Gaussian Phylogenetic reconstruction of correlations

Alejandro Lage-Castellanos, Pierre Barrat-Charlaix

April 26, 2018

1 Direct problem: phylogenetic tree under gaussian evolution

We consider a genome characterized by a vector $\vec{x}$ with an equilibrium probability distribution given by

$$P(\vec{x}) = \frac{1}{Z} \exp H(\vec{x}) \qquad \text{with } \mathcal{H}(\vec{x}) = \vec{x}^T J \vec{x} + \vec{h} \cdot \vec{x} \tag{1}$$

where the degrees of freedom $x_i \in \mathbb{R}, i \in [1 \dots N]$ are continuous variables, and the interactions among them J_{ij} are fixed, but randomly selected from a given ensemble of symmetric negative definite matrices. Notice the absense of the usual minus sign in front of the Hamiltonian. So, usually the vector $\vec{x}$ will be near the maximum of $\mathcal{H}(\vec{x})$

We will assume that the matrix J have some eigenvalues close to zero, so the genetic sequences can drift with relatively low cost in the subsapce generated by such eigenvectors.

1.1 Evolution process

We consider a phylogenetic tree construction starting from a vector $\vec{x}_0$ extracted randomly from the equilibrium distribution (1). Then at every point of our evolution algorithm we do the following

Algorithm 1 Generates tree of evolved vectors $T = \{\vec{x}_0, \vec{x}_{1,1}, \vec{x}_{1,2}, \dots\}$

Require: Root $\vec{x}_0$, number of generations G, distance to mutants d.
Ensure: Returns T

```
T = {x_0}
L = {x_0} {Leaves of the tree (the last added nodes)}
for (g = 1; g < G; g+ = 1; ) do
    L_new = {}
    for (x in L) do
        x_child1 = Monte-Carlo-evolve(x, d)
        x_child2 = Monte-Carlo-evolve(x, d)
        Append T <-= {x_child1, x_child2}
        Append L_new <-= {x_child1, x_child2}
    end for
    L = L_new
end for
return T = {x_0, x_1,1, x_1,2, ...}
```

The step named Monte-Carlo-evolve$(\vec{x}, d)$, carries out a Monte Carlo evolution starting at configuration $\vec{x}$ and until the configuration found $\vec{x}^t$ is at distance

$$d = \frac{1}{N} ||\vec{x} - \vec{x}^t||_2$$

This means that all new generation sequences are at the same distance of their parents.

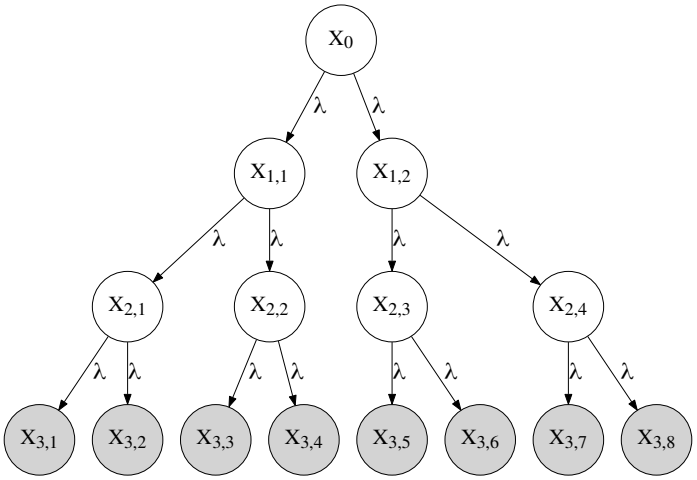

Figure 1: Phylogenetic model for genetic drifting

1.2 Likelihood of observed sequences

In figure 1 we represent the phylogenetic tree generated by an evolution process as described in algorithm 1. The observed sequences are those of the last generation $\vec{x}_{G,\cdot}$ and are portrayed in gray. We assume as known also the model $J, \vec{h}$ and the distances d at which the tree was created, but we ignore the set of sequences $\vec{x}_{g,\cdot}$ in previous generations $g < G$.

Let us define the part of the interactions concerning a node and its children as follows

$$K(\vec{x}_0, \vec{x}_{1,1}, \vec{x}_{1,2}) = \exp\left(\mathcal{H}(\vec{x}_0) + \lambda_0 \vec{x}_0 \cdot \vec{x}_{1,1} + \lambda_0 \vec{x}_0 \cdot \vec{x}_{1,1}\right) \tag{2}$$

where the coupling constants λ are there to ensure that the distance between the father and each children is consistently fixed to have the given expected value.

The full probability distribution of the process in figure 1 is

$$\begin{aligned} P(T) \quad = \quad & \frac{1}{Z} K(\vec{x}_0, \vec{x}_{1,1}, \vec{x}_{1,2}) K(\vec{x}_{1,1}, \vec{x}_{2,1}, \vec{x}_{2,2}) K(\vec{x}_{1,2}, \vec{x}_{2,3}, \vec{x}_{2,4}) \\ & K(\vec{x}_{2,1}, \vec{x}_{3,1}, \vec{x}_{3,2}) K(\vec{x}_{2,2}, \vec{x}_{3,3}, \vec{x}_{3,4}) \\ & K(\vec{x}_{2,3}, \vec{x}_{3,5}, \vec{x}_{3,6}) K(\vec{x}_{2,4}, \vec{x}_{3,7}, \vec{x}_{3,8}) \end{aligned} \tag{3}$$

If this is correct, the marginal over any of these variables have to be the same distribution (1). BUT IS NOT

1.3 Reconstruction of hidden sequences

2 Ornstein-Uhlenbeck dynamics

The system studied is the following. A tree – balanced and binary in the following – is given, with $K+1$ levels and a time Δt being assigned to each of its branches. A root configuration $\vec{x}$ is chosen with gaussian probability

$$P_{eq}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}|}} \exp\left\{ -\frac{1}{2} \vec{x}^T \boldsymbol{C}^{-1} \vec{x} \right\}.$$

It then evolves using dynamics described below. At each division event of the tree, two copies of the system are created and evolve independently along each branch. The final result of the process are the configurations of the 2^K leaves of the tree.
The inference problem is the following. With the knowledge of the leaves configurations and of the type of evolution dynamics, is it possible to reconstruct the potential $\boldsymbol{C}^{-1}$ acting on the system? The idea for this is to

- Compute the probability of observing leaves configurations given the potential $\boldsymbol{C}^{-1}$, $P(leaves|\boldsymbol{C})$.
- Invert this relation using Bayes formula:

$$P(\boldsymbol{C}|leaves) = \frac{P(leaves|\boldsymbol{C})P(\boldsymbol{C})}{Z}$$

2.1 Dynamics

The Langevin equation for a N-dimensional particle with position $\vec{x} = (x_i), \quad i \in \{1 \dots N\}$ in a harmonic potential centered in $\vec{0}$ is

$$\gamma \frac{\mathrm{d}x_i}{\mathrm{d}t} = -\sum_j \lambda_{ij} x_j + \sqrt{2kT\gamma}\xi_i(t) \tag{4}$$

with $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t')$. The diffusion coefficient is $D = kT/\gamma$. The force on the particle is represented by the stiffness matrix λ_{ij}. It can be shown that the corresponding Fokker-Planck equation is

$$\gamma \partial_t P = \left(-\sum_{i,j} \frac{\partial}{\partial x_i} \lambda_{ij} x_j + kT \frac{\partial^2}{\partial x_i^2} \right) P \tag{5}$$

The stationnary state solution of this is

$$P_{eq}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}|}} \exp\left\{ -\frac{1}{2kT} \vec{x}^T \boldsymbol{C}^{-1} \vec{x} \right\} \tag{6}$$

with $\boldsymbol{C} = \lambda^{-1}$. In other words, the potential the particle evolves in is $V(x) = \frac{1}{2}\vec{x}^T\boldsymbol{C}^{-1}\vec{x} = \frac{1}{2}\sum_{i,j}\lambda_{ij}x_i x_j$. For our case, we can set $kT = 1$ for the rest.

Following [1], we can explicitely write the solution of 5. The result is an Ornstein-Uhlenbeck process defined in the following way

$$\begin{aligned} P(\vec{x}) =& \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}|}} \exp\left\{ -\frac{1}{2}\vec{x}^T \boldsymbol{C}^{-1} \vec{x} \right\}, \\ P(\vec{x}_2|\vec{x}_1, \Delta t) =& \frac{1}{\sqrt{(2\pi)^N (1-e^{-2\Delta t})|\boldsymbol{\Sigma}^{-1}|}} \exp\left\{ -\frac{1}{2}(\vec{x}_2 - \vec{\mu_1})^T \boldsymbol{\Sigma}^{-1} (\vec{x}_2 - \vec{\mu_1}) \right\}. \end{aligned} \tag{7}$$

where

$$\mu_1 = \boldsymbol{\Lambda}\vec{x}_1, \qquad \boldsymbol{\Sigma} = \boldsymbol{C} - \boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{\Lambda}, \qquad \boldsymbol{\Lambda} = e^{-\gamma \boldsymbol{C}^{-1}\Delta t}.$$

Thus, the average and covariance of variable $\vec{x}_1$ are time dependent through matrix $\boldsymbol{\Lambda}$, which itself depends on the equilibrium properties of the process through $\boldsymbol{C}$, and on the temporal dynamics through the dynamical parameter γ. It is important to notice that $\boldsymbol{\Lambda}$, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ all commute and are symetric.

Note: On dimensions. $\boldsymbol{C} \sim [T^2]$, $\boldsymbol{\Lambda} \sim [T^{-2}]$, $\gamma \sim [T^{-1}]$ and $kT \sim [L^2T^{-2}]$.

What is the probability of observing two configurations $\vec{x}_1$ and $\vec{x}_2$ of the system knowing that they are separated by time Δt? Using the identity $\boldsymbol{C}^{-1} = \boldsymbol{\Sigma}^{-1}(1 - \boldsymbol{\Lambda}^2)$, we find

$$\begin{aligned} \log P(\vec{x}_2, \vec{x}_1, \Delta t) &\propto -\frac{1}{2}\left[\vec{x}_2^T \boldsymbol{\Sigma}^{-1}\vec{x}_2 - 2\vec{x}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\vec{x}_1 + 2\vec{x}_1^T\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\vec{x}_1 + \vec{x}_1^T\boldsymbol{C}^{-1}\vec{x}_1\right] \\ &\propto -\frac{1}{2}\left[\vec{x}_2^T \boldsymbol{\Sigma}^{-1}\vec{x}_2 + \vec{x}_1^T\boldsymbol{\Sigma}^{-1}\vec{x}_1 - 2\vec{x}_1^T(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})\vec{x}_2\right], \end{aligned} \tag{8}$$

which is a gaussian distribution with a block correlation matrix, having $\boldsymbol{\Sigma}^{-1}$ on the diagonal blocks and $-\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}$ on the off-diagonal part. Inverting this matrix – like a 2×2 matrix, since everything commutes – yields the following expression for the covariance of configurations $\vec{x}_1$ and $\vec{x}_2$ separated by Δt:

$$\langle \vec{x}_1 \vec{x}_2^T \rangle = \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\Sigma}^2(1 - \boldsymbol{\Lambda}^2)^{-1} = \boldsymbol{\Lambda}\boldsymbol{C}. \tag{9}$$

2.2 Small trees

As an exercise, let us compute probability of the smallest non trivial tree, *ie* root $\vec{x}_0$ with children $\vec{x}_1$ and $\vec{x}_2$ and branch length Δt. The probability of observing given configurations on this topology is

$$\begin{aligned} P(\vec{x}_0, \vec{x}_1, \vec{x}_2; \Delta t) &= P(\vec{x}_1|\vec{x}_0)P(\vec{x}_2|\vec{x}_0)P(\vec{x}_0) \\ &\propto \exp -\frac{1}{2}\left\{\vec{x}_1\boldsymbol{\Sigma}^{-1}\vec{x}_1 + \vec{x}_2\boldsymbol{\Sigma}^{-1}\vec{x}_2 - 2(\vec{x}_1 + \vec{x}_2)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\vec{x}_0 + \vec{x}_0\boldsymbol{\Sigma}^{-1}(1 + \boldsymbol{\Lambda}^2)\vec{x}_0\right\}, \end{aligned} \tag{10}$$

where the identity $\boldsymbol{C}^{-1} = \boldsymbol{\Sigma}^{-1}(1 - \boldsymbol{\Lambda}^2)$ has been used. Integrating this over all values of $\vec{x}_0$ using eq. (11), and remembering that $\boldsymbol{\Sigma}(2\Delta t) = \boldsymbol{C} - \boldsymbol{\Lambda}^2\boldsymbol{C}\boldsymbol{\Lambda}^2$, we recover equation 8 with $\Delta t \to 2\Delta t$, that is with $\boldsymbol{\Lambda} \to \boldsymbol{\Lambda}^2$.

Note: Gaussian integration

$$\int \exp -\frac{1}{2}\left\{\vec{x}^T A\vec{x} + B^T\vec{x}\right\} \mathrm{d}^n x = \left(\frac{(2\pi)^n}{|A|}\right)^{1/2} \exp\left(\frac{1}{8}B^T A^{-1}B\right) \tag{11}$$

Let us do the same thing for a tree with two levels. Nodes are labelled from 0 to 6, with $\vec{x}_1$ and $\vec{x}_2$ being the children of root $\vec{x}_0$, and so on. The full probability can be written

$$P(\{\vec{x}_i\}_{i=0\ldots6}) = P(\vec{x}_3, \vec{x}_4|\vec{x}_1)P(\vec{x}_5, \vec{x}_6|\vec{x}_2)P(\vec{x}_1|\vec{x}_0)P(\vec{x}_2|\vec{x}_0)P_{eq}(\vec{x}_0) \tag{12}$$

Integrating this over $\vec{x}_0$ gives equation 8 for variables 1 and 2, with $\Delta t \to 2\Delta t$, while the part concering variables 3 to 6 remains untouched. Thus, we have to perform the following integration.

$$\begin{aligned} P(\{\vec{x}_i\}_{i=3\ldots6}) &= \int \mathrm{d}\vec{x}_1\mathrm{d}\vec{x}_2 P(\vec{x}_3, \vec{x}_4|\vec{x}_1)P(\vec{x}_5, \vec{x}_6|\vec{x}_2) \cdot P(\vec{x}_2, \vec{x}_1, 2\Delta t) \\ &\propto \int \mathrm{d}\vec{x}_1\mathrm{d}\vec{x}_2 \exp -\frac{1}{2}\left\{(\vec{x}_{3/4} - \boldsymbol{\Lambda}\vec{x}_1)\boldsymbol{C}^{-1}(1 - \boldsymbol{\Lambda}^2)^{-1}(\vec{x}_{3/4} - \boldsymbol{\Lambda}\vec{x}_1) + (\vec{x}_{5/6} - \boldsymbol{\Lambda}\vec{x}_2)\boldsymbol{C}^{-1}(1 - \boldsymbol{\Lambda}^2)^{-1}(\vec{x}_{5/6} - \boldsymbol{\Lambda}\vec{x}_2)\right. \\ &\qquad \left. + \vec{x}_{1/2}\boldsymbol{C}^{-1}(1 - \boldsymbol{\Lambda}^4)^{-1}\vec{x}_{1/2} - 2\vec{x}_1\boldsymbol{C}^{-1}\boldsymbol{\Lambda}^2(1 - \boldsymbol{\Lambda}^4)^{-1}\vec{x}_2\right\} \end{aligned} \tag{13}$$

where the notation $\vec{x}_{1/2}$ means that the corresponding term must be repeated for variables 1 and 2. For example, $\vec{x}_{1/2}\boldsymbol{U}\vec{x}_{1/2} = \vec{x}_1\boldsymbol{U}\vec{x}_1 + \vec{x}_2\boldsymbol{U}\vec{x}_2$. Here, we used the fact that when $\Delta t \to 2\Delta t$, $\boldsymbol{\Lambda} \to \boldsymbol{\Lambda}^2$. Since we have to integrate over 1 and 2, let us count linear and quadratic terms in those parameters to apply eq (11). For instance, $\vec{x}_1$:

- Quadratic: $2\boldsymbol{C}^{-1}\boldsymbol{\Lambda}^2(1 - \boldsymbol{\Lambda}^2)^{-1} + \boldsymbol{C}^{-1}(1 - \boldsymbol{\Lambda}^4)^{-1}$
- Linear: $-2\boldsymbol{C}^{-1}\boldsymbol{\Lambda}(1 - \boldsymbol{\Lambda}^2)(\vec{x}_3 + \vec{x}_4) - 2\boldsymbol{C}^{-1}\boldsymbol{\Lambda}^2(1 - \boldsymbol{\Lambda}^4)^{-1}\vec{x}_2$

From this point, it is quite clear that the result will be gaussian. The output of the integration will give quadratic and cross terms in $\vec{x}_{3/4/5/6}$, with a correlation matrix depending on $\boldsymbol{C}$ and $\boldsymbol{\Lambda}$. The exact expression has to be computed, though.

Let's try in another way. Since matrices $\boldsymbol{C}$ and $\boldsymbol{\Lambda}$ commute, we can express all the previous equations as scalars ones. Eq (13) can formally be expressed as

$$P(\{\vec{x}_i\}_{i=3\dots6}) \propto \int \mathrm{d}x_1 \mathrm{d}x_2 \exp -\frac{1}{2}\left\{\kappa\left(x_1^2+x_2^2\right)+2\nu x_1 x_2+2\alpha_1 x_1+2\alpha_2 x_2+c\right\} \tag{14}$$

where

$$\begin{aligned}
&\boldsymbol{C}\cdot\kappa = 2\boldsymbol{\Lambda}^2(1-\boldsymbol{\Lambda}^2)^{-1}+(1-\boldsymbol{\Lambda}^4)^{-1},\\
&\boldsymbol{C}\cdot\alpha = -\boldsymbol{\Lambda}(1-\boldsymbol{\Lambda}^2)^{-1}, \text{ with } \alpha_1=\alpha(x_3+x_4) \text{ and } \alpha_2=\alpha(x_5+x_6),\\
&\boldsymbol{C}\cdot\nu = -\boldsymbol{\Lambda}^2(1-\boldsymbol{\Lambda}^4)^{-1},\\
&\boldsymbol{C}\cdot c = (1-\boldsymbol{\Lambda}^2)^{-1}(x_3^2+x_4^2+x_5^2+x_6^2).
\end{aligned}$$

The integral can be computed easily, seeing that it is the integration of a vectorial gaussian variable (x_1, x_2) with correlation matrix having κ on the diagonal blocks and ν on the off-diagonal. Using (11), we get

$$\begin{aligned}
P(\{\vec{x}_i\}_{i=3\dots6}) \propto \exp -\frac{1}{2}\Big\{&(1-\boldsymbol{\Lambda}^2)(x_3^2+x_4^2+x_5^2+x_6^2)\\
&-\frac{1}{\kappa^2-\nu^2}\left(\kappa\alpha^2(x_3+x_4)^2+\kappa\alpha^2(x_5+x_6)^2+2\nu\alpha^2(x_3+x_4)(x_5+x_6)\right)\Big\},
\end{aligned} \tag{15}$$

where the $\boldsymbol{C}^{-1}$ in each quadratic term is omitted here for consision.
To obtain the correlation matrix caracterizing the leaves, one has to develop the above expression, and group quadratic terms, nearest neighbours terms x_3x_4 and x_5x_6, and further neighbours terms such as x_3x_5. The symetry of the equation makes clear the fact that the cross correlations should only depend on the distance of the corresponding pair in the tree.

2.3 General tree?

Let us consider a binary tree with $K+1$ levels, labelled $k \in \{0\dots K\}$. Nodes at level k are written $X^{\{k\}} = (X^i)_{i=1\dots2^k}$. In the following, we consider two levels of the tree, k and $k+1$, and introduce the following notation: nodes in level k are written $X^{\{k\}} = (X_0^i)_{i=1\dots2^k}$ and nodes in $k+1$ $X^{\{k+1\}} = \left((X_1^i, X_2^i)\right)_{i=1\dots2^k}$, where (X_1^i, X_2^i) are children of node X_0^i.
What we want to know is the probability of a configuration of nodes at level $k+1$, independently of the parent configurations. We assume that this quantity is known for level k, and we try to propagate it down the tree. Furthermore, we assume that $P(X^{\{k\}}) = P(X_0^1\dots X_0^{2^k})$ is a gaussian. In this case, the probability of observing a given configuration in level $k+1$ is

$$P(X^{\{k+1\}}) = \int \left(\prod_{i=1}^{2^k}\mathrm{d}X_0^i\right) P(X_0^1\dots X_0^{2^k}) \prod_{i=1}^{2^k} P(X_1^i, X_2^i|X_0^i). \tag{16}$$

Since $P(X_1^i, X_2^i|X_0^i)$ is a gaussian in all three variables, and $P(X_0^1\dots X_0^{2^k})$ is gaussian by assumption, the resulting distribution for nodes at level $k+1$ will also be a gaussian. And since the root of the tree is by assumption a gaussian variable with correlation matrix $\boldsymbol{C}$, the leaves of the tree are correlated gaussian variable by recursion. Quite clearly, the correlation matrix of the leaves is a combination of $\boldsymbol{C}^{-1}$ and of $\boldsymbol{\Lambda}$. Given the expression found for the 2-levels tree and the 3-levels tree (not finished, but it points in this direction), it is likely that this combination involves a product of $\boldsymbol{C}^{-1}$ with fraction

of polynomials of $\mathbf{\Lambda}$. Branch length and number of children do not change this, but only the nature of $\mathbf{\Lambda}$.
Thus, what would be analytically useful is to propagate the correlation matrix through the tree by finding some recursive equation. In other words, given correlation matrix Ξ_k for level k, find Ξ_{k+1} as a function of Ξ_k and $\mathbf{\Lambda}$ by performing the integration in eq (16). Solving this numerically would then give us the likelihood of the data given parameters $\boldsymbol{C}$ in the form of a gaussian function. This could then be inverted by Bayes theorem to obtain an estimator of $\boldsymbol{C}$.

Idea on how to do that. $P(X_0^1 \dots X_0^{2^k})$ is a gaussian, and we can write 16 in the following form

$$P(X^{\{k+1\}}) \propto \int \left(\prod_{i=1}^{2^k} \mathrm{d}X_0^i \right) \exp -\frac{1}{2} \left\{ \kappa^{(k)} \sum_{i=1}^{2^k} (X_0^i)^2 + 2 \sum_{i=1}^{2^k} \sum_{\langle ij \rangle = d} \nu_d^k X_0^i X_0^j + 2 \sum_{i=1}^{2^k} \alpha_i^{(k)} X_0^i + c \right\}, \tag{17}$$

where notation is inspired by that of 14. κ^k represents the diagonal block of the correlation matrix of $P(X_0^1 \dots X_0^{2^k})$ and $\nu_d^{(k)}$ represents the off-diagonal blocks. The subscript (k) is here to remind that these coefficients are specific to level k of the tree. Index d represents the distance $\langle ij \rangle$ in the tree of two nodes i and j: it is the number of levels one should go up in the tree to find a common ancestor to i and j. The assumption here is that the correlation between i and j only depends on this quantity, in the sense that all nodes distant of d have the same correlation block matrix ν_d. Coefficients α_i can be written $\alpha(X_1^i + X_1^j)$. Finally, c is just the sum of $(X_1^i)^2 + (X_1^j)^2$ with a factor $(1 - \mathbf{\Lambda}^2)^{-1} \boldsymbol{C}^{-1}$.

3 Message Passing

Belief propagation is an exact algorithm for finding marginal distributions on trees. The interactions among our variables (circular nodes) are represented in Fig. 2 by squares nodes. In the Ornstein-Uhlenbeck process described in the previous section, both the equilibrium distribution P_{eq} and the propagator $G(\Delta t)$ at time Δt are gaussians. The observed nodes are also gaussians, since a delta distribution is a zero variance gaussian. Then we hope to solve all message passing computations within gaussian integrals.

Following standard Belief Propagation method [], the exact free energy (likelihood) of the model can be written in terms of two variables joint distributions $b(x_1, x_2)$, when x_1 and x_2 are first neighbors in the graph, and one variable distributions $b(x)$, as

$$\begin{aligned} F[\{b(\cdot)\}] \quad = \quad & \sum_{(i-j)} \sum_{x_i, x_j} (E_{i,j}(x_i, x_j) b_{i,j}(x_i, x_j) + b_{i,j}(x_i, x_j) \log b_{i,j}(x_i, x_j)) \\ & -2 \sum_i \sum_{x_i, x_j} (E_i(x_i) b_i(x_i) + b_i(x_i) \log b_i(x_i)) \end{aligned}$$

The consistency between both types of distributions, when they share the same variable implies

$$b(x_1) = \sum_{x_2} b(x_1, x_2)$$

and is guaranteed by a set of lagrange multipliers $m_{2 \to 1}(x_1)$ for every posible marginalization constraint.

After introducing the messages in the free energy, we get a Lagrangian, and extremization in terms

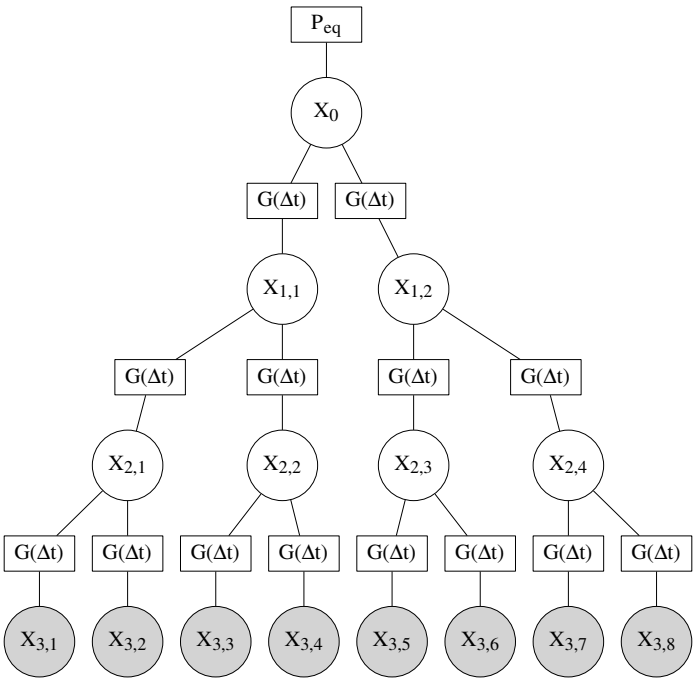

Figure 2: Phylogenetic model for genetic drifting

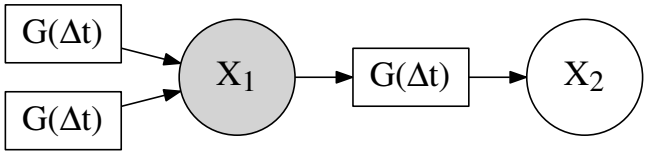

Figure 3: Message passing step

of beliefs give the relation

$$b_{x_1,x_2} = \frac{1}{Z_{1,2}} G_{\Delta t}(x_1,x_2) \prod_{j\in\partial 1\setminus 2} m_{j\to 1}(x_1) \prod_{j\in\partial 2\setminus 1} m_{j\to 2}(x_2) \tag{18}$$

$$b_{x_1} = \frac{1}{Z_1} \prod_{j\in\partial 1} m_{j\to 1}(x_1) \tag{19}$$

Replacing this relations in the free energy, we get a "variational" free energy in terms of the messages that appear in the normalization constants:

$$F[\{m_{a\to b}\}] = \sum_{(i-j)} Z_{i,j}[\{m\}] - 2\sum_i Z_i[\{m\}]$$

Variational here means that the true values of $m_{i\to j}$ are the one solving

$$\frac{\partial F[\{m\}]}{\partial m} = 0$$

resulting in

$$m_{1\to 2}(x_2) = \int \mathrm{d}x_1 G_{\Delta t}(x_1, x_2) m_{i,\to 1}(x_1) m_{j,\to 1}(x_1)$$

Assuming that messages are gaussians (not necessarily normalized)

$$m(x) = \exp\left(-\frac{1}{2}\vec{x}^T A\vec{x} + \vec{B}\cdot x\right)$$

and the propagator

$$G_{\Delta t}(x_1, x_2) = \exp\left(-\frac{1}{2}[x_1^T\Sigma^{-1}x_1 + x_2^T\Sigma^{-1}x_2 - 2x_1^T(\Lambda\Sigma^{-1})x_2]\right)$$

then the message passing step is only the following update

$$\begin{aligned}
m_{1\to 2}(x_2) &= \int \mathrm{d}x_1 \exp\left(-\frac{1}{2}[x_1^T\Sigma^{-1}x_1 + x_2^T\Sigma^{-1}x_2 - 2x_1^T(\Lambda\Sigma^{-1})x_2] - \frac{1}{2}x_1^T A x_1 + \vec{B}^T x_1\right) \\
&= \exp\left(-\frac{1}{2}x_2^T\Sigma^{-1}x_2\right)\int \mathrm{d}x_1 \exp\left(-\frac{1}{2}x_1^T(\Sigma^{-1}+A)x_1 + x_1^T(B + (\Lambda\Sigma^{-1})x_2)\right) \\
&= \exp\left(-\frac{1}{2}x_2^T\Sigma^{-1}x_2\right)\exp\ -\frac{1}{2}\left((B + (\Lambda\Sigma^{-1})x_2)^T(\Sigma^{-1}+A)^{-1}(B + (\Lambda\Sigma^{-1})x_2)\right) \\
&= \exp\left(-\frac{1}{2}x_2^T\left[\Sigma^{-1} + (\Lambda\Sigma^{-1})^T(\Sigma^{-1}+A)^{-1}(\Lambda\Sigma^{-1})\right]x_2 - B^T(\Sigma^{-1}+A)^{-1}(\Lambda\Sigma^{-1})x_2\right)
\end{aligned}$$

Calling $A = A_{i\to 1} + A_{j\to 1}$ y $B = B_{i\to 1} + B_{j\to 1}$, the message update rule is

$$\begin{aligned}
A_{1\to 2} &= \Sigma^{-1} + (\Lambda\Sigma^{-1})^T(\Sigma^{-1}+A)^{-1}(\Lambda\Sigma^{-1}) \\
B_{1\to 2} &= B^T(\Sigma^{-1}+A)^{-1}(\Lambda\Sigma^{-1})
\end{aligned}$$

Such update step has to be implemented $3*N$, where N is the number of free variables in the tree (not considering the observed leaves). Since the graph is a tree, there is no need for a fixed point iteration. If the procedure is carried from the leaves inwards, it will solve all messages in a single swept.

3.1 Inverse problem: learning C

Based on the exactness of the Bethe Free Energy for a tree, then

$$F[\{m\}] = -\log Z$$

where

$$Z[x^k] = \int_{x_0,x_1,x_2,\ldots} \Psi(x_0)G(x_0,x_1)G(x_0,x_2)G(x_1,x_3)G(x_1,x_4)\ldots G\ldots$$

is the partition function but also the likelihood of the data at level k. For this reason, $F[\{m\}]$ is the log-likelihood, and a model can be learned by maximizing it with respect to parameters.

4 Maximum likelihood

4.1 Expression of the likelihood

The full system, *i.e.* all nodes in the tree, will be noted $\vec{X} = \{X_i\}_{i=1\dots N}$. Each X_i is a gaussian vector. X_0 is the root of the tree. We decompose the nodes in two groups, leaves and internal nodes: $\vec{X} = \{\vec{X}_h, \vec{X}_d\}$, with $\vec{X}_h$ being the internal nodes and $\vec{X}_d$ being the leaves.
The assumption on the distribution of $\vec{X}$ is that it follows the Ornstein-Uhlenbeck process, that is

$$\begin{aligned} P^0(X_i) &\propto \exp -\frac{1}{2}\left\{X_i \boldsymbol{C}^{-1} X_i\right\} \\ P^0(X_j|X_i;\Delta t) &\propto \exp -\frac{1}{2}\left\{X_j \boldsymbol{\Sigma}^{-1} X_j + X_i \boldsymbol{\Lambda}^2 \boldsymbol{\Sigma}^{-1} X_i - 2X_i \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} X_j\right\} \end{aligned} \tag{20}$$

The probability of a configuration $\vec{X}$ of the full system can be written as

$$P^0(\vec{X}) = P^0(X_0)P^0(X_1|X_0)P^0(X_2|X_0)P^0(X_3|X_1)P^0(X_4|X_1)\dots$$

assuming the above mentionned pairwise tree topology. Expanding this expression, it is easy to see that probability P^0 can be written as

$$P^0(\vec{X}) = \frac{1}{Z^0}\exp\left\{\sum_{i<j} X_i J_{ij} X_j + \sum_{i=1}^{N} X_i H_i X_i\right\} \tag{21}$$

where

$$J_{ij} = \begin{cases} \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} & \text{if } i \text{ and } j \text{ are in contact,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$H_i = -\frac{1}{2}\cdot\begin{cases} (1+\boldsymbol{\Lambda}^2)\boldsymbol{\Sigma}^{-1} & \text{if } i = 0, \\ \boldsymbol{\Sigma}^{-1} & \text{if } i \text{ is a leaf,} \\ (1+2\boldsymbol{\Lambda}^2)\boldsymbol{\Sigma}^{-1} & \text{otherwise.} \end{cases}$$

In other words, the system is described by a pairwise hamiltonian

$$\mathcal{H}^0 = -\sum_{i<j} X_i J_{ij} X_j - \sum_{i=1}^{N} X_i H_i X_i. \tag{22}$$

It is important to note that thanks to the properties of the OU process, one can write the single and pairwise marginals of P^0. Indeed, using 20 immediatly gives

$$\begin{aligned} P_i^0(X_i) &\propto \exp -\frac{1}{2}\left\{X_i \boldsymbol{C}^{-1} X_i\right\} \\ P_{ij}^0(X_i, X_j;\Delta t) &\propto \exp -\frac{1}{2}\left\{X_i \boldsymbol{\Sigma}^{-1} X_i + X_j \boldsymbol{\Sigma}^{-1} X_j - 2X_i \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1} X_j\right\} \end{aligned} \tag{23}$$

This, put together with the fact that the topology of interaction of J_{ij} is a tree, means we can write

the free energy exactly using Bethe's formula

$$\begin{aligned}
-\log Z^0 =& \langle H^0 \rangle_{P^0} - S(P^0) \\
=& \sum_{i<j} \int_{X_i, X_j} P^0_{ij}(X_i, X_j) \left[\log P^0_{ij}(X_i, X_j) - X_i J_{ij} X_j \right] \\
&+ \sum_{i=1}^{N} \int_{X_i} P^0_i(X_i) \left[-(r_i - 1) \log P^0_i(X_i) - X_i H_i X_i \right] \\
=& \sum_{i<j} \int_{X_i, X_j} P^0_{ij}(X_i, X_j) \left[\log P^0_{ij}(X_i, X_j) + E_{ij}(X_i, X_j) \right] \\
&- \sum_{i=1}^{N} (r_i - 1) \int_{X_i} P^0_i(X_i) \left[\log P^0_i(X_i) + E_i(X_i) \right],
\end{aligned} \tag{24}$$

by defining $E_{ij}(X_i, X_j) = -X_i J_{ij} X_j + X_i H_i X_i + X_j H_j X_j$ and $E_i(X_i) = -X_i H_i X_i$.

Let us now write the probability of observing a certain configuration $\vec{X_d}$ of the leaves. The likelihood is

$$\begin{aligned}
P(\vec{X_d}|\boldsymbol{C}) =& \log \int_{\vec{X_h}} P^0(\vec{X}) \\
=& -\log Z^0 + \log \int_{\vec{X_h}} e^{-\mathcal{H}^0(\vec{X_h}, \vec{X_d})}.
\end{aligned} \tag{25}$$

The first member of the right hand side is the free energy associated to $\mathcal{H}^0$, described by equation 24. The second member is the sum of Boltzmann weights over a sub ensemble of the system, with leaves fixed to a given configuration. This can be seen as the partition function of a second hamiltonian $\mathcal{H}^d(\vec{X_h}) = \mathcal{H}^0(\vec{X_h}, \vec{X_d})$. Analytical expression of $\mathcal{H}^d$ can easily be written

$$\mathcal{H}^d(\vec{X_h}) = -\sum_{1 \leq i < j \leq N_h} X_i J_{ij} X_j - \sum_{i=1}^{N_h} (X_i H_i X_i + \mu_i X_i) + cte, \tag{26}$$

where the fields and couplings $\{J_{ij}, H_i\}$ are equal to those of $\mathcal{H}^0$, and

$$\mu_i = \begin{cases} (X_i^{(1)} + X_i^{(2)}) \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} & \text{if } i \text{ is in contact with two leaves } X_i^{(1)} \text{ and } X_i^{(2)}, \\ 0 & \text{otherwise.} \end{cases}$$

The constant *cte* depends only on the data, and can be written as

$$cte = -\sum_{i=1}^{N_d} X_i H_i X_i. \tag{27}$$

As a consequence, the likelihood can be written as

$$P(\vec{X_d}|\boldsymbol{C}) = -\frac{1}{2} \sum_{i \in leaves} X_i \boldsymbol{\Sigma}^{-1} X_i + \log Z^d - \log Z^0. \tag{28}$$

The second member of this equation can be computed using the analytical formula 24. Since the hamiltonian $\mathcal{H}^d$ is also pairwise, and with a tree-like topology of interaction, we can also write its free energy using the Bethe approximation. However, in this case, we do not know the pairwise marginals $P^d_{ij}(X_i, X_j)$. Thus, we will use message passing to compute those and write the free energy.

4.2 Gradient of Z^0

The free energy of $\mathcal{H}^0$ can be expressed as a sum of pairwise free energies F^0_{ij} and single site free energies F^0_i. Expanding equation 24 using definitions of couplings J_{ij} and fields H_i as well as the characteristics of the Ornstein-Uhlenbeck process, one finds

$$F^0 = -\sum_{i<j} \log Z^0_{ij} + \sum_{i=1}^{N} \log Z^0_i, \tag{29}$$

where

$$\begin{aligned} Z^0_i &= \int \mathrm{d}X_i \exp -\frac{1}{2} X_i \boldsymbol{C}^{-1} X_i \\ Z^0_{ij} &= \int \mathrm{d}X_i \mathrm{d}X_j \exp -\frac{1}{2} \left\{ X_i \boldsymbol{\Sigma}^{-1} X_i + X_j \boldsymbol{\Sigma}^{-1} X_j - 2 X_i \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} X_j \right\}. \end{aligned} \tag{30}$$

4.3 Message passing

Message from i to j, in the case that i is a leaf of the tree

$$\begin{aligned} m_{ij}(X_j) &= \int \mathrm{d}X_i e^{X_i H_i X_i + (X_j J_{ij} + \mu_i) X_i} \prod_{k \in \partial i \setminus j} m_{ki}(X_i) \\ &= \left(\frac{\pi^L}{|H_i|} \right)^{1/2} e^{-\frac{1}{4}(X_j J_{ij} + \mu_i) H_i^{-1} (J_{ij} X_j + \mu_i)}. \end{aligned} \tag{31}$$

From this, it is clear that messages will always be a gaussian function of their variable, since integrating products of gaussian functions always yields a new gaussian function. Thus, we can write messages in the following way

$$m_{ij}(X_j) \propto e^{X_j \Phi_{ij} X_j + \Psi_{ij} X_j} \tag{32}$$

Plugging this into the message defining equation, we obtain

$$\begin{aligned} m_{ij}(X_j) &= \int \mathrm{d}X_i e^{X_i H_i X_i + (X_j J_{ij} + \mu_i) X_i} \prod_{k \in \partial i \setminus j} m_{ki}(X_i) \\ &\propto \int \mathrm{d}X_i \exp \left\{ \left(X_j J_{ij} + \mu_i + \sum_{k \in \partial i \setminus j} \Psi_{ki} \right) X_i + X_i \left(H_i + \sum_{k \in \partial i \setminus j} \Phi_{ki} \right) X_i \right\} \end{aligned} \tag{33}$$

and thus arriving to the following recurrence relation after performing the integration

$$\begin{aligned} \Phi_{ij} &= -\frac{1}{4} J_{ij} \left(H_i + \sum_{k \in \partial i \setminus j} \Phi_{ki} \right)^{-1} J_{ij} \\ \Psi_{ij} &= -\frac{1}{2} J_{ij} \left(H_i + \sum_{k \in \partial i \setminus j} \Phi_{ki} \right)^{-1} \left(\mu_i + \sum_{k \in \partial i \setminus j} \Psi_{ki} \right) \end{aligned} \tag{34}$$

5 Annex

5.1 Gaussian integration

$$Z_{12} = \int \mathrm{d}X_1 \mathrm{d}X_2 \exp \left\{ X_1 A_{11} X_1 + X_2 A_{22} X_2 + X_1 A_{12} X_2 + B_1 X_1 + B_2 X_2 \right\}$$

can be written as

$$\begin{aligned} Z_{12} &= \int \mathrm{d}\boldsymbol{Y} \exp -\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{m})\boldsymbol{K}(\boldsymbol{Y}-\boldsymbol{m}) \cdot \exp\left(2\boldsymbol{m}\boldsymbol{K}\boldsymbol{m}\right) \\ &= \left(\frac{(2\pi)^{2L}}{|\boldsymbol{K}|}\right) \cdot \exp\left(2\boldsymbol{m}\boldsymbol{K}\boldsymbol{m}\right) \end{aligned} \tag{35}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \boldsymbol{K} = -\begin{pmatrix} 2A_{11} & A_{12} \\ A_{12} & 2A_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{m} = \begin{pmatrix} B_2A_{12} - 2B_1A_{22} \\ B_1A_{12} - 2B_2A_{11} \end{pmatrix} .* (4A_{22}A_{11} - A_{12}^2)^{-1}$$

References

[1] Rajesh Singh, Dipanjan Ghosh, and R. Adhikari. Fast bayesian inference of the multivariate ornstein-uhlenbeck process. *arxiv:1706.04961*, 2017.