Accelerated inference in a complex phylogenetic model

Exact inference with massive systems of ODEs

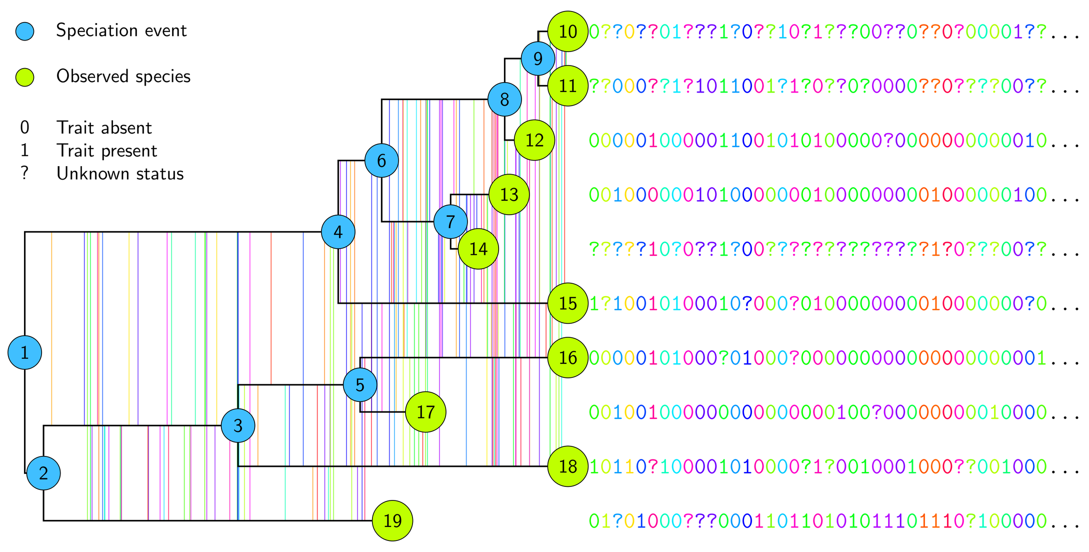
Luke Kelly and Geoff Nicholls

University of Oxford, Department of Statistics

Motivation

Lateral trait transfer is a form of evolutionary activity whereby traits pass through non-ancestral relationships between contemporary species.

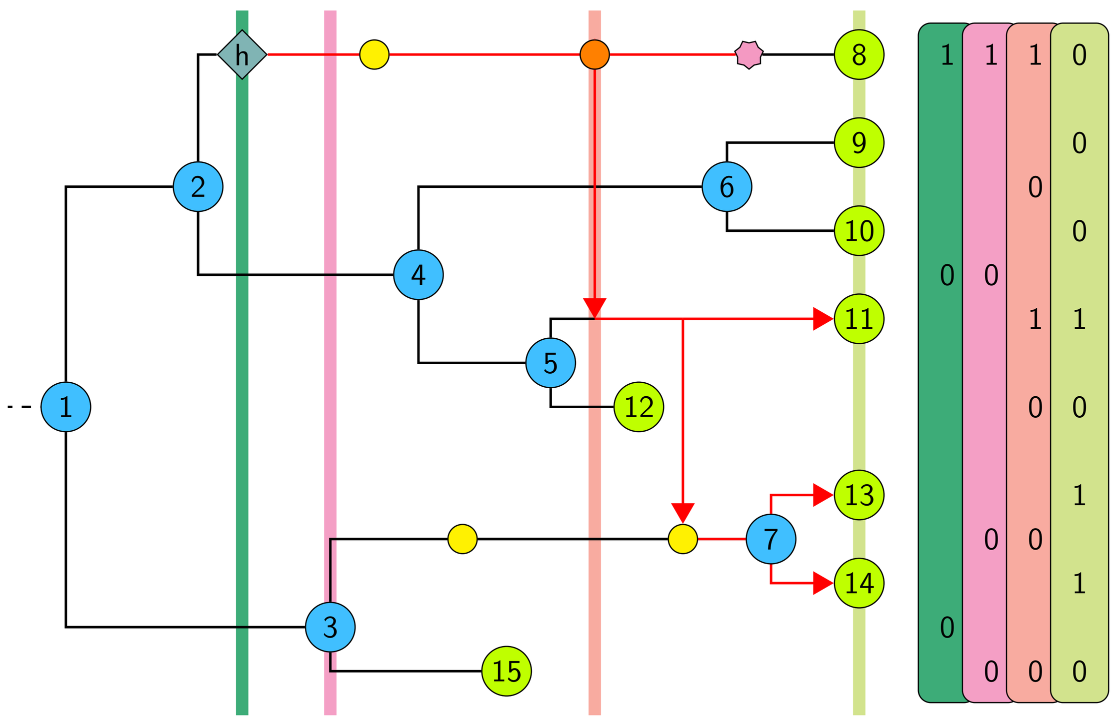
Although tree-like, the histories of transferred traits conflict with the overall phylogeny so models based solely on ancestral inheritance are misspecified here.



Process

A branching process on sets of traits is the phylogeny of the observed taxa.

- Species evolve new traits at rate λ.
- Trait instances die independently at rate μ and transfer to other species at per capita rate β.



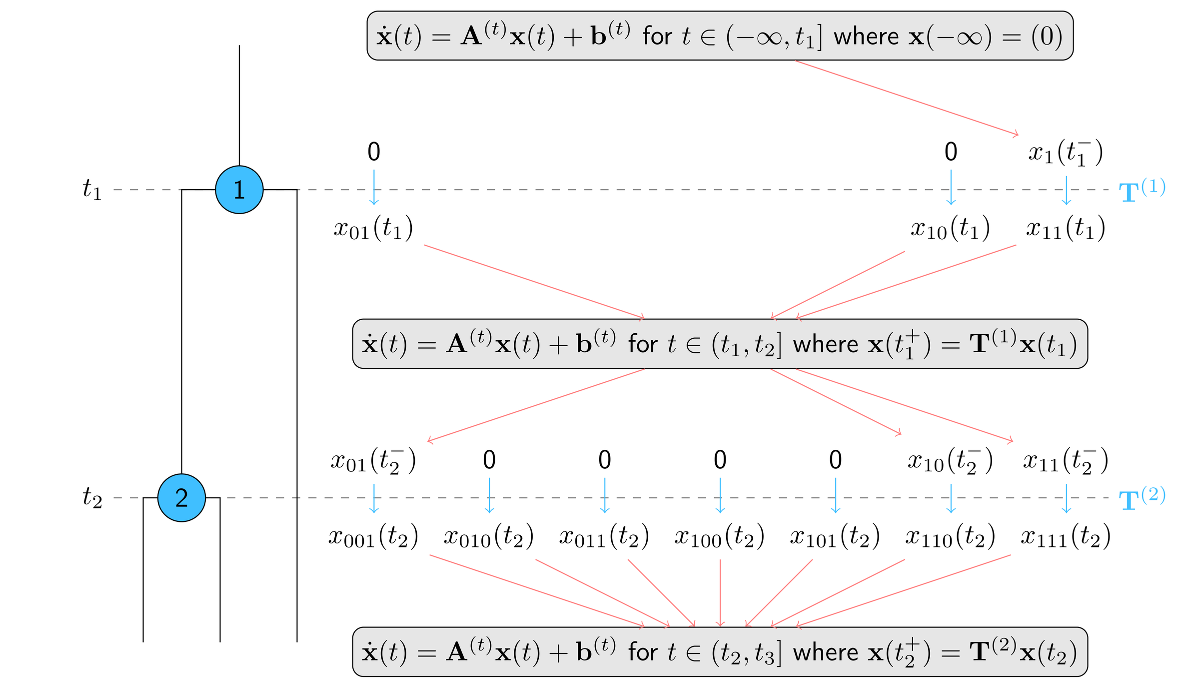
A phylogenetic tree and trait history drawn from our process with snapshots of the corresponding pattern process. Catastrophe nodes represent spikes in activity.

A trait h displays a pattern $\mathbf{p}^h(t) \in \mathcal{P}^{(t)} = \{0,1\}^{L^{(t)}} \setminus \{\mathbf{0}\}$ of presence or absence across $L^{(t)}$ species at time t. Traits are exchangeable so we model the terminal pattern frequencies $\mathbf{N}(T) = (N_{\mathbf{p}}(T))_{\mathbf{p}\in\mathcal{P}^{(T)}}$.

Inference

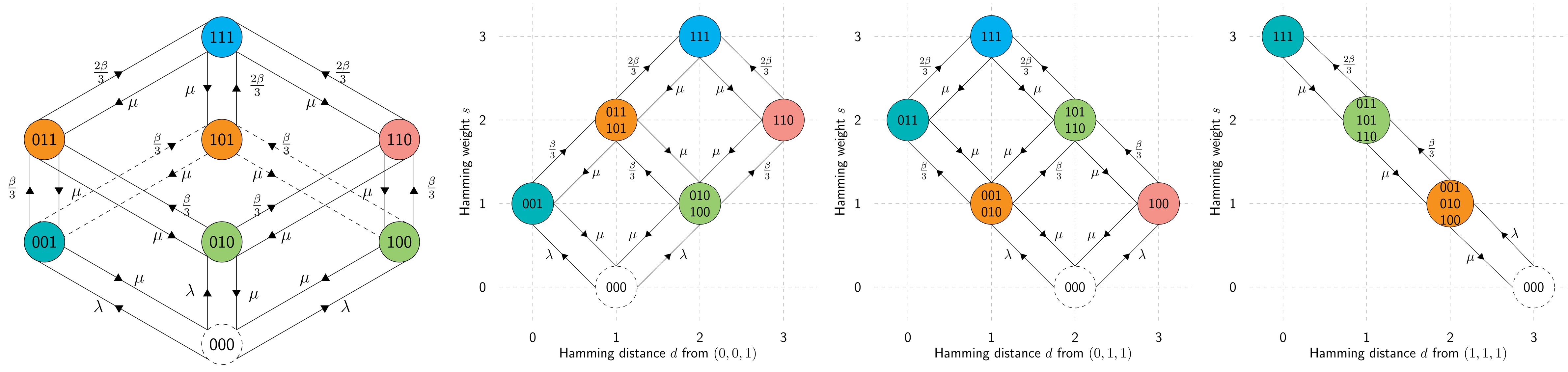
After integrating out the birth rate λ and unobserved trait histories, $\mathbf{N}(T)$ is multinomial with unnormalised weights $\mathbf{x}(T) = (x_{\mathbf{p}}(T))_{\mathbf{p}\in\mathcal{P}^{(T)}}$, the solution of a sequence of initial value problems across the tree.

The ODEs have dimension $\mathcal{O}(2^{L^{(t)}})$ so exact inference with an ODE solver quickly becomes intractable as $L^{(t)}$ increases.



We want to perform MCMC so develop a fast method for computing the parameters $\mathbf{x}(T)$.

Symmetries and equivalence classes



For $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L^{(t)})}$ and $\mathbf{z}$ solving IVPs on $\mathcal{O}(L^{(t)2})$ equivalence classes, we have

$$x_{\mathbf{p}}(t+\Delta) = \sum_{\mathbf{q}\in\mathcal{P}^{(t)}} y^{s(\mathbf{q})}_{s(\mathbf{p}),d(\mathbf{p},\mathbf{q})}(\Delta)\, x_{\mathbf{q}}(t) + z_{\mathbf{p}}(\Delta), \qquad \left\{\begin{array}{ll}\text{Hamming weight} & s\\ \text{Hamming distance} & d\end{array}\right\}.$$

The computational cost of this exact approach is $\mathcal{O}(2^{2L^{(t)}})$ as we form $\exp(\mathbf{A}^{(t)}\Delta)$ explicitly.

Sparse estimator $\mathbf{G}(\Delta)$ of $\exp(\mathbf{A}^{(t)}\Delta)$,

$$G_{\mathbf{p},\mathbf{q}}(\Delta) = \begin{cases} y^{s(\mathbf{q})}_{s(\mathbf{q}),d(\mathbf{p},\mathbf{q})}(\Delta), & d(\mathbf{p},\mathbf{q}) \le 1,\\ 0, & \text{otherwise,}\end{cases}$$

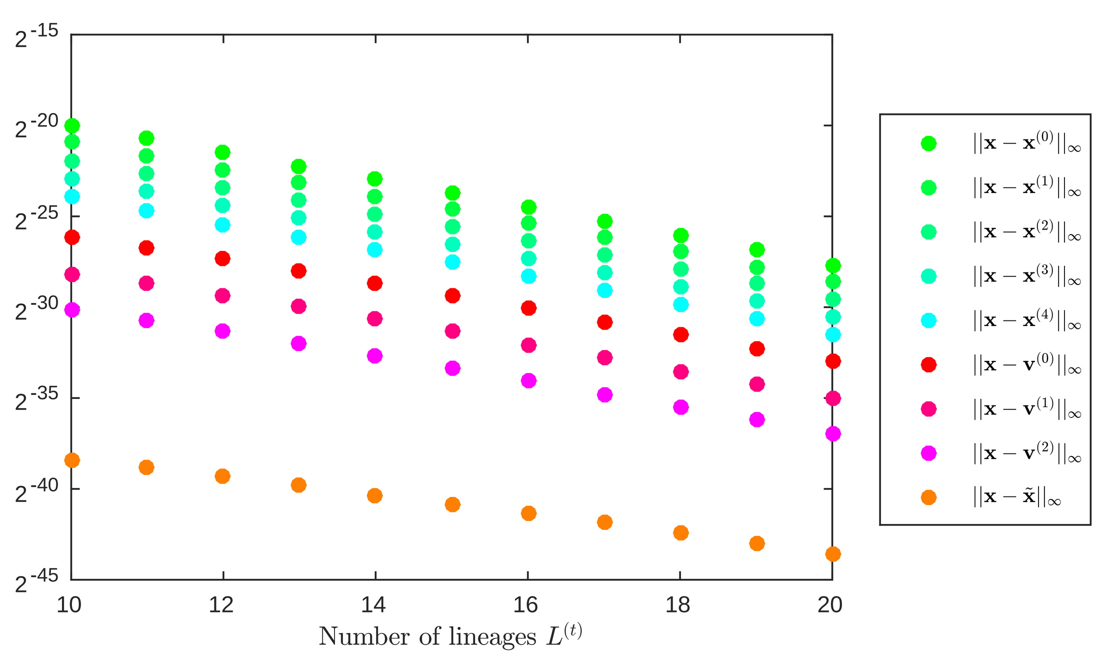
and construct $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots \to \mathbf{x}$ by

$$\mathbf{x}^{(k)}(t+\Delta) = \mathbf{G}(\Delta 2^{-k})^{2^k}\mathbf{x}(t) + \mathbf{Z}(\Delta).$$

Linear convergence: $\mathbf{x} - \mathbf{x}^{(k)} = \mathcal{O}(2^{-k})\mathbf{x}$.

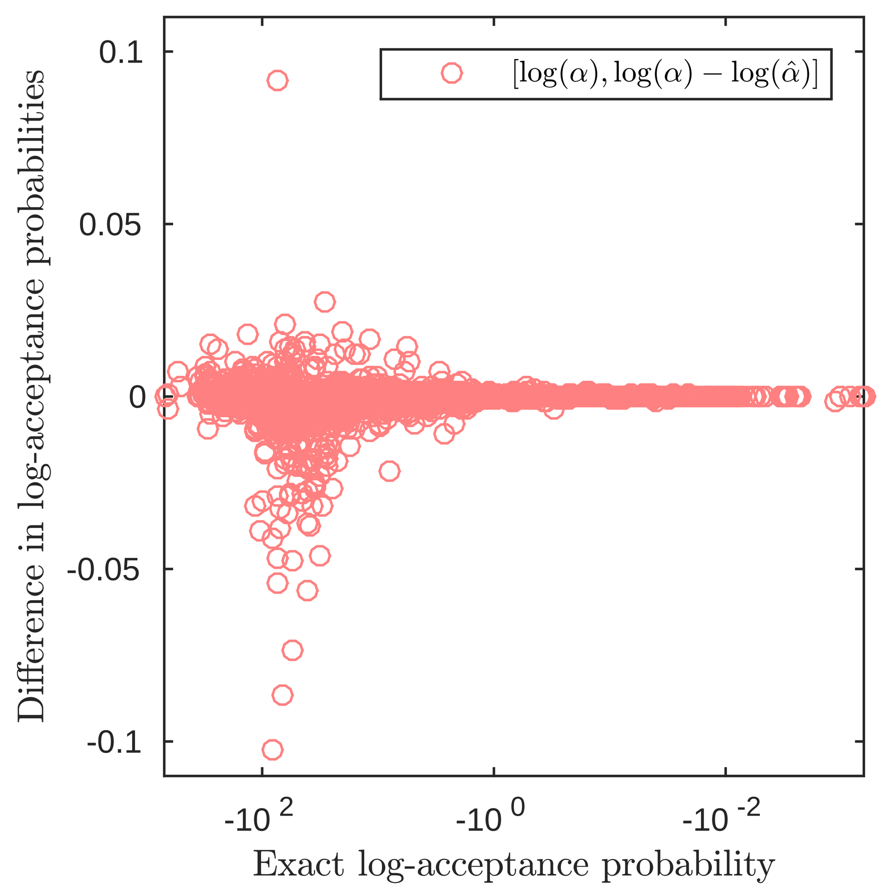
Acceleration scheme

Jennings' transformation, a stable, non-linear extrapolation for vector sequences, significantly reduces the error in our estimates with negligible computational cost.



$\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)} \xrightarrow{\text{Jennings}} \mathbf{v}^{(0)}, \mathbf{v}^{(1)}, \mathbf{v}^{(2)} \xrightarrow{\text{Jennings}} \tilde{\mathbf{x}}$.

Construct an unbiased likelihood estimator and run pseudo-marginal MCMC.



- Our accelerated inference scheme is exact in a MCMC sense.
- The effective sample size per unit time is an order of magnitude higher than computing parameters with a standard ODE solver and running the Metropolis–Hastings algorithm.

References

A. Jennings. *J. Inst. Maths. Applics.*, 1971.

L.J. Kelly. PhD thesis, University of Oxford, 2016.

L.J. Kelly and G.K. Nicholls. *Ann. Appl. Stat.*, 2017.

Acknowledgements

St John's College and the EPSRC.