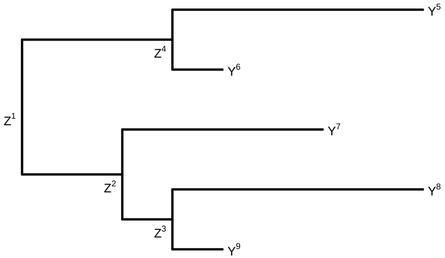
CD4: CD4+ T cells decline rate
VL: Set point viral load



Fraser et al. (2014)

Questions: Is virulence "heritable"? What model of trait evolution?

Heritability

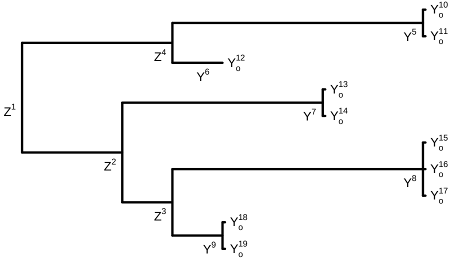


$$\mathbf{X} = \begin{cases} \mathbf{Z} & : \text{latent nodes} \\ \mathbf{Y} & : \text{tips} \end{cases}$$

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

$$\mathbf{X}^j \,\Big|\, \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}_{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{tips and nodes}$$

Heritability



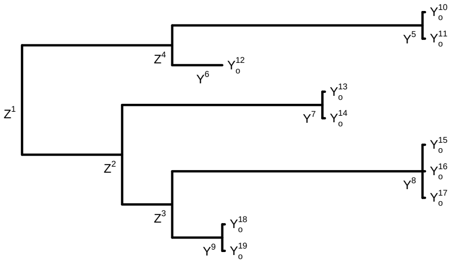
$$\mathbf{X} = \begin{cases} \mathbf{Z} & : \text{latent nodes} \\ \mathbf{Y} & : \text{latent tips} \\ \mathbf{Y}_o & : \text{observed traits} \end{cases}$$

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

$$\mathbf{X}^j \mid \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}_{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{latent tips and nodes}$$

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

Heritability



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

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

$$\mathbf{X}^j \mid \mathbf{X}^{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_i \mathbf{X}_{\mathrm{pa}(i)} + \mathbf{r}_i,\ \boldsymbol{\Sigma}_i\right) \qquad \text{latent tips and nodes}$$

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

"Heritability": $$h_k^2 = \frac{V(\mathbf{Y}_{\cdot k})}{V(\mathbf{Y}_{o\cdot k})} \approx \frac{\sigma_k^2 \tilde{t}}{\sigma_k^2 \tilde{t} + s_k}$$

Models

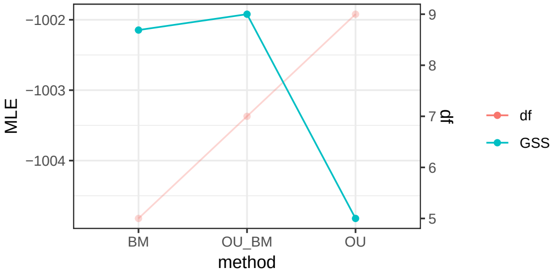
(Alizon et al., 2010)

We use three different models:

- **BM** no selection on the traits.
- **OU-BM** selection on VL, not on CD4.
- **OU** selection on both traits.

- Each model is fitted using a MCMC.
- Estimated Marginal likelihood is used to compare them.

Results



Heritability (using OU-BM):

VL $h^2 = 17\%$ $[0.007, 82.5]\%$ (95% CI)

CD4 $h^2 = 0.02\%$ $[0.0024, 0.16]\%$

"Consistent" with previous estimates.

Conclusion and Perspectives

A general framework for trait evolution with dated tips.

Main Features:

- Flexible models
- Flexible implementation
- Efficient algorithms
- Applicable to virology

Perspectives:

- Develop HMC
- Apply to larger datasets
- Other questions: geographical spread, comparative studies, ...

Bibliography

Alizon, von Wyl, Stadler, et al. 2010. *PLoS Pathog.* 6.

Aristide, Bastide, dos Reis, et al. 2018. *Evolution*. 72:2697–2711.

Betancourt. 2017. *arXiv e-print*. .

Dellicour, Rose, Faria, et al. 2017. *Mol. Biol. Evol.* 34:2563–2571.

Fraser, Lythgoe, Leventhal, et al. 2014. *Science (80-.).* 343:1243727–1243727.

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

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

Lewandowski, Kurowicka, Joe. 2009. *J. Multivar. Anal.* 100:1989–2001.

Suchard, Lemey, Baele, et al. 2018. *Virus Evol.* 4:1–5.

Vrancken, Lemey, Rambaut, et al. 2015. *Methods Ecol. Evol.* 6:67–82.

Photo Credits:

Thank you for listening

Kasteel van Arenberg

Appendices

Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) = \text{Potential energy} + \text{Kinetic energy}$$

Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned}\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy}\\ &= U(\mathbf{q}) + K(\mathbf{p})\\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}}\end{aligned}$$

Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned} \text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy} \\ &= U(\mathbf{q}) + K(\mathbf{p}) \\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}} \end{aligned}$$

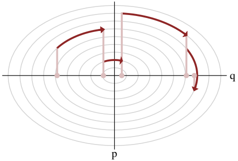
H **total energy** invariant by Hamiltonian dynamic: $\begin{cases} \dfrac{d\mathbf{p}}{dt} = \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \dfrac{d\mathbf{q}}{dt} = -\nabla_{\mathbf{p}} K(\mathbf{p}) \end{cases}$

Hamiltonian Monte Carlo (Betancourt, 2017)

Idea: Introduce the "momentum" $\mathbf{p}$ of the parameters $\mathbf{q}$.

$$\begin{aligned}\text{Hamiltonian } H(\mathbf{q}, \mathbf{p}) &= \text{Potential energy} + \text{Kinetic energy} \\ &= U(\mathbf{q}) + K(\mathbf{p}) \\ &= -\log \underbrace{p(\mathbf{Y}|\mathbf{q})p(\mathbf{q})}_{\text{posterior}} - \log \underbrace{\phi(\mathbf{p})}_{\text{Gaussian}}\end{aligned}$$

H **total energy** invariant by Hamiltonian dynamic: $\begin{cases} \dfrac{d\mathbf{p}}{dt} = \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \dfrac{d\mathbf{q}}{dt} = -\nabla_{\mathbf{p}} K(\mathbf{p}) \end{cases}$



1. Draw random moments $\mathbf{p}$.
2. Propose a new $\mathbf{q}$ from the dynamic.

back