

Scalable approximate inference for the phylogenetic threshold model

Nathaniel S. Pope¹ Eric Goolsby²

27 June 2017

¹University of Texas at Austin, nspope@utexas.edu

²Brown University

the threshold model

Quantitative genetics for categorical trait evolution¹

- *latent* continuous traits \mathbf{x} follow Brownian motion
- *observed* binary (categorical) phenotypes \mathbf{y}
- $y_i = 1$ if $x_i > 0$, else $y_i = 0$

¹Felsenstein 2005, Wright 1934

threshold model

Quantitative genetics for categorical trait evolution¹

- *latent* continuous traits \mathbf{x} follow Brownian motion
- *observed* binary (categorical) phenotypes \mathbf{y}
- $y_i = 1$ if $x_i > 0$, else $y_i = 0$

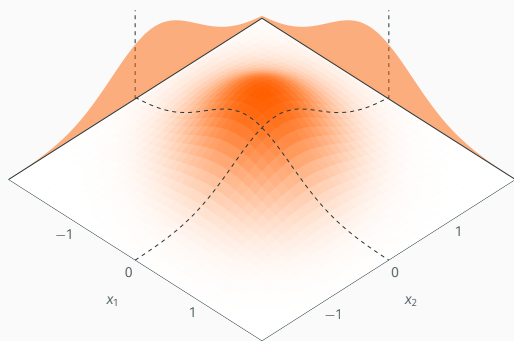
$$|2\pi\mathbf{K}|^{-\frac{1}{2}} \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \exp\{-0.5(\mathbf{x} - \boldsymbol{\mu})' \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\mu})\} d\mathbf{x}_1 \dots d\mathbf{x}_n$$

Likelihood is Gaussian integral

- $\{l_i, u_i\} = \{0, \infty\}$ if $y_i = 1$
- phylogenetic v.c.v. \mathbf{K} , ancestral states $\boldsymbol{\mu}$

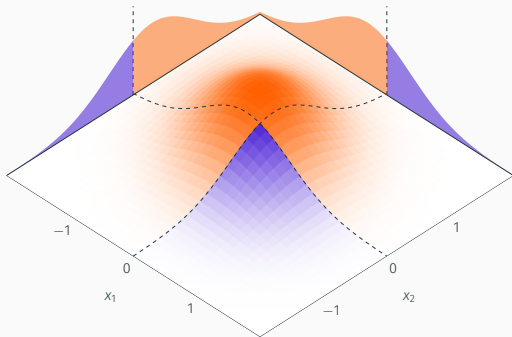
¹Felsenstein 2005, Wright 1934

visual example



$$p(x_1, x_2 | \mu, K) \equiv \mathcal{N}(\mu, K)$$

visual example



$$\Pr(y_1 = 1, y_2 = 0 | \mu, K) = \int_0^\infty \int_{-\infty}^0 p(x_1, x_2 | \mu, K) dx_2 dx_1$$

difficult integration problem

Felsenstein 2005

- MCMC + expectation-maximization

Hadfield 2015

- efficient MCMC

Hiscott *et al* 2016

- single trait across tree with n taxa
- numeric integration on N -grid
- complexity $\mathcal{O}(nN^2)$

Scalable approximation to likelihood

- “scalable” in tips n , traits m
- flexible; e.g. mixed continuous/categorical observed traits
- complexity $\mathcal{O}(dnm)$, d is avg. # edges from root to tips

Scalable approximation to likelihood

- “scalable” in tips n , traits m
- flexible; e.g. mixed continuous/categorical observed traits
- complexity $\mathcal{O}(dnm)$, d is avg. # edges from root to tips

Strategy

1. vector multiplication with covariance matrix in $\mathcal{O}(dnm)$
2. variational approximation of likelihood, closed form is linear equation

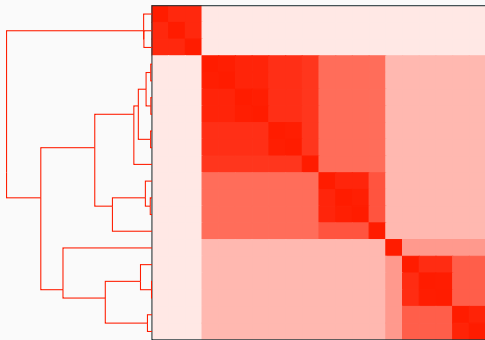
first ingredient

Full covariance matrix:

- V : phylogenetic covariance matrix
- Λ : trait covariance across taxa
- Σ : trait covariance within taxa

$$K = \Lambda \otimes V + \Sigma \otimes I$$

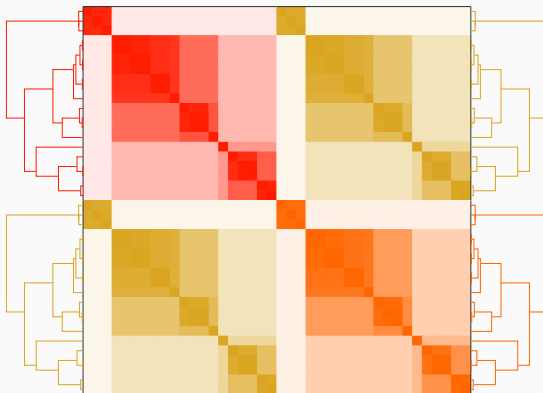
covariance model



$K =$

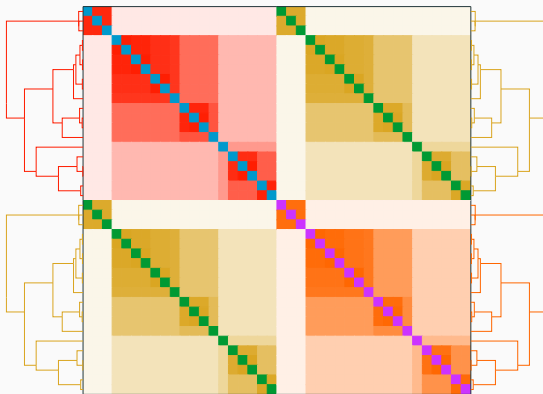
V

covariance model



$$K = \begin{bmatrix} \Lambda_{1,1} & \Lambda_{2,1} \\ \Lambda_{1,2} & \Lambda_{2,2} \end{bmatrix} \otimes V$$

covariance model



$$K = \begin{bmatrix} \Lambda_{1,1} & \Lambda_{2,1} \\ \Lambda_{1,2} & \Lambda_{2,2} \end{bmatrix} \otimes V + \begin{bmatrix} \Sigma_{1,1} & \Sigma_{2,1} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{bmatrix} \otimes I$$

Full covariance matrix:

- V : phylogenetic covariance matrix
- Λ : trait covariance across taxa
- Σ : trait covariance within taxa

$$K = \Lambda \otimes V + \Sigma \otimes I$$

Ho and Ané 2014:

- $x'K^{-1}$ in $\mathcal{O}(dnm^3)$
- $|K^{-1}|, x'K^{-1}x$ in $\mathcal{O}(nm^3)$

a low rank extension

Low rank covariance matrix:

- $\Lambda^* = LL'$ where $\text{rank}(L) \ll \text{rank}(\Lambda)$
- $\Sigma^* = QQ'$ where $\text{rank}(Q) \ll \text{rank}(\Sigma)$
- E is diagonal

$$K = \Lambda^* \otimes V + (\Sigma^* + E) \otimes I$$

a low rank extension

Low rank covariance matrix:

- $\Lambda^* = LL'$ where $\text{rank}(L) \ll \text{rank}(\Lambda)$
- $\Sigma^* = QQ'$ where $\text{rank}(Q) \ll \text{rank}(\Sigma)$
- E is diagonal

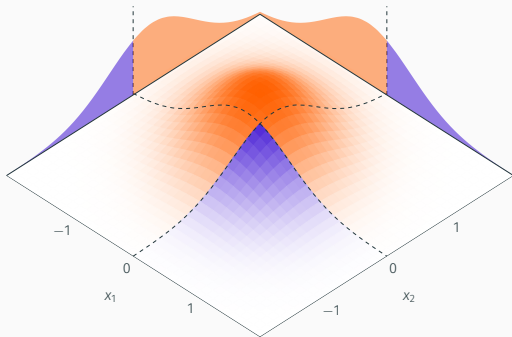
$$K = \Lambda^* \otimes V + (\Sigma^* + E) \otimes I$$

For fixed rank ...

- $x'K^{-1}$ in $\mathcal{O}(dnm)$
- $|K^{-1}|, x'K^{-1}x$ in $\mathcal{O}(nm)$
- possible to efficiently model many *correlated* traits

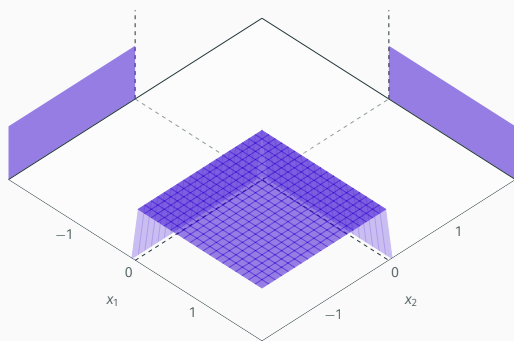
second ingredient

construction with indicator function



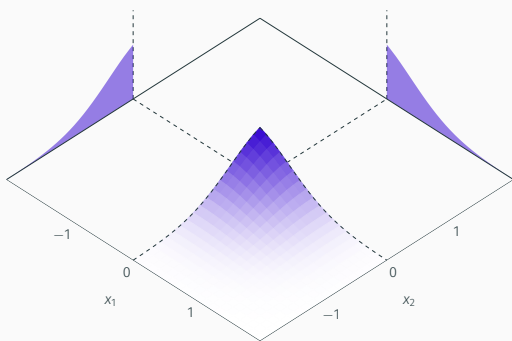
$$\Pr(y_1 = 1, y_2 = 0 | \mu, K) = \int_0^\infty \int_{-\infty}^0 p(x_1, x_2 | \mu, K) dx_2 dx_1$$

construction with indicator function



$$q_1(x_1; y_1) \cdot q_2(x_2; y_2) = \mathbb{I}[x_1 > 0] \cdot \mathbb{I}[x_2 < 0]$$

construction with indicator function



$$\Pr(\mathbf{y}|\mu, K) = \int p(\mathbf{x}|\mu, K) \prod q_i(x_i; y_i) d\mathbf{x}$$

Target function:

$$f(\mathbf{x}; \mathbf{y}) = \underbrace{p(\mathbf{x}; \mu, \Sigma)}_{\text{Gaussian}} \prod \underbrace{q_i(x_i; y_i)}_{\text{indicators}}$$

variational approximation

Target function:

$$f(\mathbf{x}; \mathbf{y}) = \underbrace{p(\mathbf{x}; \mu, \Sigma)}_{\text{Gaussian}} \prod \underbrace{q_i(x_i; y_i)}_{\text{indicators}}$$

Variational approximation:

$$f^*(\mathbf{x}; \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*) = \underbrace{p(\mathbf{x}; \mu, \Sigma)}_{\text{Gaussian}} \prod \underbrace{q_i^*(x_i; \mu_i^*, \sigma_i^*)}_{\text{approximations}}$$

- q_i^* are univariate Gaussians with params μ_i^*, σ_i^*
- the q_i^* approximate the q_i

How to approximate?

Find variational parameters μ^*, σ^* that minimize KL-divergence:

$$\text{KL}(f||f^*) = \int f(x; y) \ln \frac{f(x; y)}{f^*(x; \mu^*, \sigma^*)} dx$$

²Minka 2001, Cunningham *et al* 2011

How to approximate?

Find variational parameters μ^*, σ^* that minimize KL-divergence:

$$\text{KL}(f||f^*) = \int f(x; y) \ln \frac{f(x; y)}{f^*(x; \mu^*, \sigma^*)} dx$$

Expectation propagation²

- approximately minimizes KL-divergence
- iterative scheme, each iteration is $\mathcal{O}(dnm)$

²Minka 2001, Cunningham *et al* 2011

what's the point?

Gaussians are tractable!

Approximation to original integral problem:

$$\ln \int f(\mathbf{x}; \mathbf{y}) d\mathbf{x} \approx \ln \int f^*(\mathbf{x}; \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*) d\mathbf{x}$$

- r.h.s. evaluates to linear function
- calculated in $\mathcal{O}(dnm)$ with Ho-Ané algorithm

what's the point?

Gaussians are tractable!

Approximation to original integral problem:

$$\ln \int f(\mathbf{x}; \mathbf{y}) d\mathbf{x} \approx \ln \int f^*(\mathbf{x}; \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*) d\mathbf{x}$$

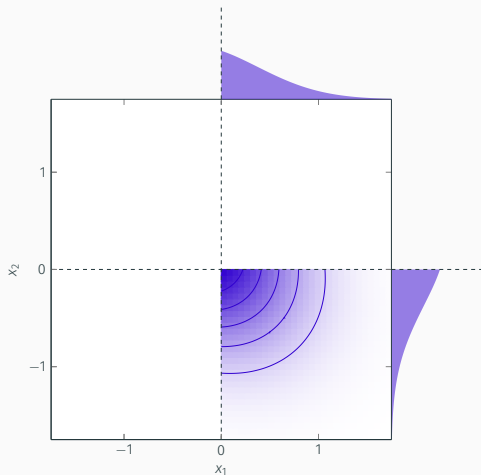
- r.h.s. evaluates to linear function
- calculated in $\mathcal{O}(dnm)$ with Ho-Ané algorithm

Take-home

- original function intractable ...
- ... so replace with tractable approximate function
- expectation propagation used to find best approximation

in practice

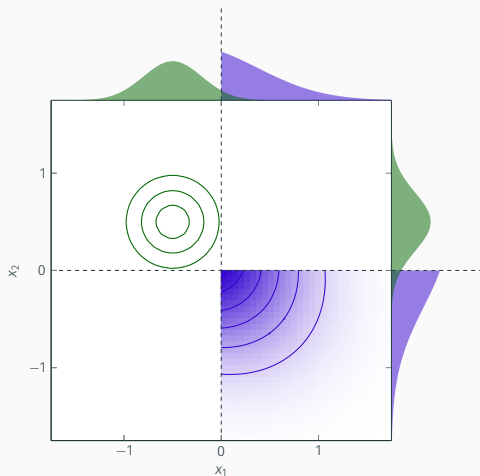
expectation propagation



log true integral

-1.8920

expectation propagation (iteration 0)



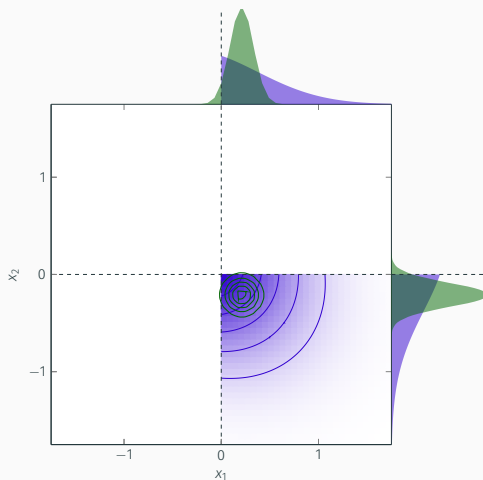
log approx. integral

-3.0261

log true integral

-1.8920

expectation propagation (iteration 1)



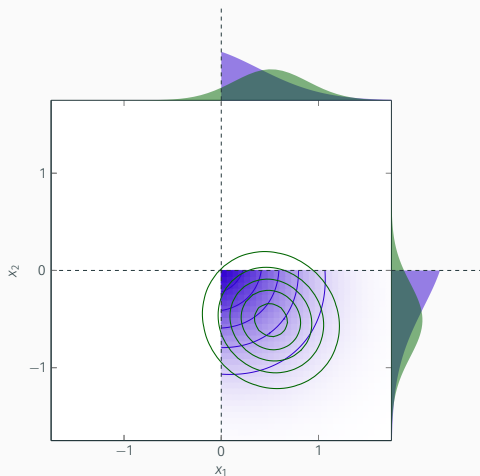
log approx. integral

-1.8929

log true integral

-1.8920

expectation propagation (iteration 2)



log approx. integral

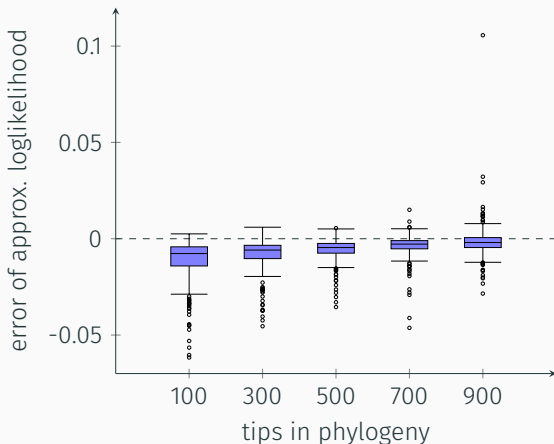
-1.8921

log true integral

-1.8920

accuracy (one trait)

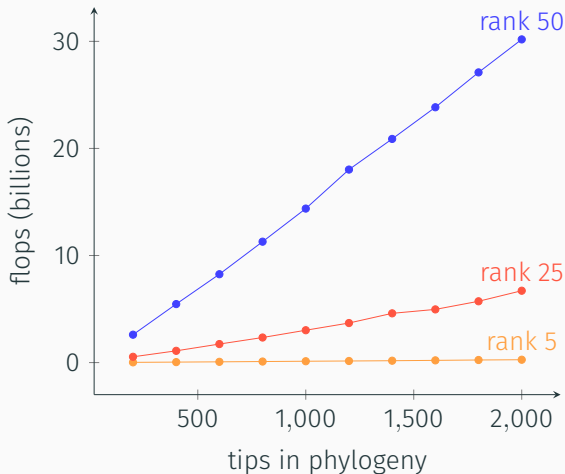
Error in approximation³; simulated data with **single trait**:



³using integration method from Hiscott *et al.* 2016 as ground truth

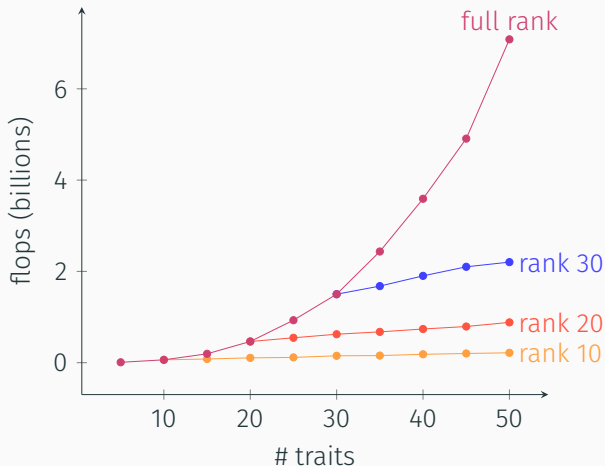
scalability with number of tips, trait covariance rank

Simulated data with 50 binary traits, low rank approximation:



scalability with number of tips, trait covariance rank

Simulated data with 500 tips, low rank approximation:



An accurate approximate likelihood for the threshold model that scales with traits, taxa.

Advantages

- arbitrarily many traits, tips
- low-cost partial derivatives
- mixed continuous-categorical traits
- different flavours of BM (Ornstein-Uhlenbeck, etc.)
- accommodates missing trait data

An accurate approximate likelihood for the threshold model that scales with traits, taxa.

Ongoing work

- MLE via geometric optimization
- full variational inference
- analytic error bounds
- C++/R code (in progress) at GitHub:

`github.com/nspope/epee`



Thanks!

we relied heavily on results from the work of:

- J. Cunningham & colleagues
- T. Minka
- C. Ané & L. Ho

Observed continuous traits \mathbf{z} , latent continuous traits \mathbf{x}

Joint distribution:

$$p(\mathbf{z}, \mathbf{x}) \equiv \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_z \\ \boldsymbol{\mu}_x \end{bmatrix}, \begin{bmatrix} \mathbf{K}_z & \mathbf{K}_{zx} \\ \mathbf{K}_{xz} & \mathbf{K}_x \end{bmatrix} \right)$$

Observed continuous traits \mathbf{z} , latent continuous traits \mathbf{x}

Factors as:

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z}) \rightarrow$$

$$p(\mathbf{z}) \equiv \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{K}_z),$$

$$p(\mathbf{x}|\mathbf{z}) \equiv \mathcal{N}(\boldsymbol{\mu}_x + \mathbf{K}_{xz}\mathbf{K}_z^{-1}(\mathbf{z} - \boldsymbol{\mu}_z), \mathbf{K}_x - \mathbf{K}_{xz}\mathbf{K}_z^{-1}\mathbf{K}_{zx})$$

Observed continuous traits \mathbf{z} , latent continuous traits \mathbf{x}

Likelihood becomes:

$$\int_{l(y)}^{u(y)} p(\mathbf{z}, \mathbf{x}) d\mathbf{x} = p(\mathbf{z}) \int_{l(y)}^{u(y)} p(\mathbf{x}|\mathbf{z}) d\mathbf{x}$$

With modification, Ho-Ané algorithm applies to conditional distribution $p(\mathbf{x}|\mathbf{z})$