Bayesian phylogenetic inference for big data

Claudia Solís-Lemus, PhD
Emory University

August 16, 2019

THE FOLLOWING **PREVIEW** HAS BEEN APPROVED FOR

ALL AUDIENCES

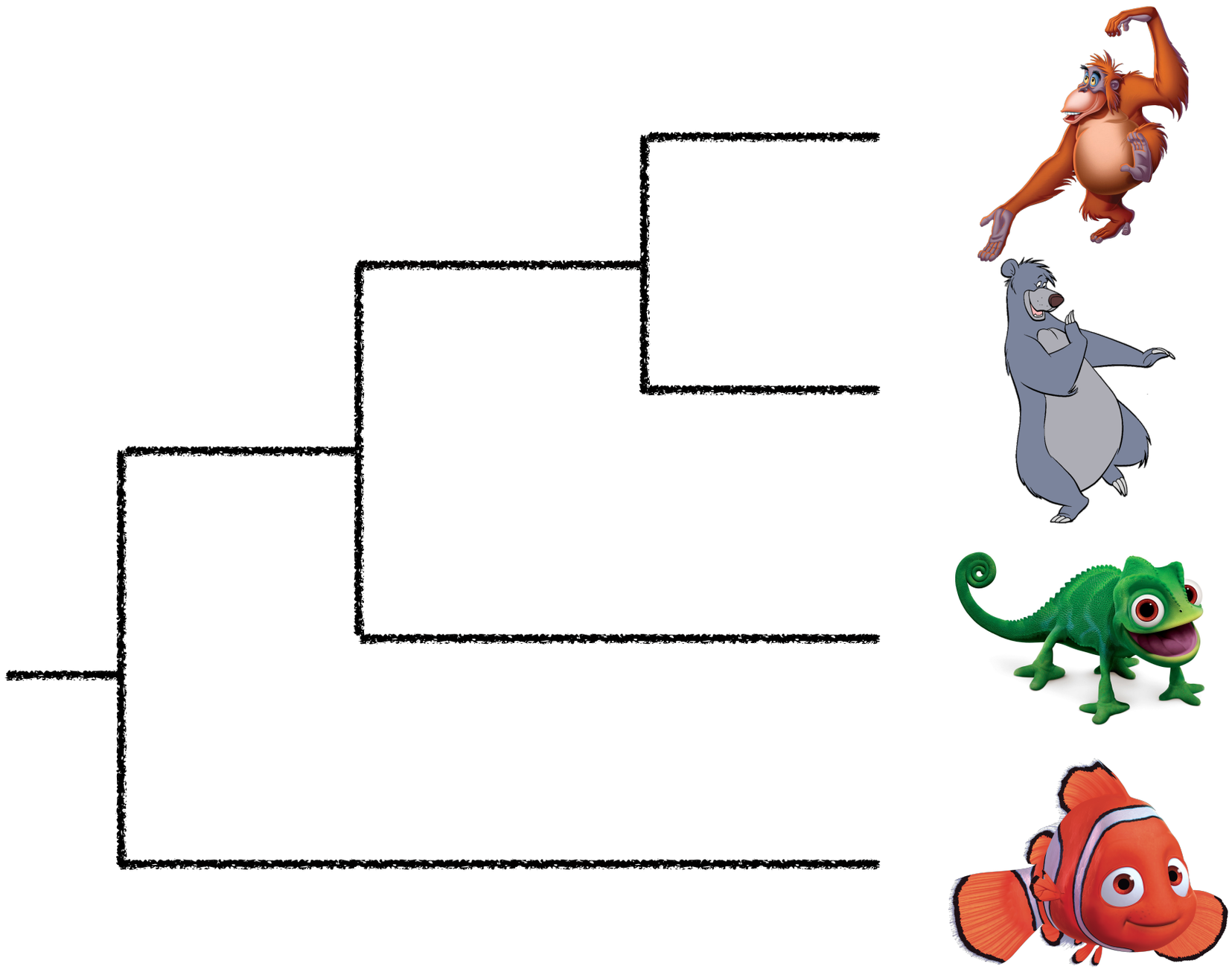
BY THE MOTION PICTURE ASSOCIATION OF AMERICA INC.

THE FILM ADVERTISED HAS BEEN RATED

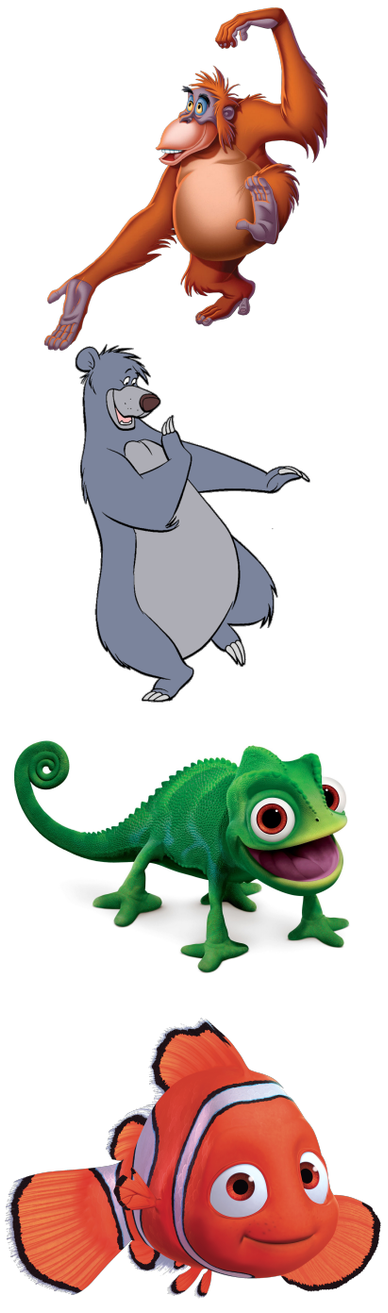
R	RESTRICTED
	UNDER 17 REQUIRES ACCOMPANYING PARENT OR GUARDIAN
Mathematical models and formulas	

www.filmratings.com www.mpaa.org

Phylogenetic inference



Phylogenetic inference



Phylogenetic inference

AAGTCTAG

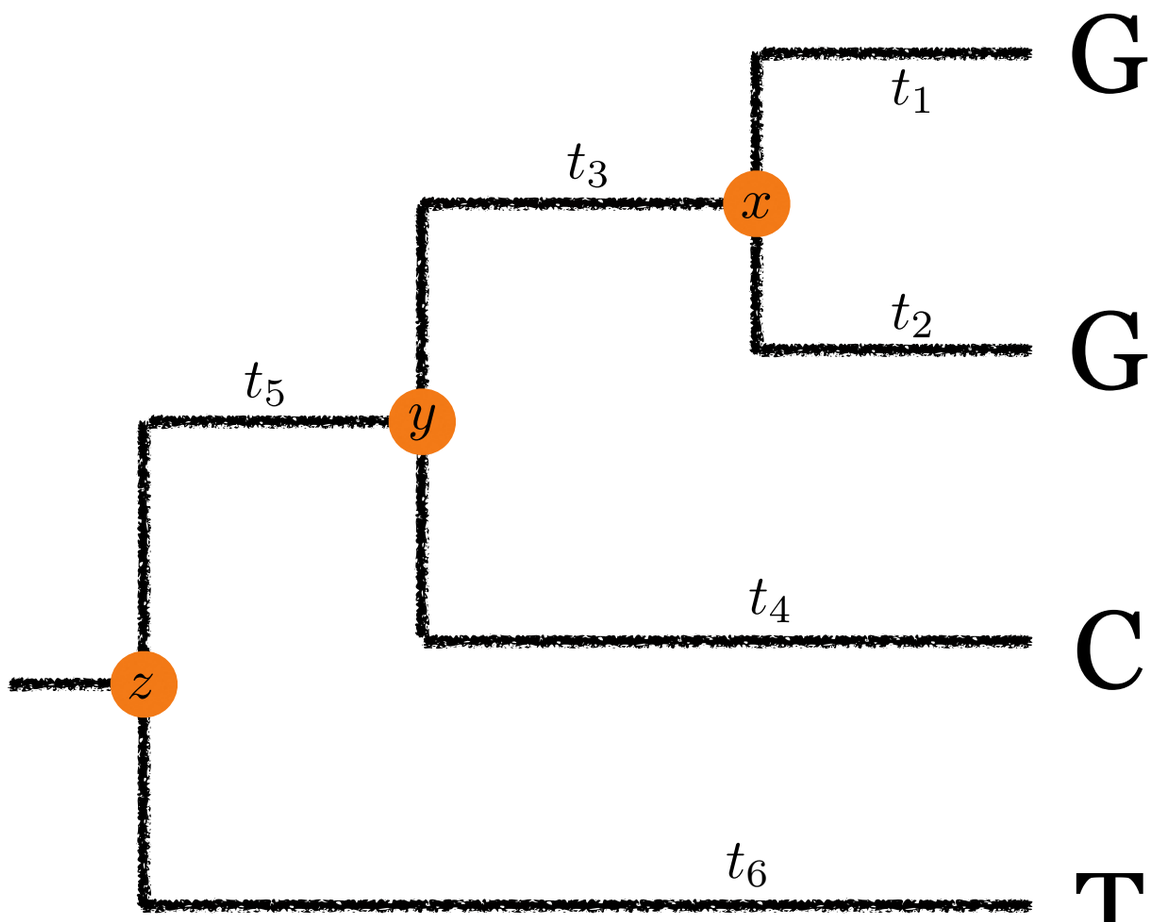
AAGTCTAG

AACTCTAG

AATTCTAG

Phylogenetic inference

$$\mathbf{Q} = \begin{array}{c} \begin{matrix} A & C & G & T \end{matrix} \\ \begin{bmatrix} * & a\pi_C & b\pi_G & c\pi_T \\ a\pi_A & * & d\pi_G & e\pi_T \\ b\pi_A & d\pi_C & * & f\pi_T \\ c\pi_A & e\pi_C & f\pi_G & * \end{bmatrix} \end{array} \begin{matrix} A \\ C \\ G \\ T \end{matrix}$$



$$L = \sum_z \sum_y \sum_x \pi(z) P_{t_6}(z,T) P_{t_5}(z,y) P_{t_4}(y,C) P_{t_3}(y,x) P_{t_2}(x,G) P_{t_1}(x,G)$$

Likelihood model: continuous-time Markov chain

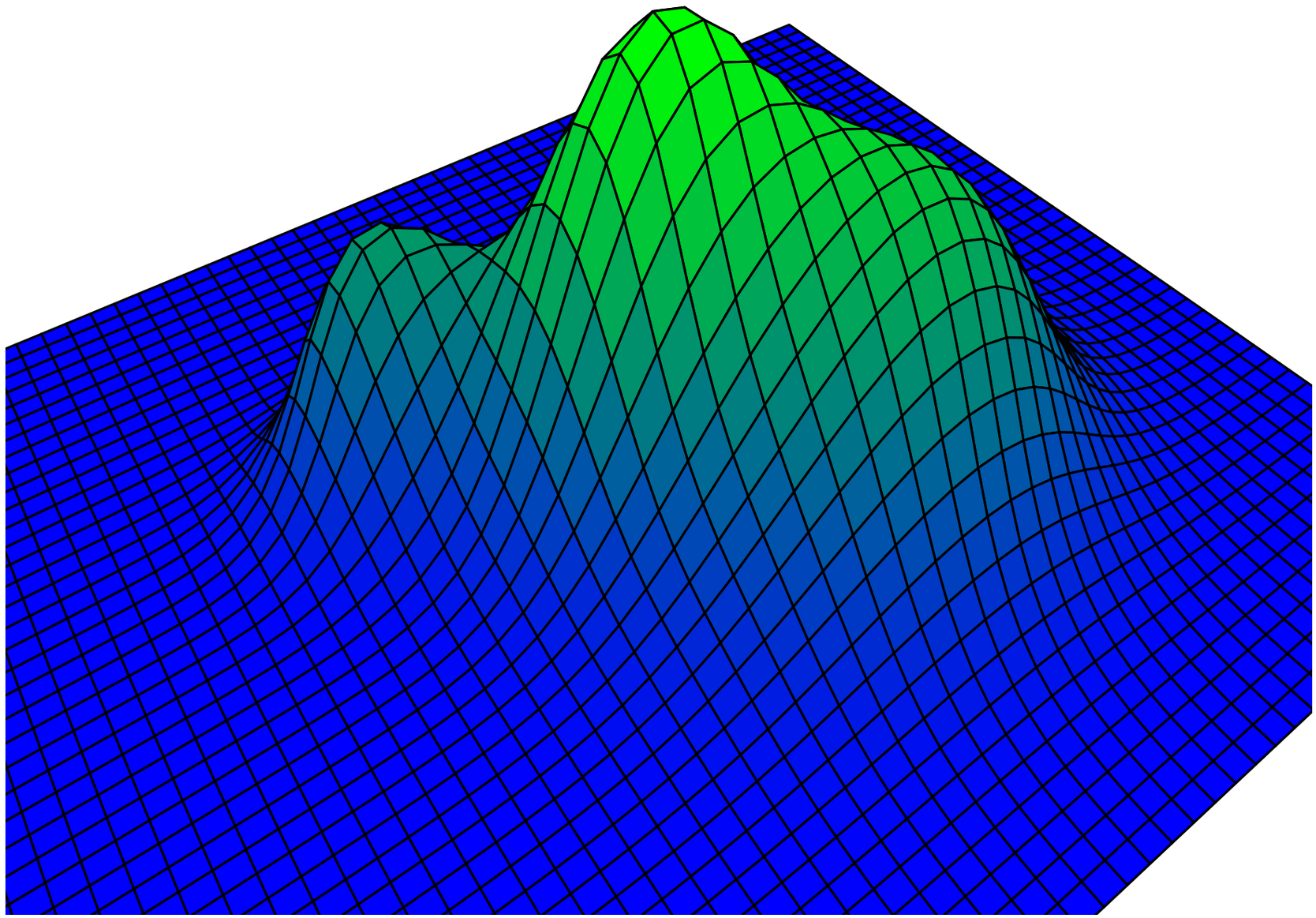
Phylogenetic inference

Maximum likelihood	**Bayesian inference**
Heuristic search	MCMC

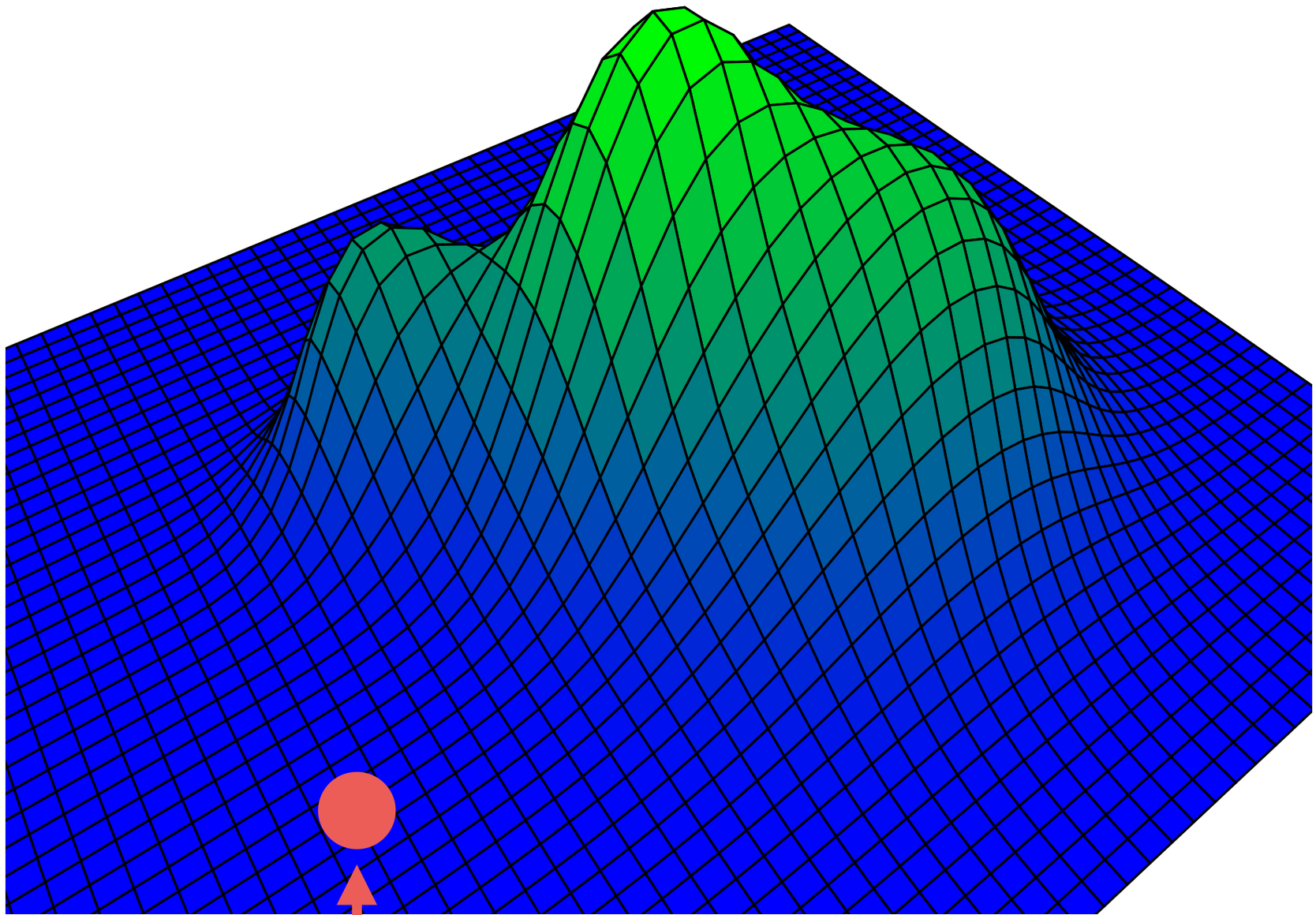
Why is MCMC so slow?

Why is MCMC so slow? Traverse tree space

Why is MCMC so slow? Traverse tree space

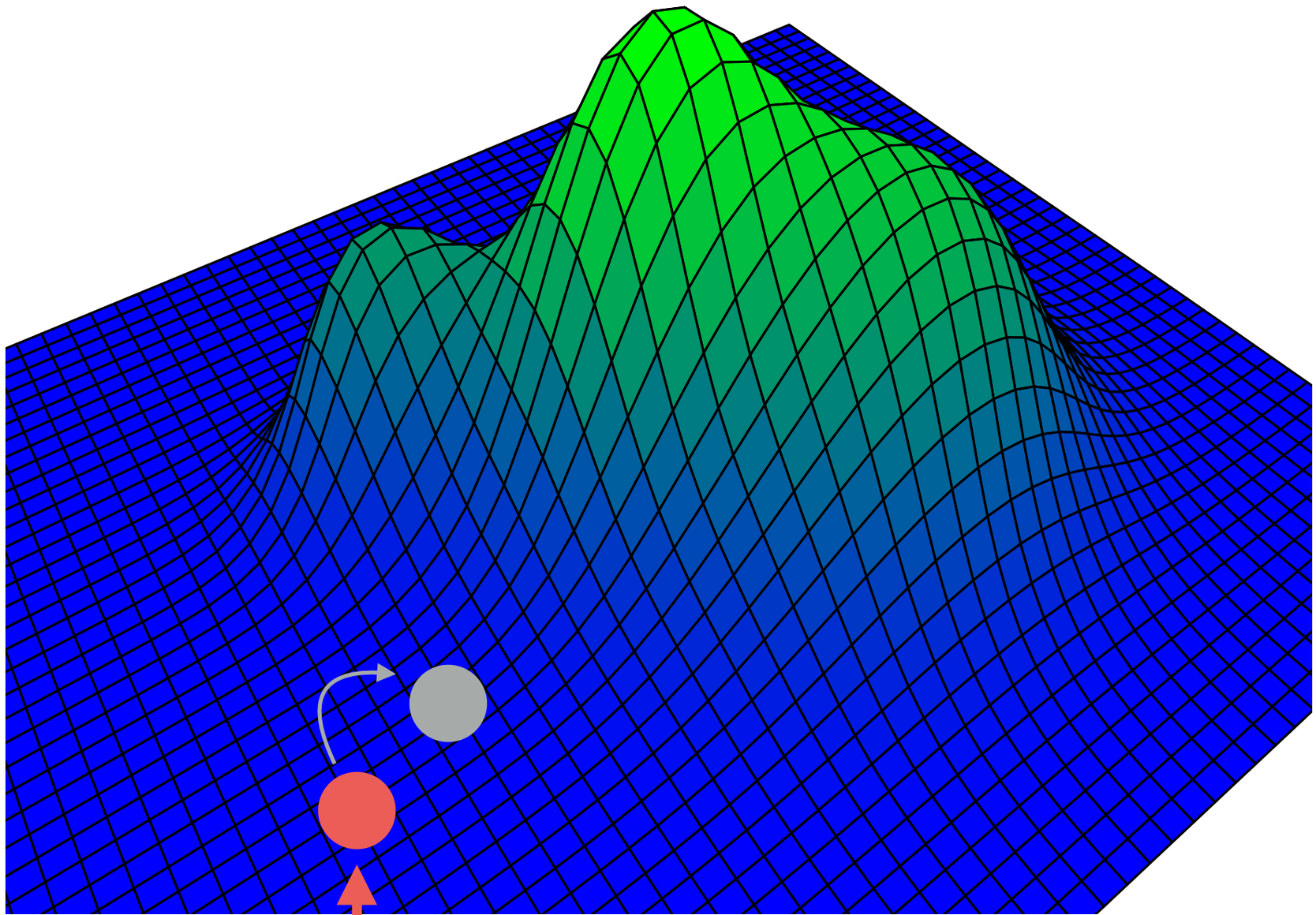


Why is MCMC so slow? Traverse tree space



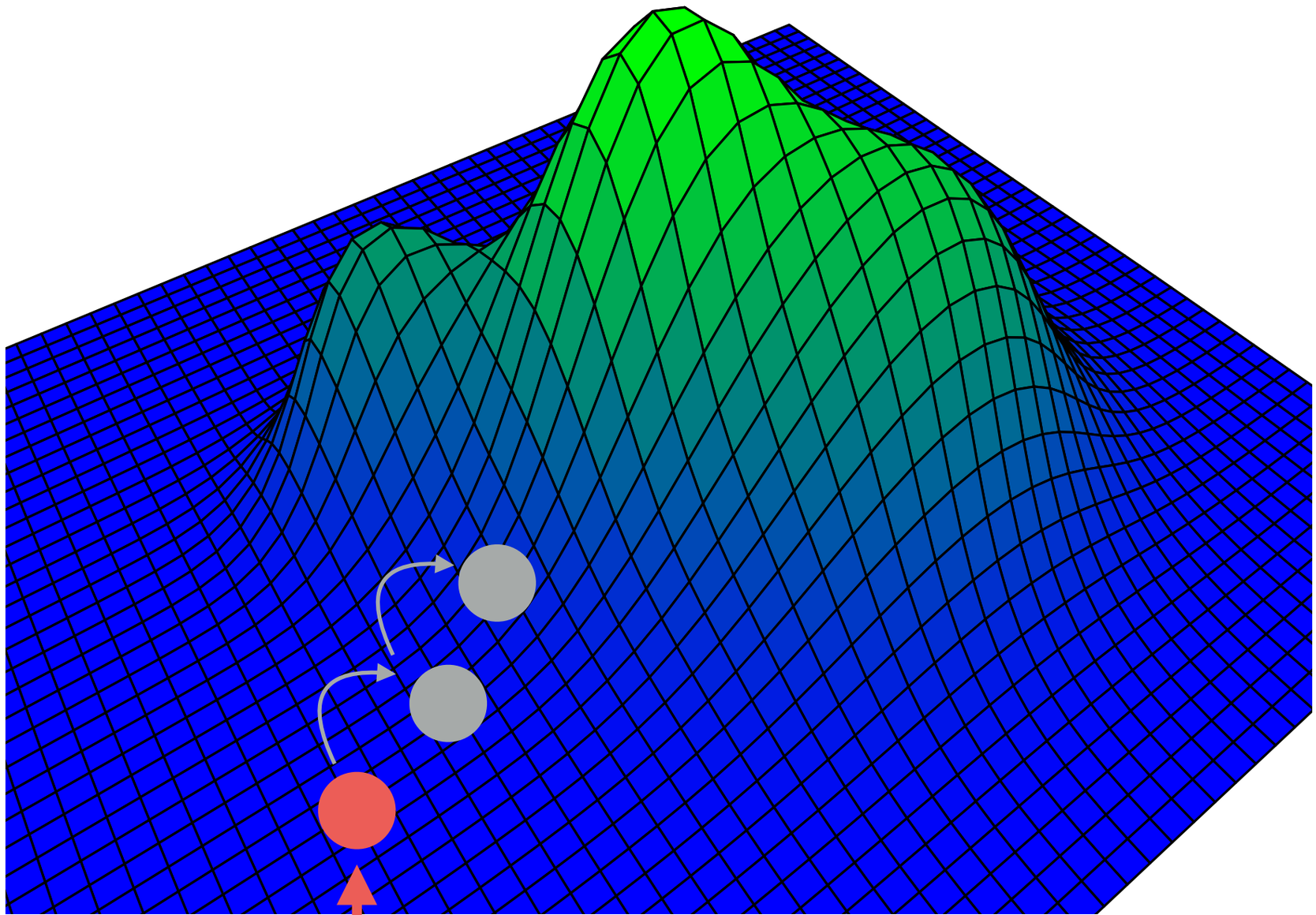
Starting tree

Why is MCMC so slow? Traverse tree space



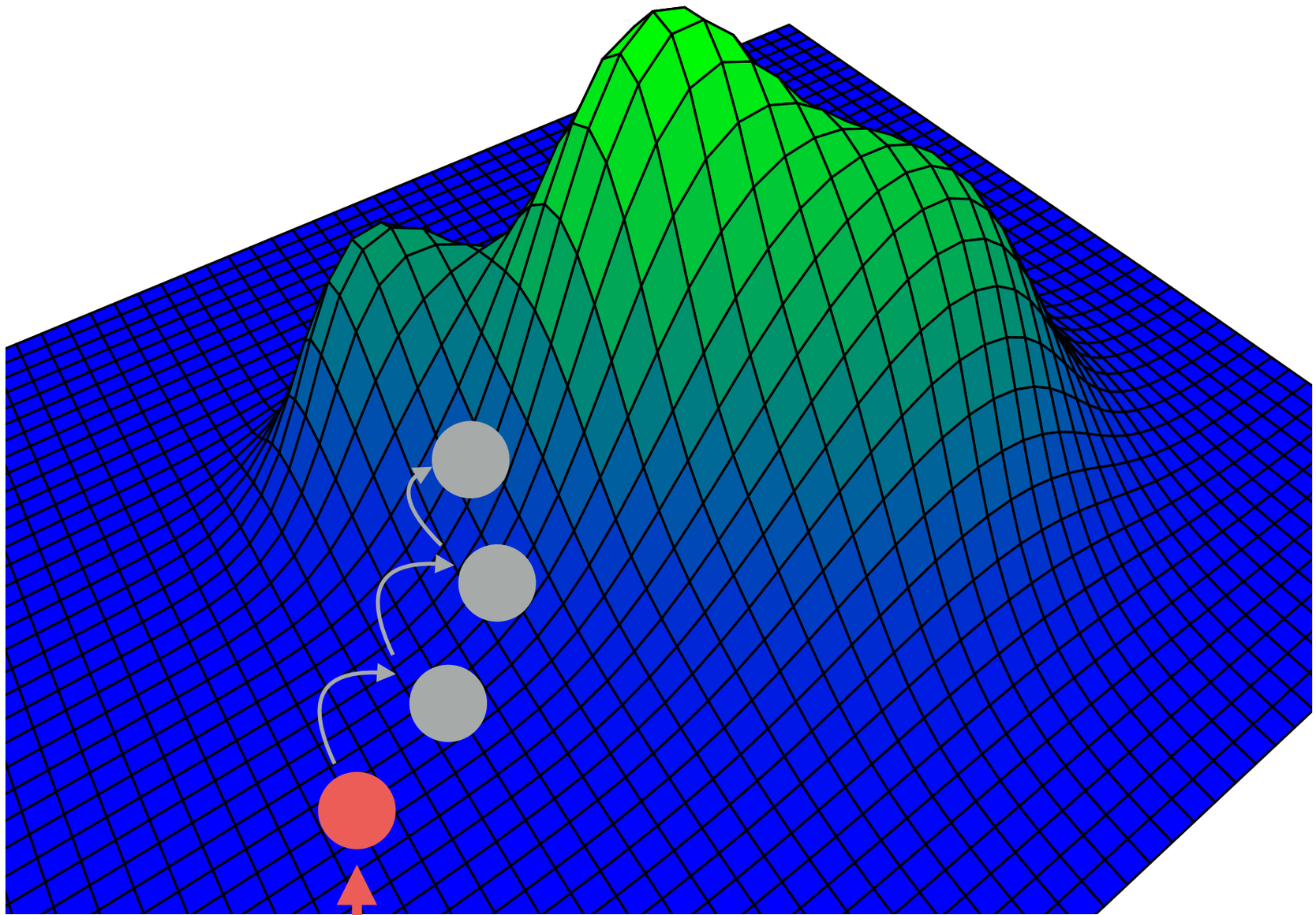
Starting tree

Why is MCMC so slow? Traverse tree space



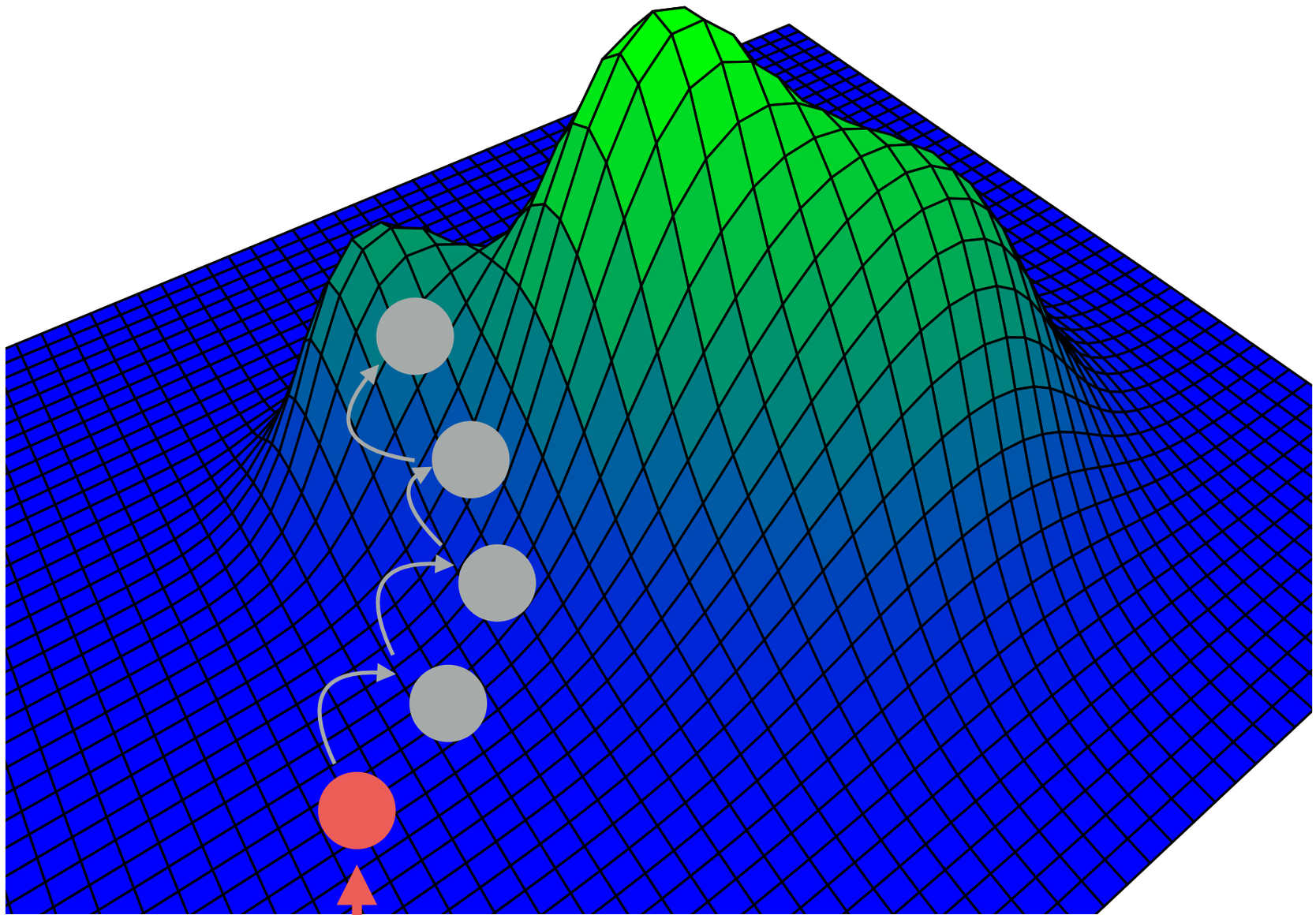
Starting tree

Why is MCMC so slow? Traverse tree space



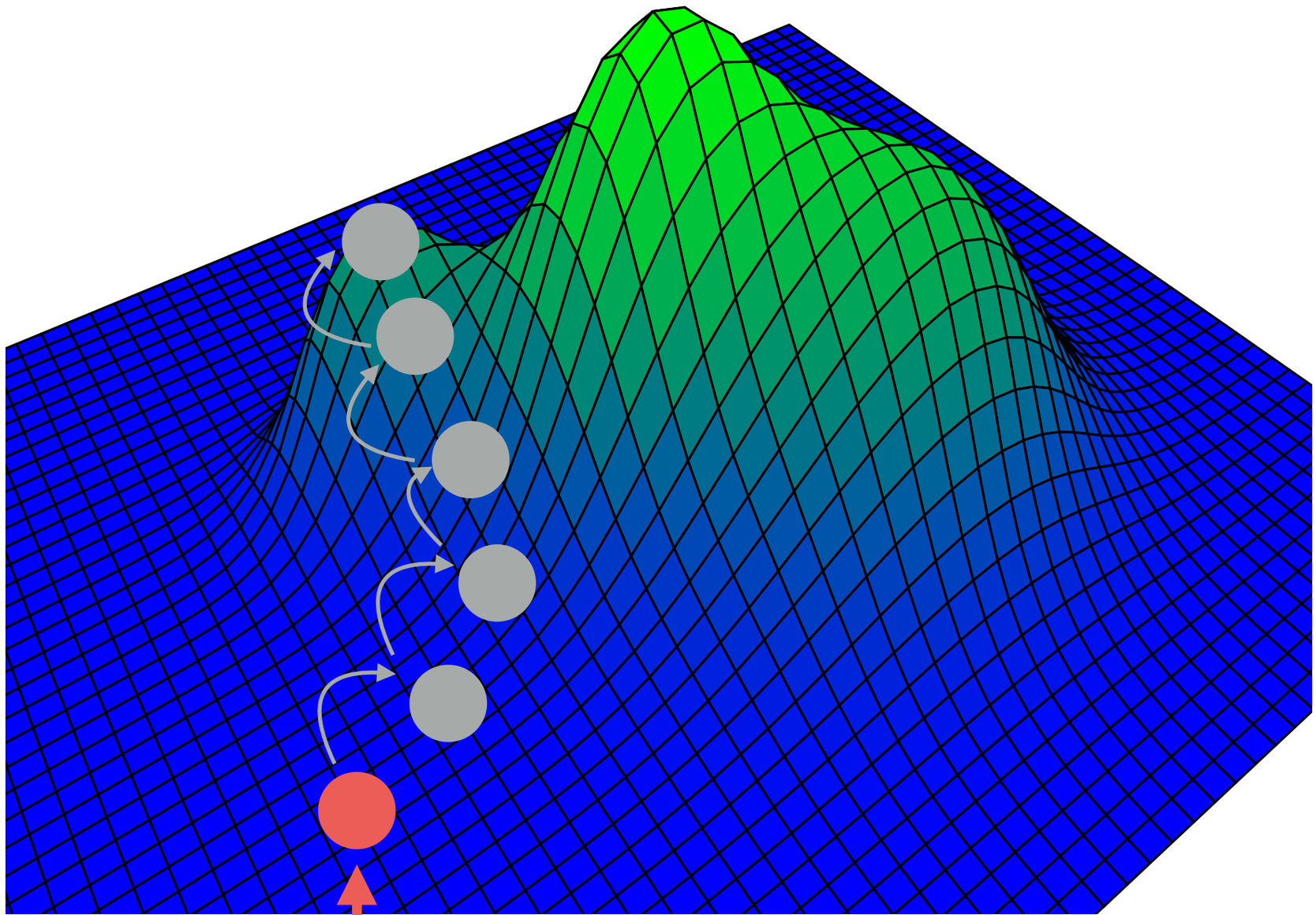
Starting tree

Why is MCMC so slow? Traverse tree space



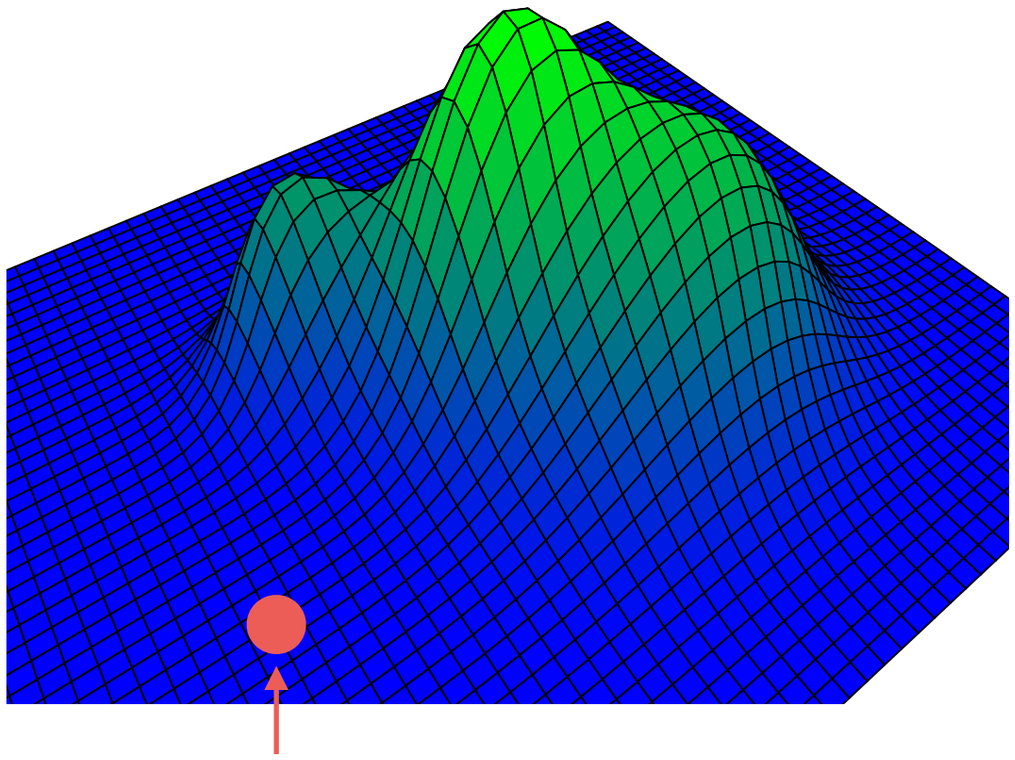
Starting tree

Why is MCMC so slow? Traverse tree space

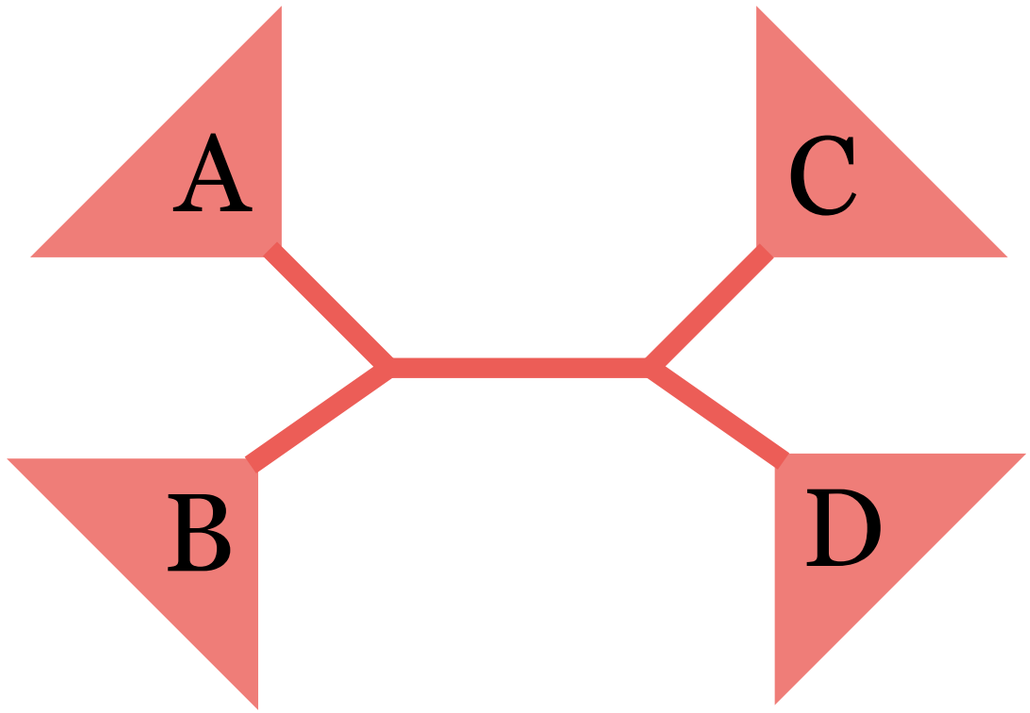


Starting tree

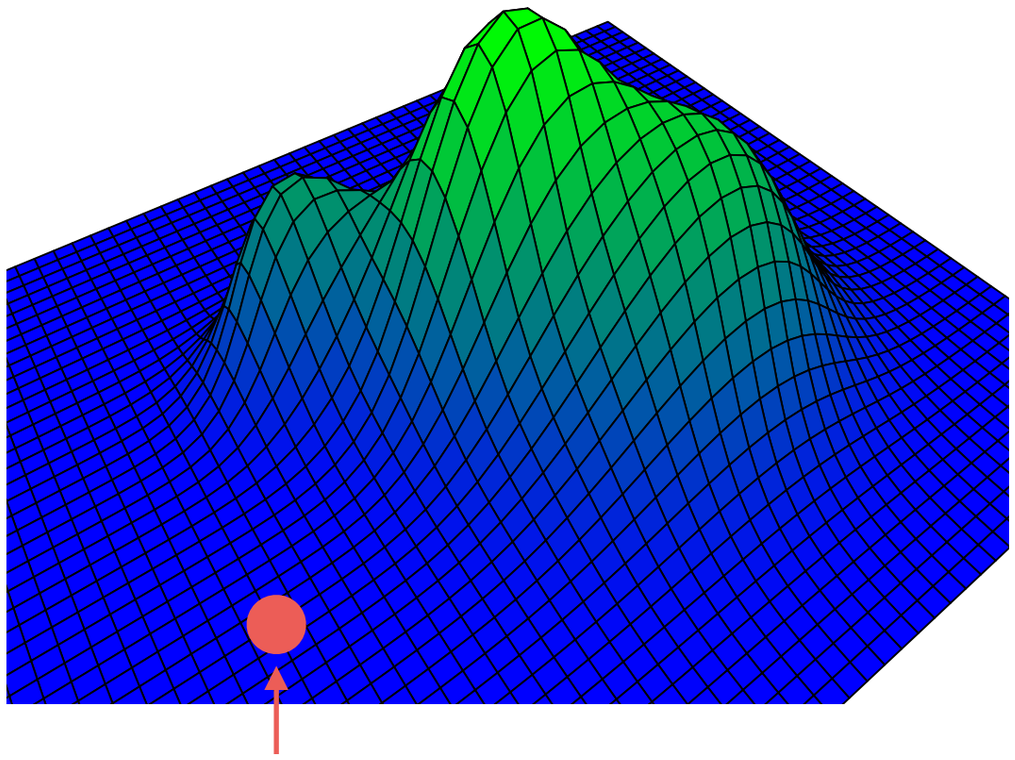
Why is MCMC so slow? Traverse tree space



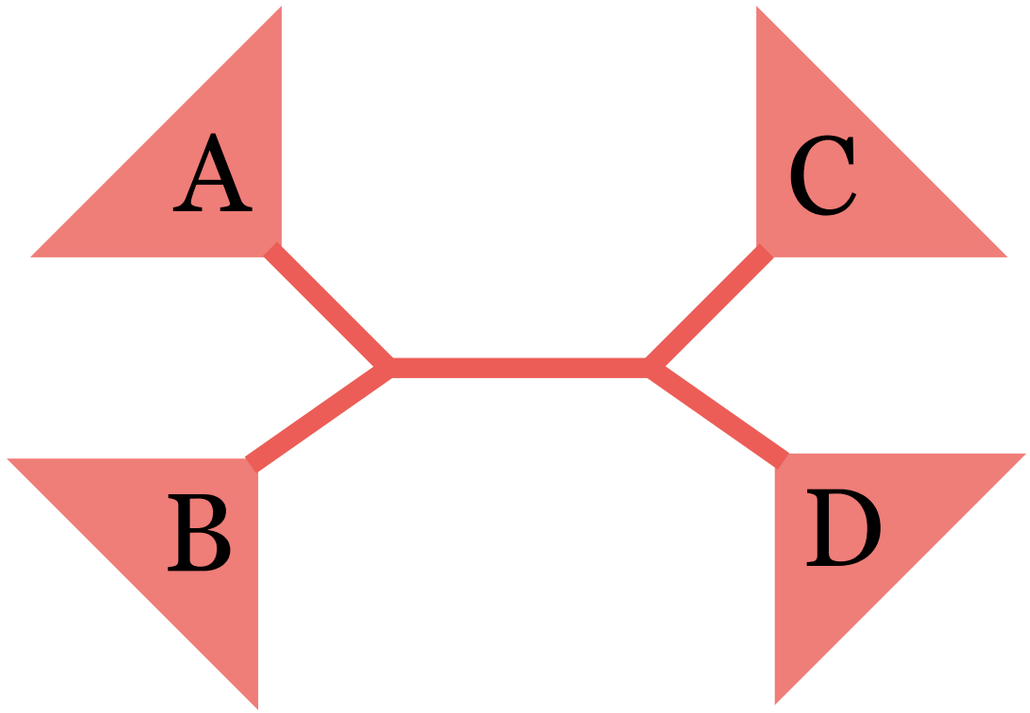
Starting tree



Why is MCMC so slow? Traverse tree space

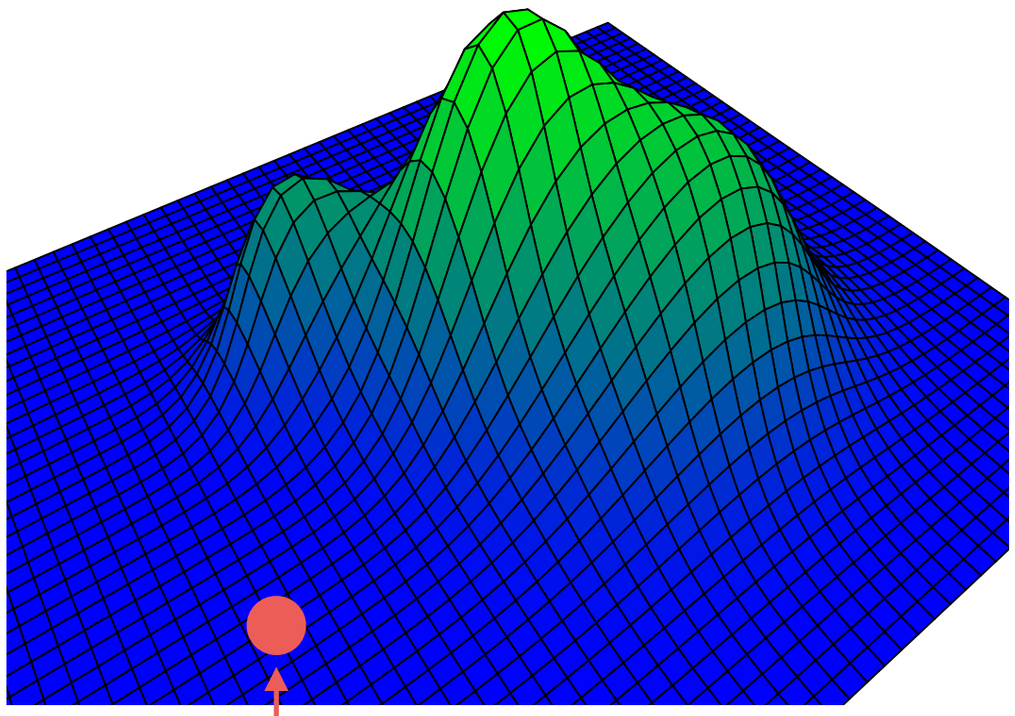


Starting tree

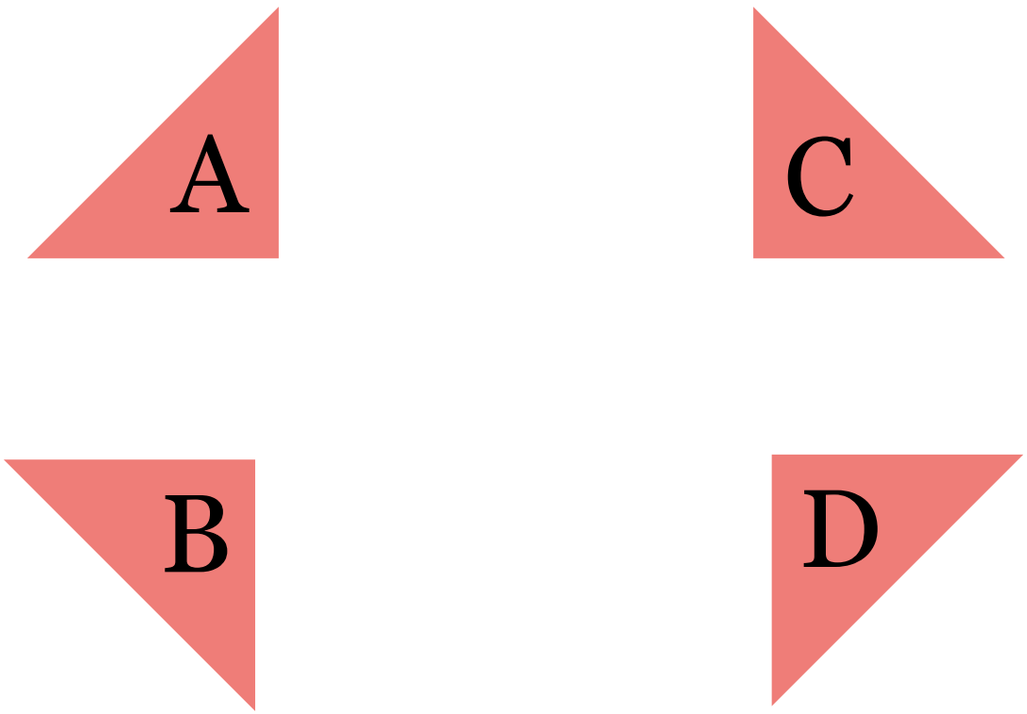


Nearest Neighbor Interchange

Why is MCMC so slow? Traverse tree space



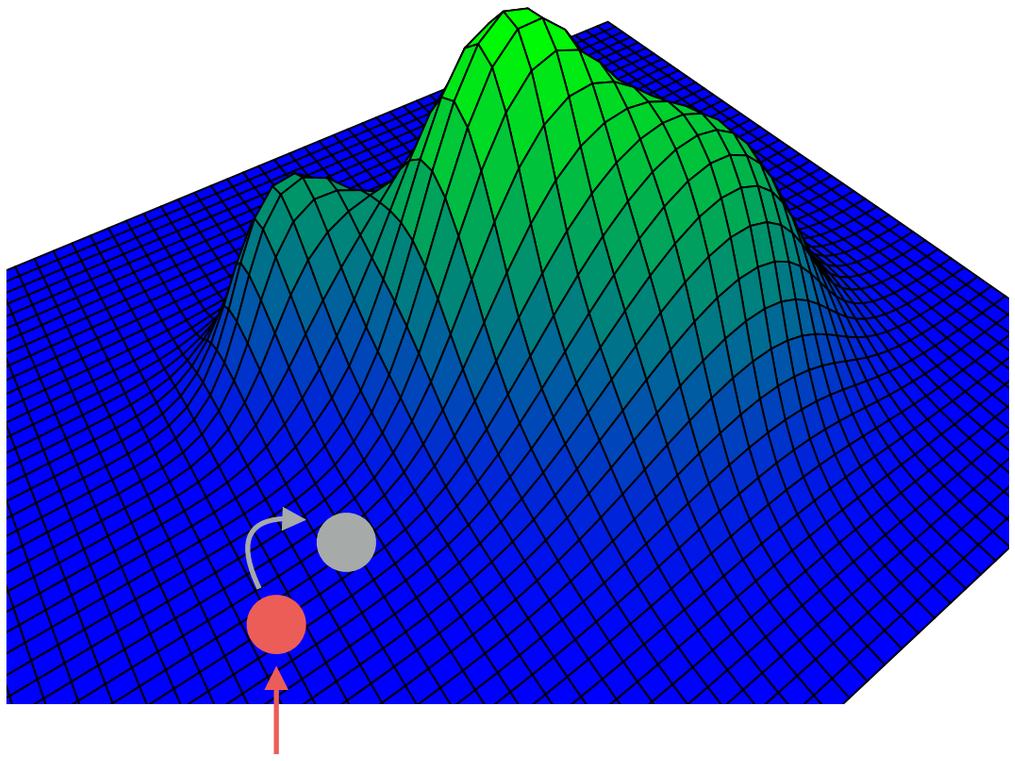
Starting tree



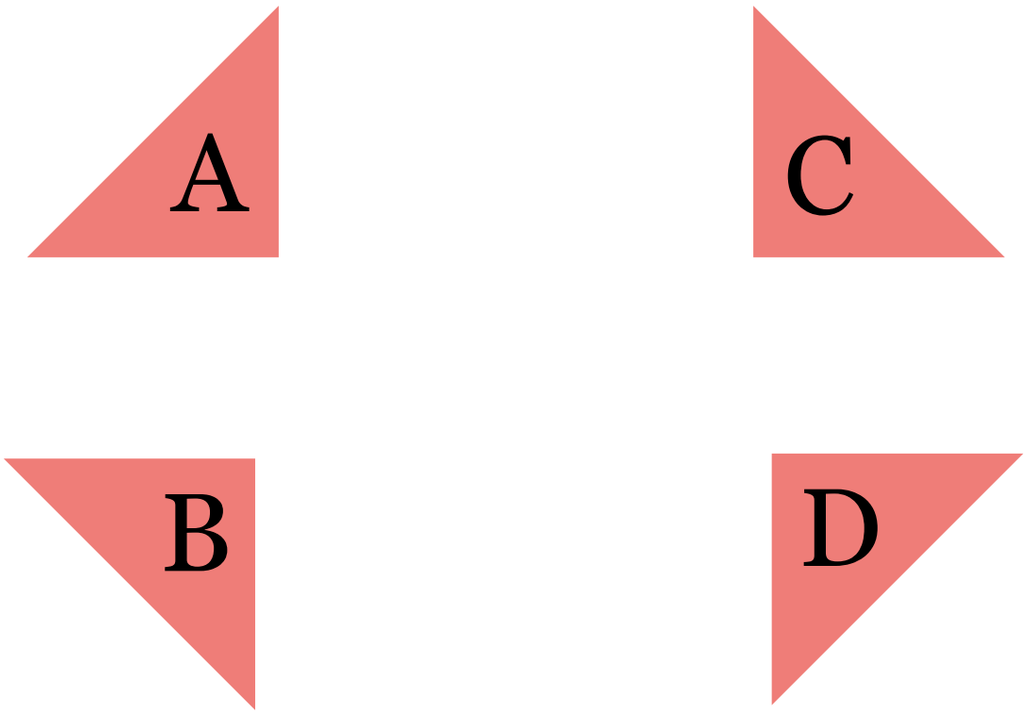
Nearest Neighbor Interchange

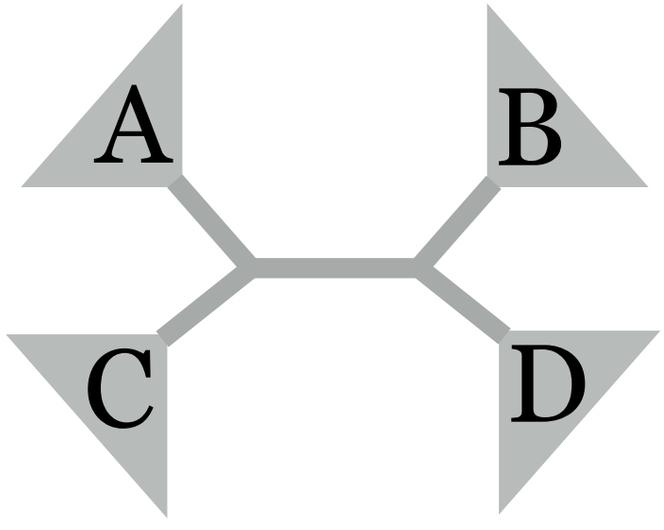
Why is MCMC so slow?

Traverse tree space



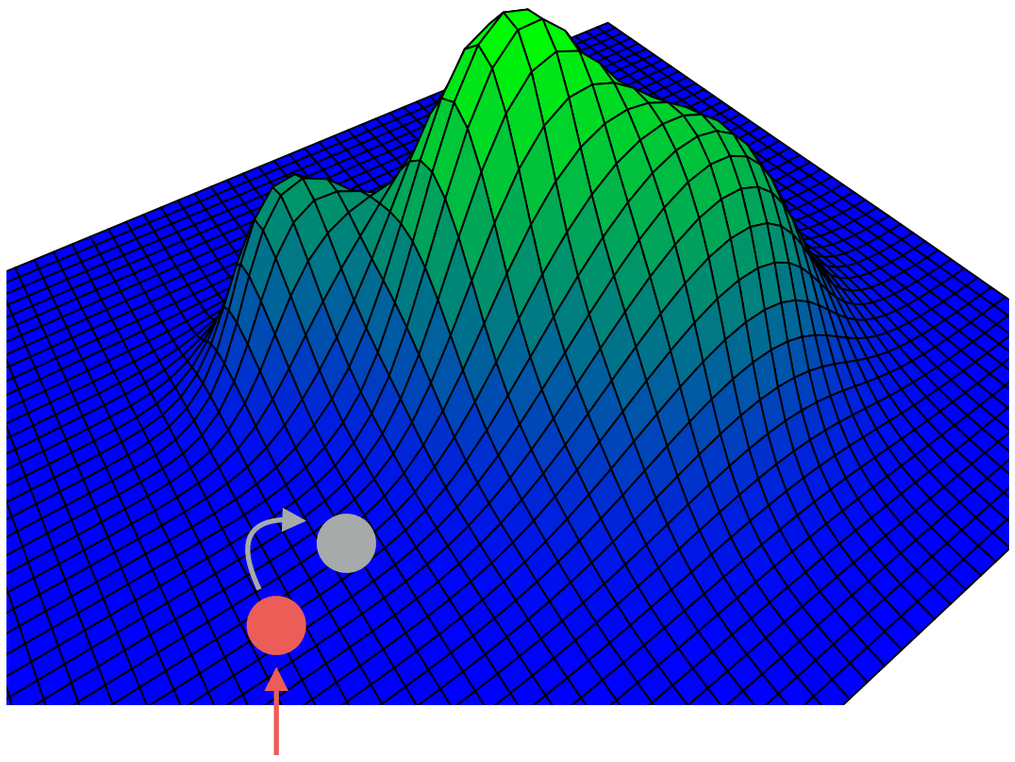
Starting tree



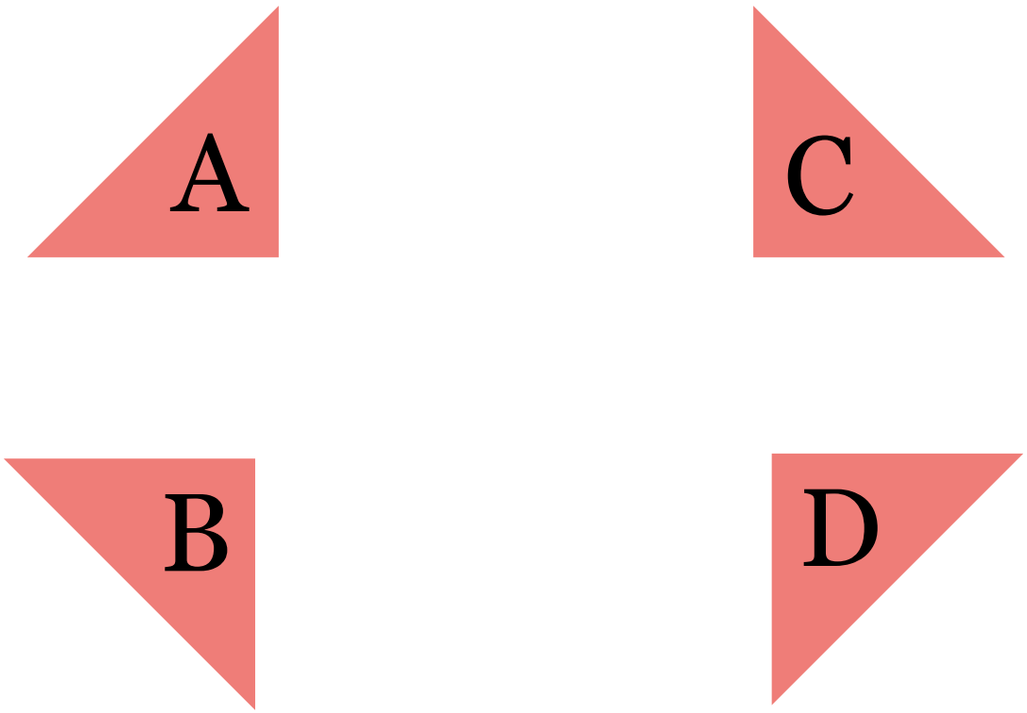


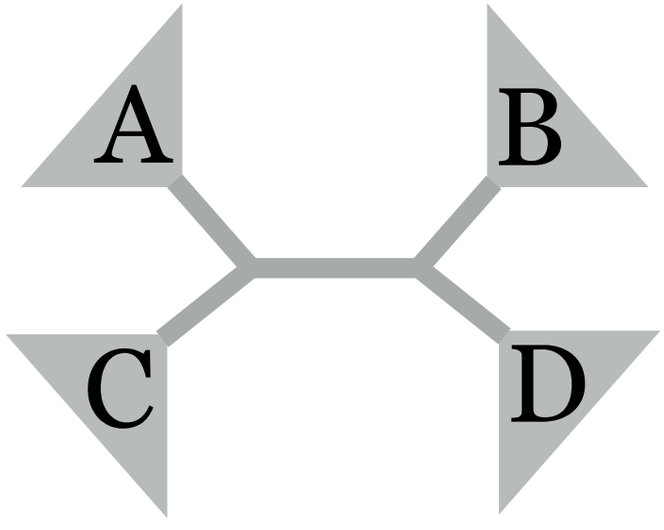
Nearest Neighbor Interchange

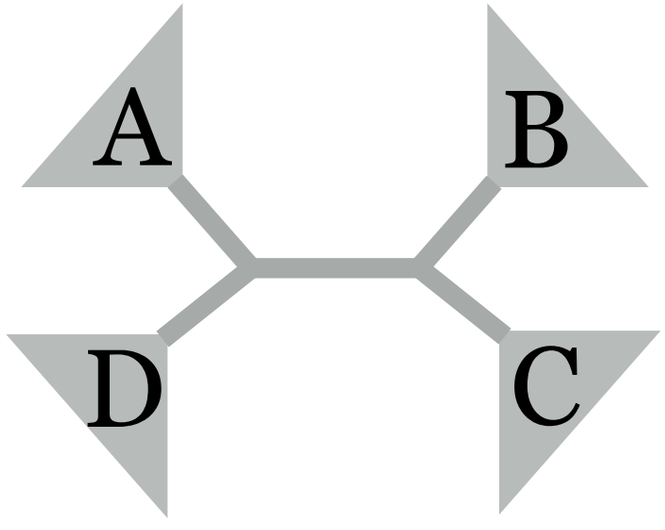
Why is MCMC so slow? Traverse tree space



Starting tree







Nearest Neighbor Interchange

Why is MCMC so slow?

Why is MCMC so slow? Tree space is huge

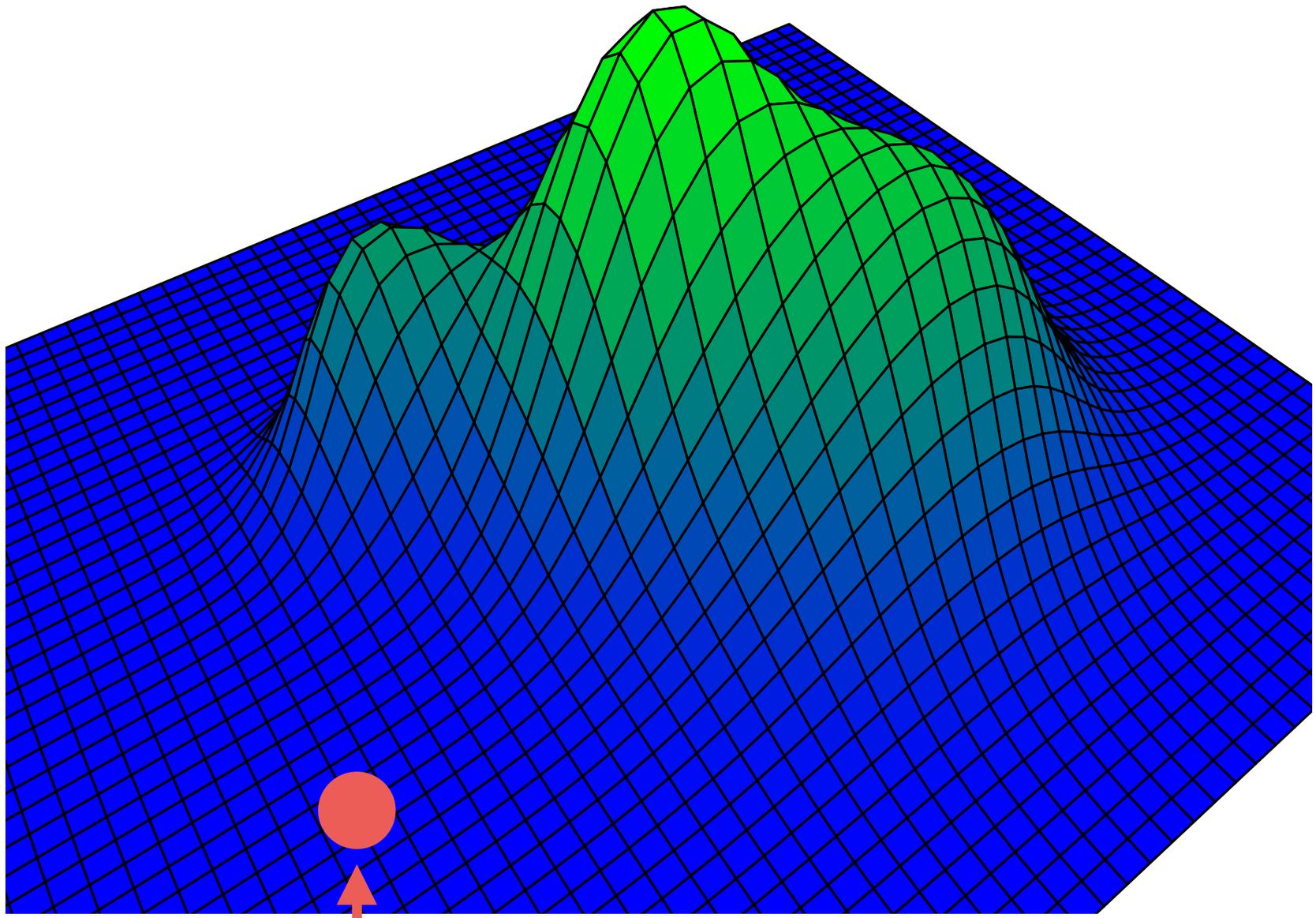
Why is MCMC so slow? Tree space is huge

# Species	# Unrooted trees	# Rooted trees
1	1	1
2	1	1
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395
8	10,395	135,135
9	135,135	2,027,025
10	2,027,025	34,459,425
11	34,459,425	654,729,075
12	654,729,075	13,749,310,575
13	13,749,310,575	316,234,143,225
⋮	⋮	⋮
52	$>$ # atoms in universe	

Why is MCMC so slow?

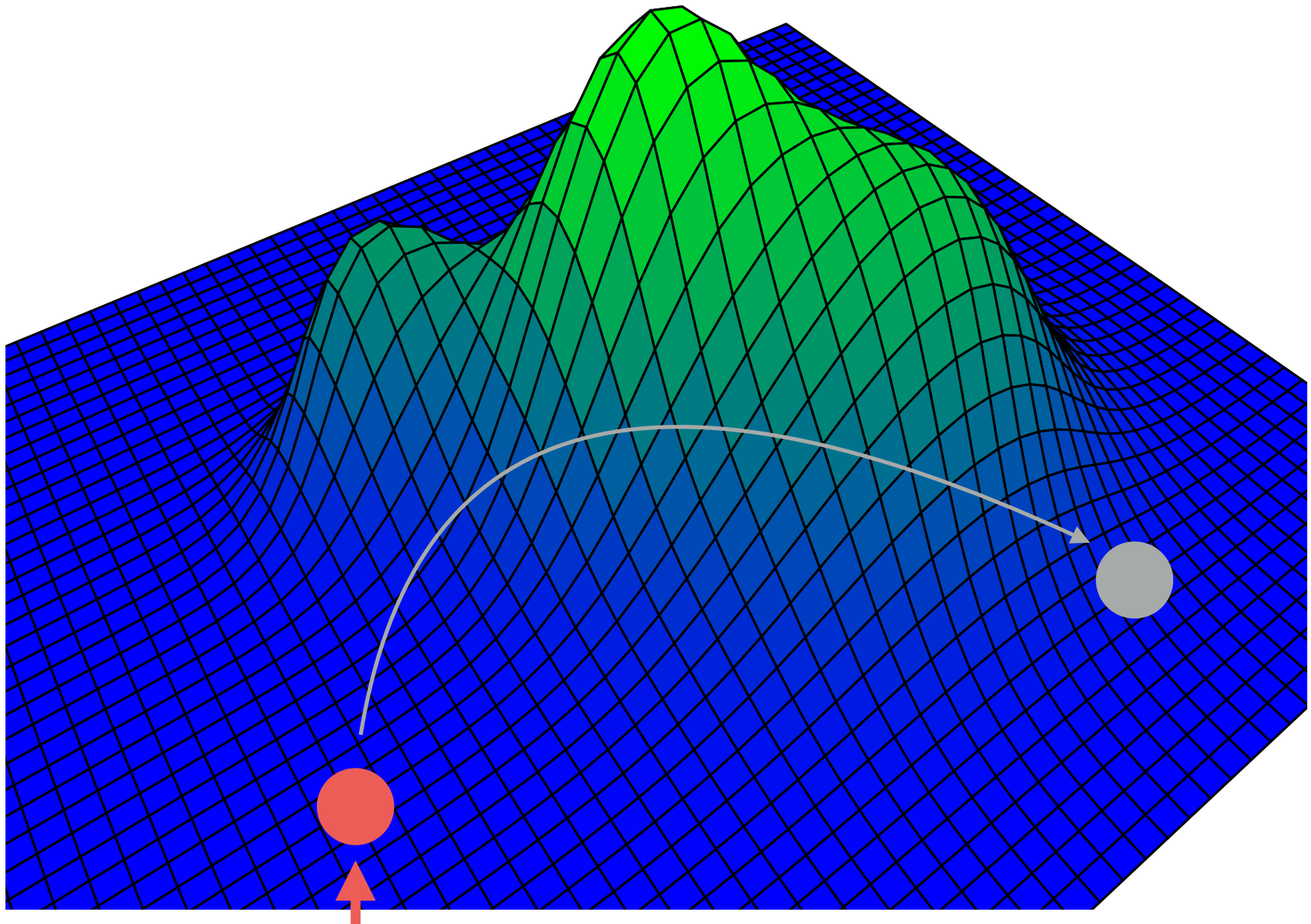
Why is MCMC so slow? Low acceptance of moves

Why is MCMC so slow? Low acceptance of moves



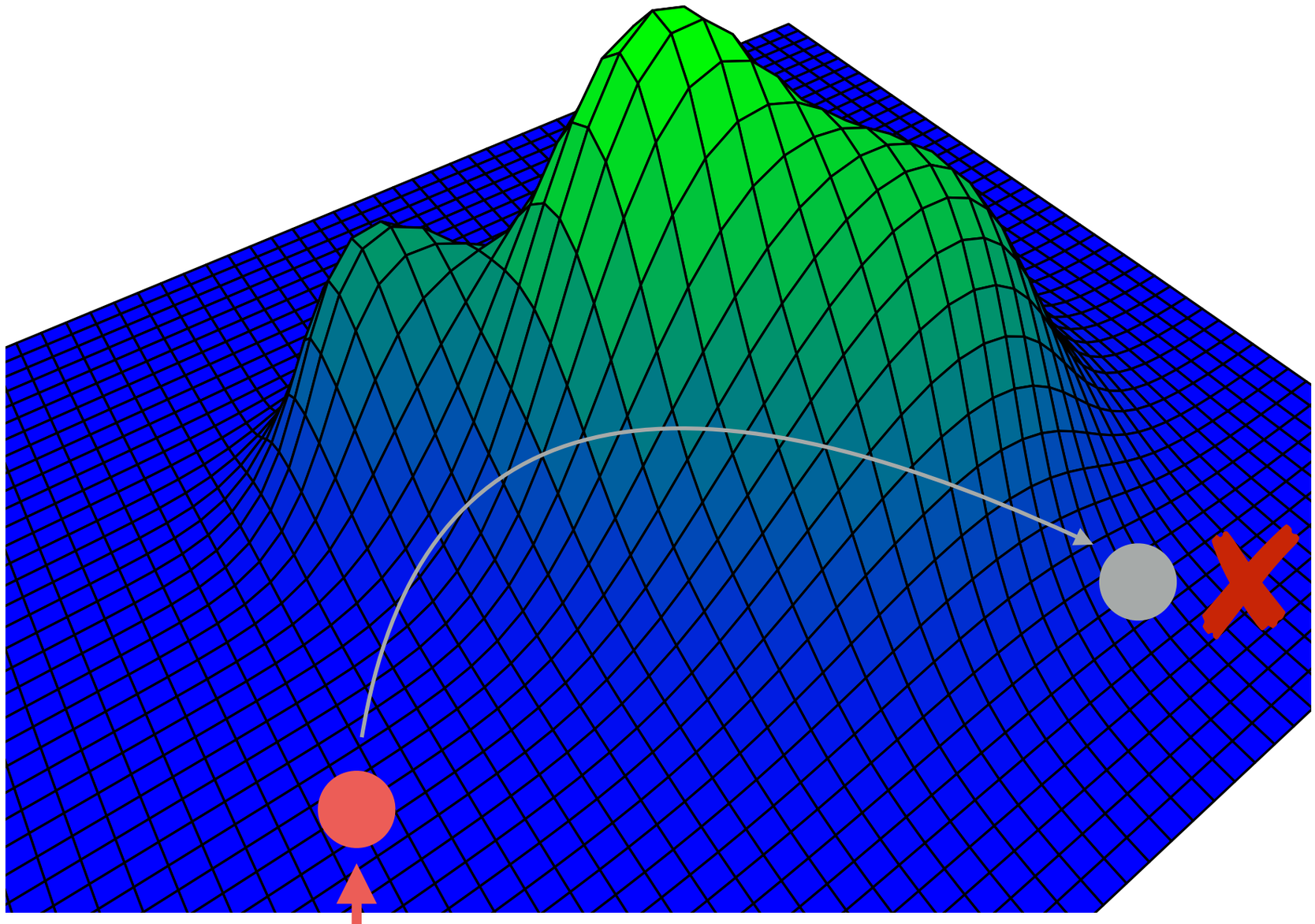
Starting tree

Why is MCMC so slow? Low acceptance of moves



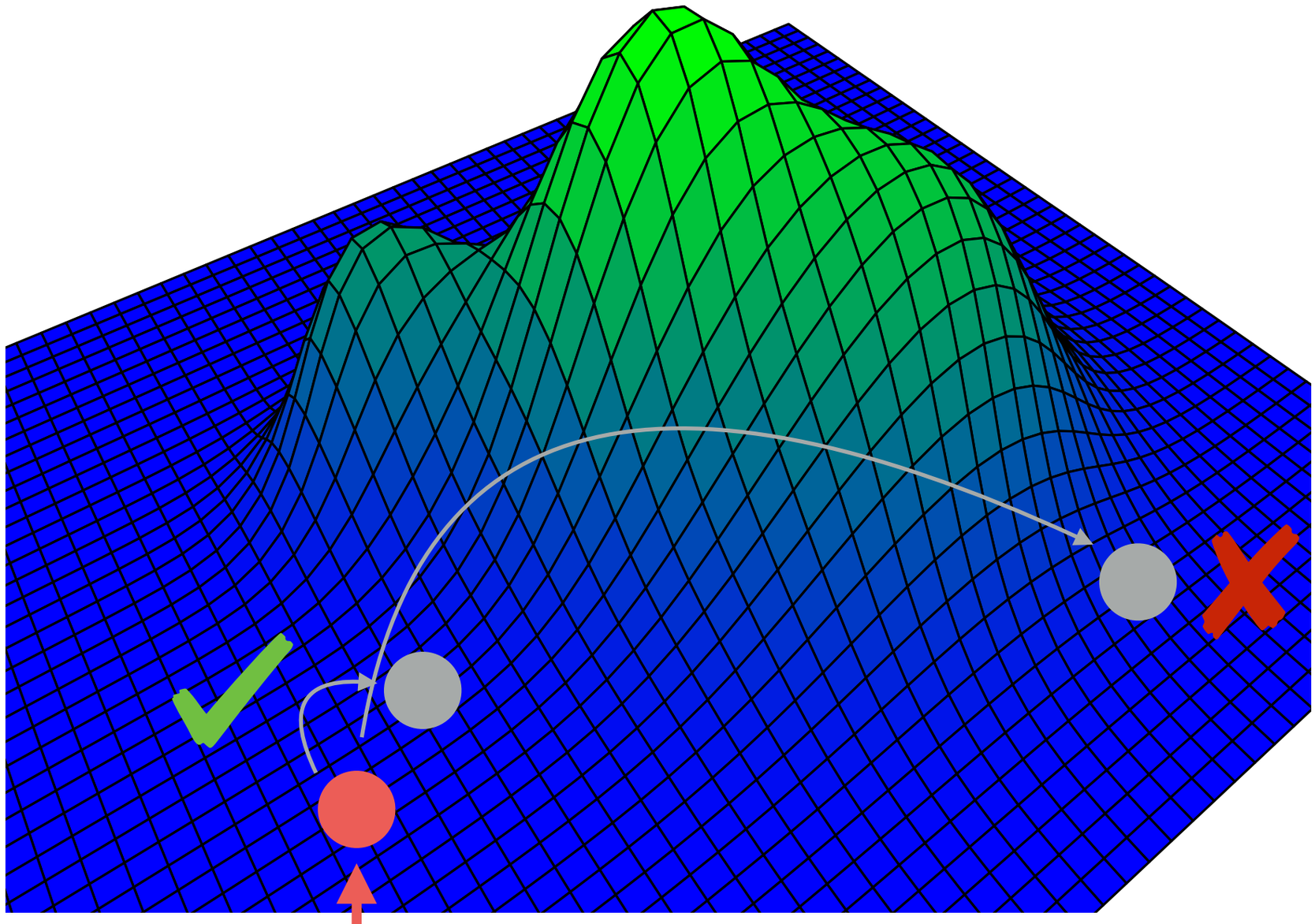
Starting tree

Why is MCMC so slow? Low acceptance of moves



Starting tree

Why is MCMC so slow? Low acceptance of moves



Starting tree

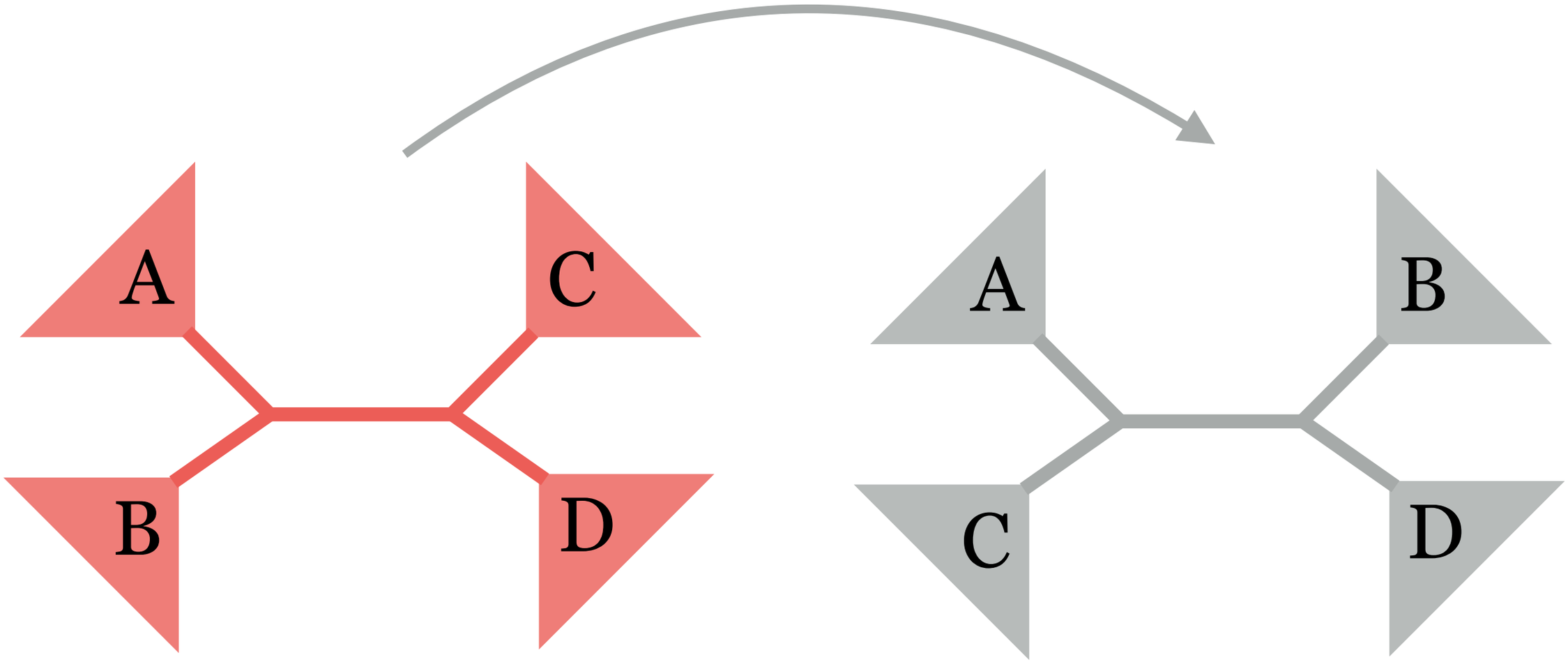
Why is MCMC so slow?

Why is MCMC so slow?

Small neighborhood implies very dependent sample

Why is MCMC so slow?

Small neighborhood implies very dependent sample



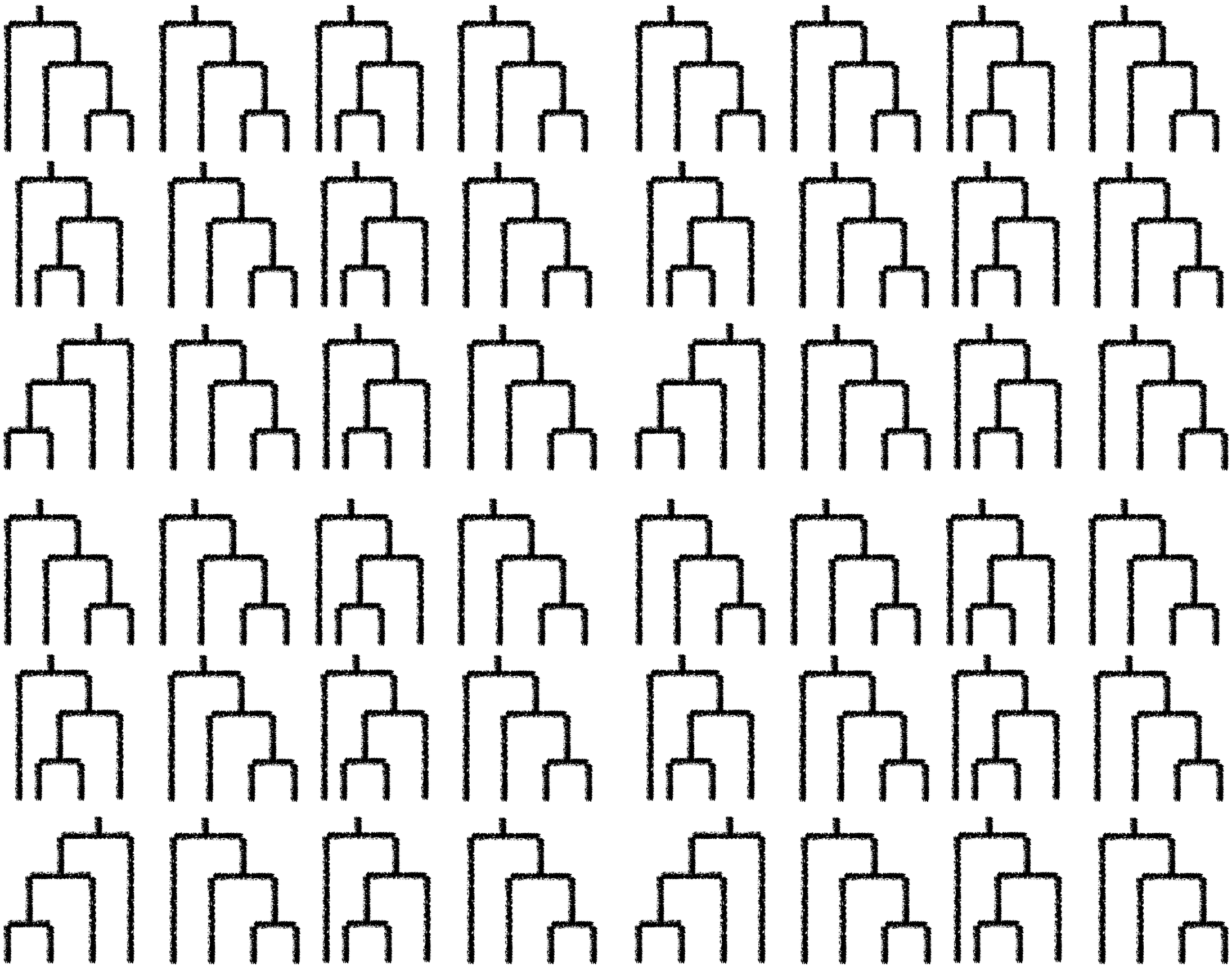
Why is MCMC so slow?

1) Huge tree space size
2) Low acceptance of moves unless small neighborhood
3) Small neighborhood implies very dependent sample, which means small effective sample size

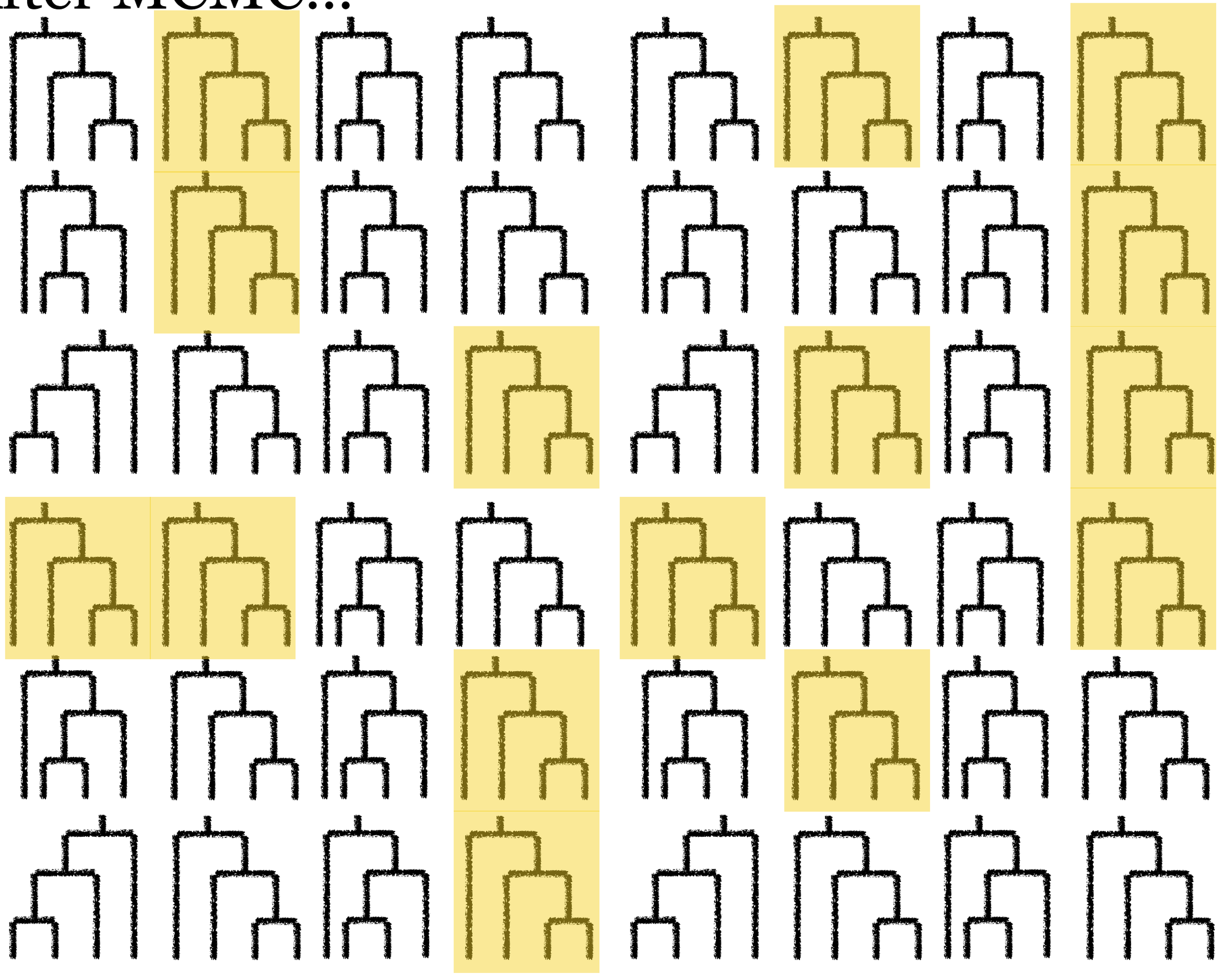
We need a gigantic chain because the space is huge and we are making tiny moves

After MCMC…

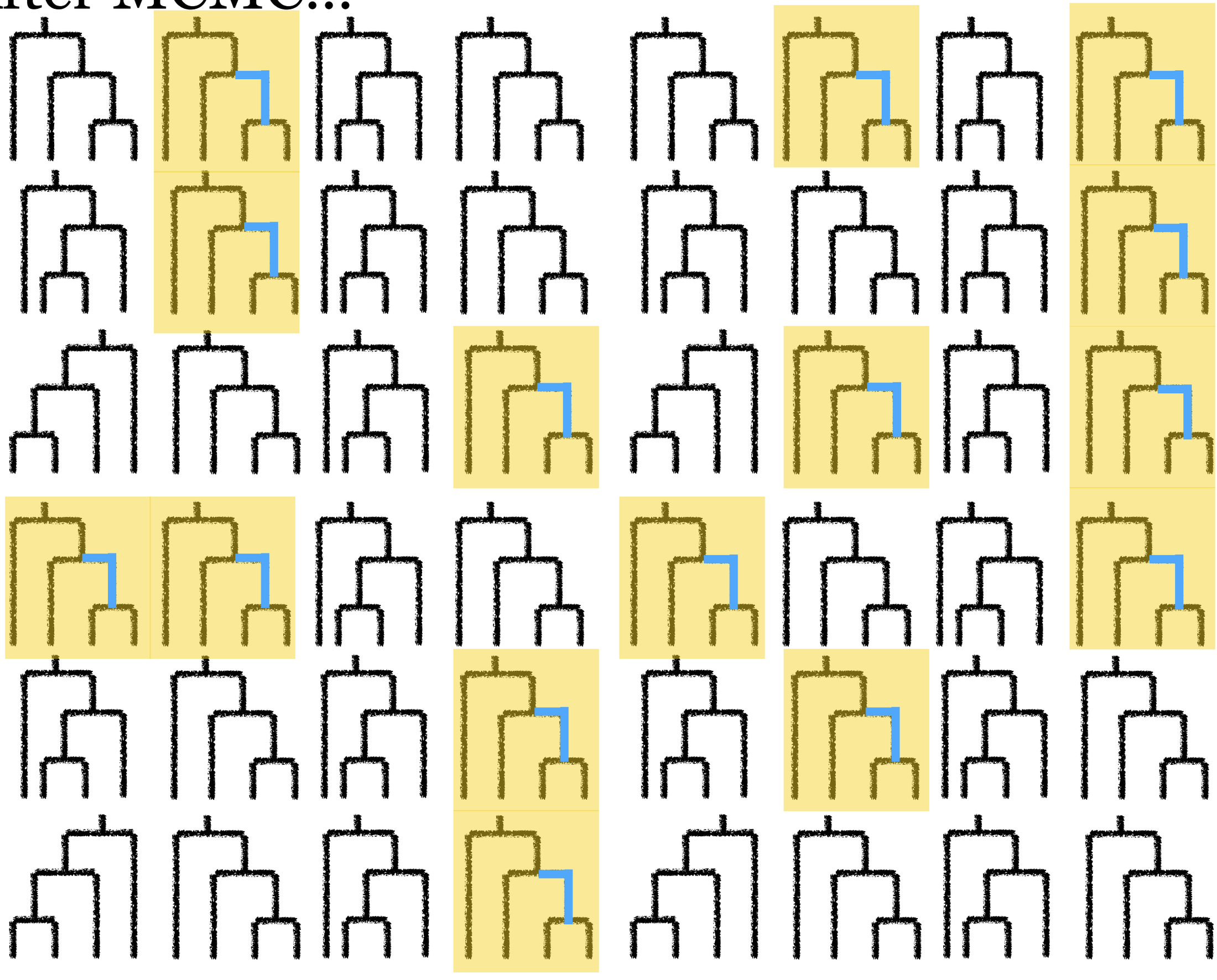
After MCMC…

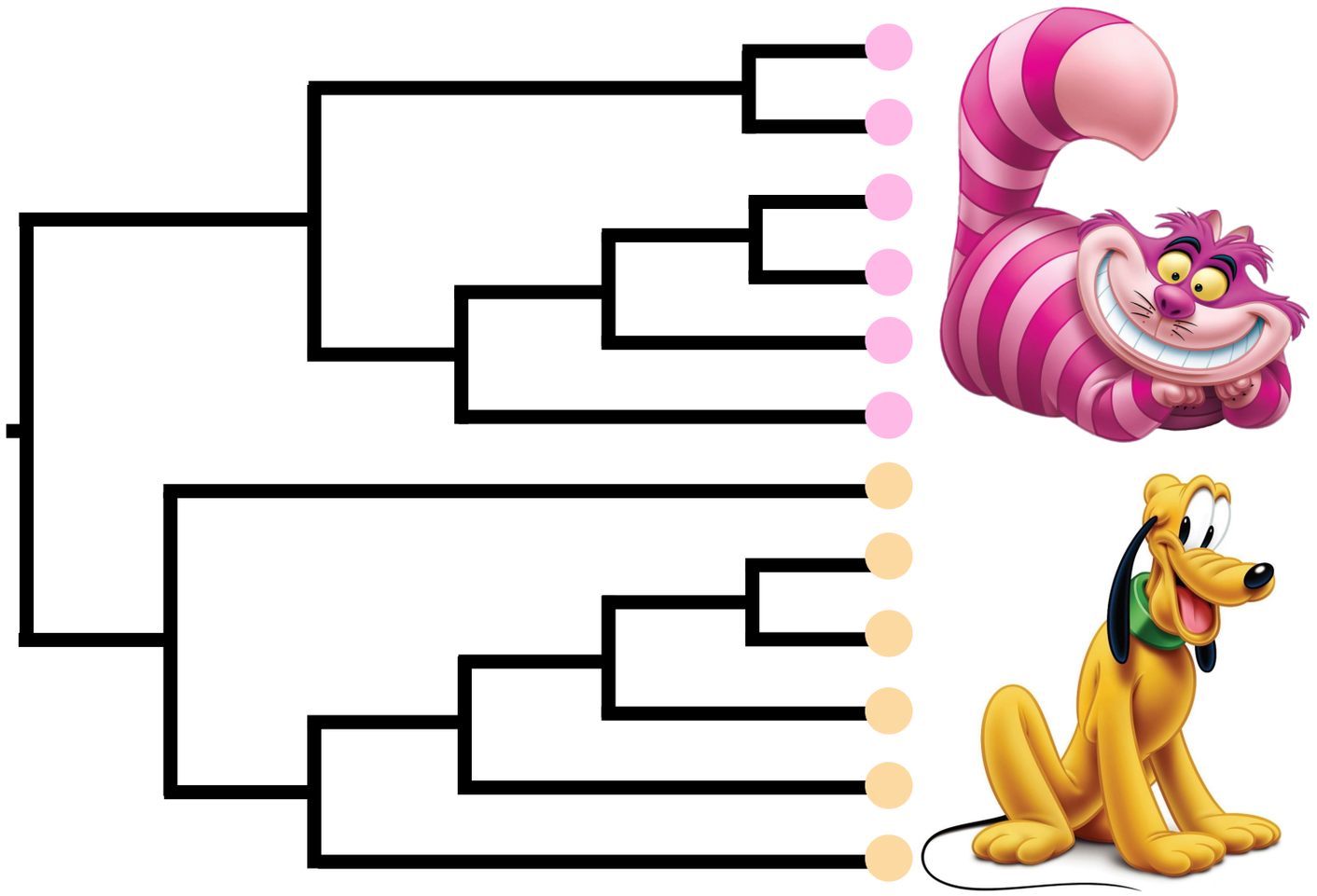


After MCMC…



After MCMC…





12 taxa *Carnivora*

MCMC efficiency ~0.025%

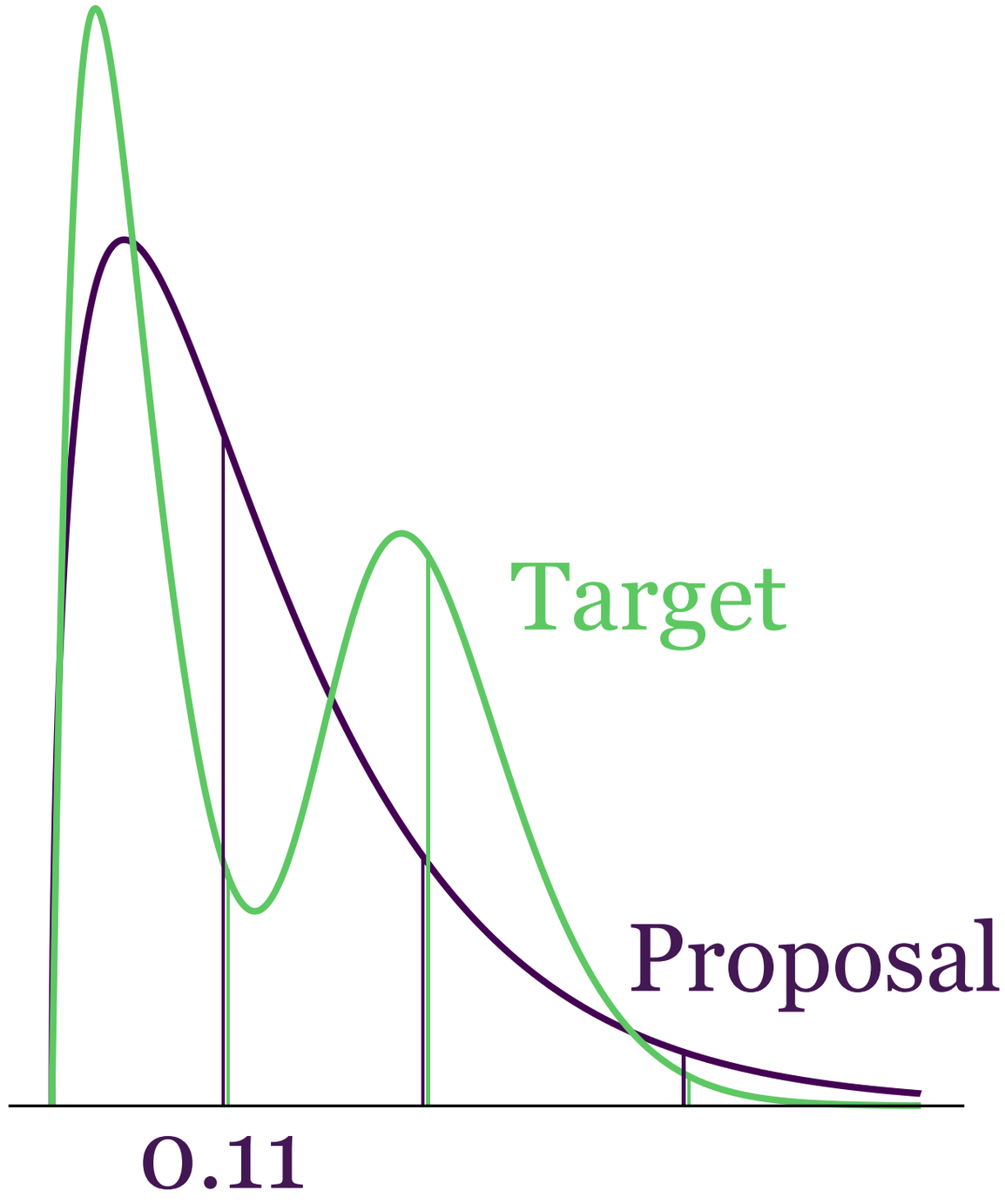
(250 from 1 million post-burnin generations)

(Larget, 2013)

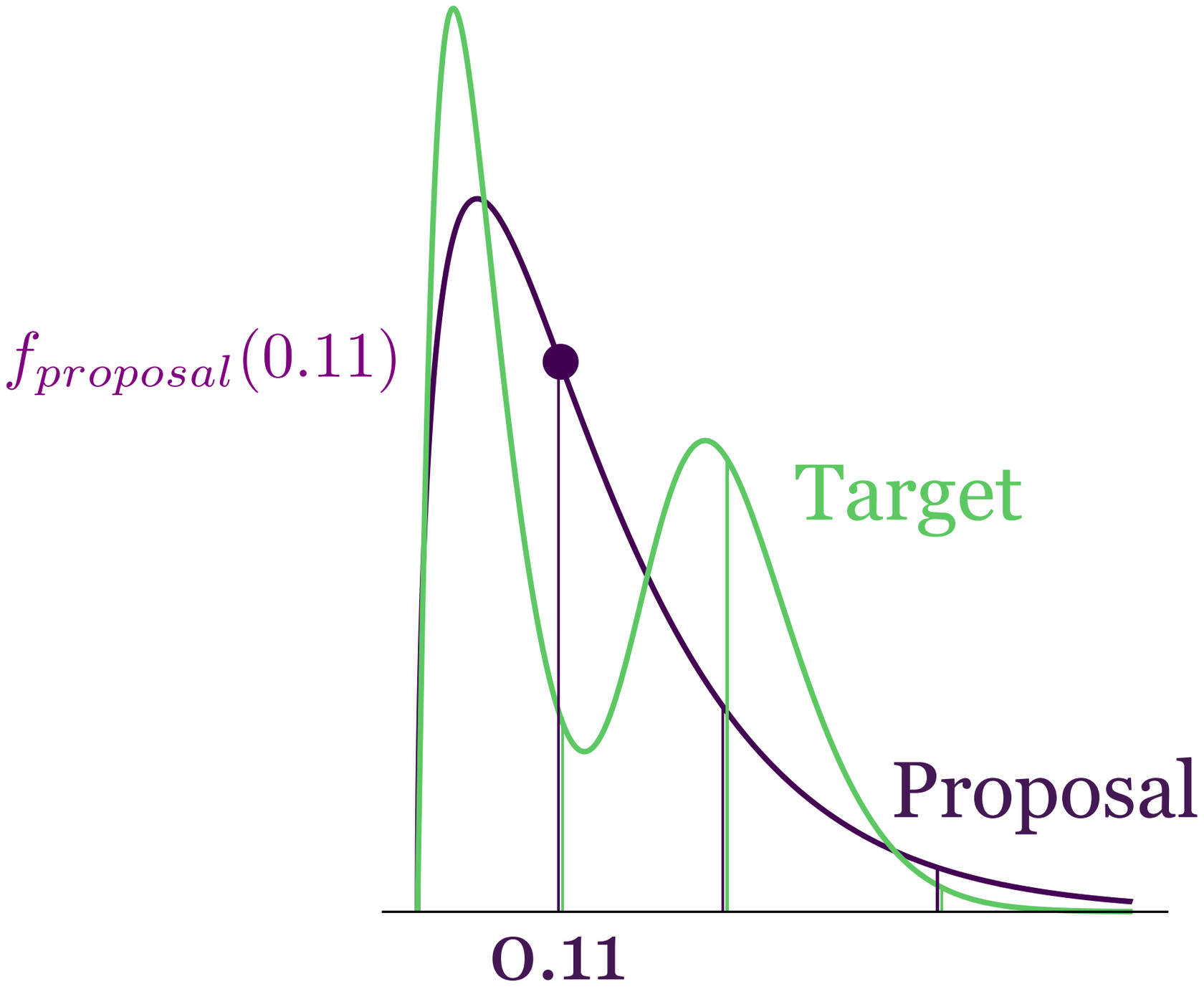
What if we could sample from the posterior more efficiently?

Importance sampling

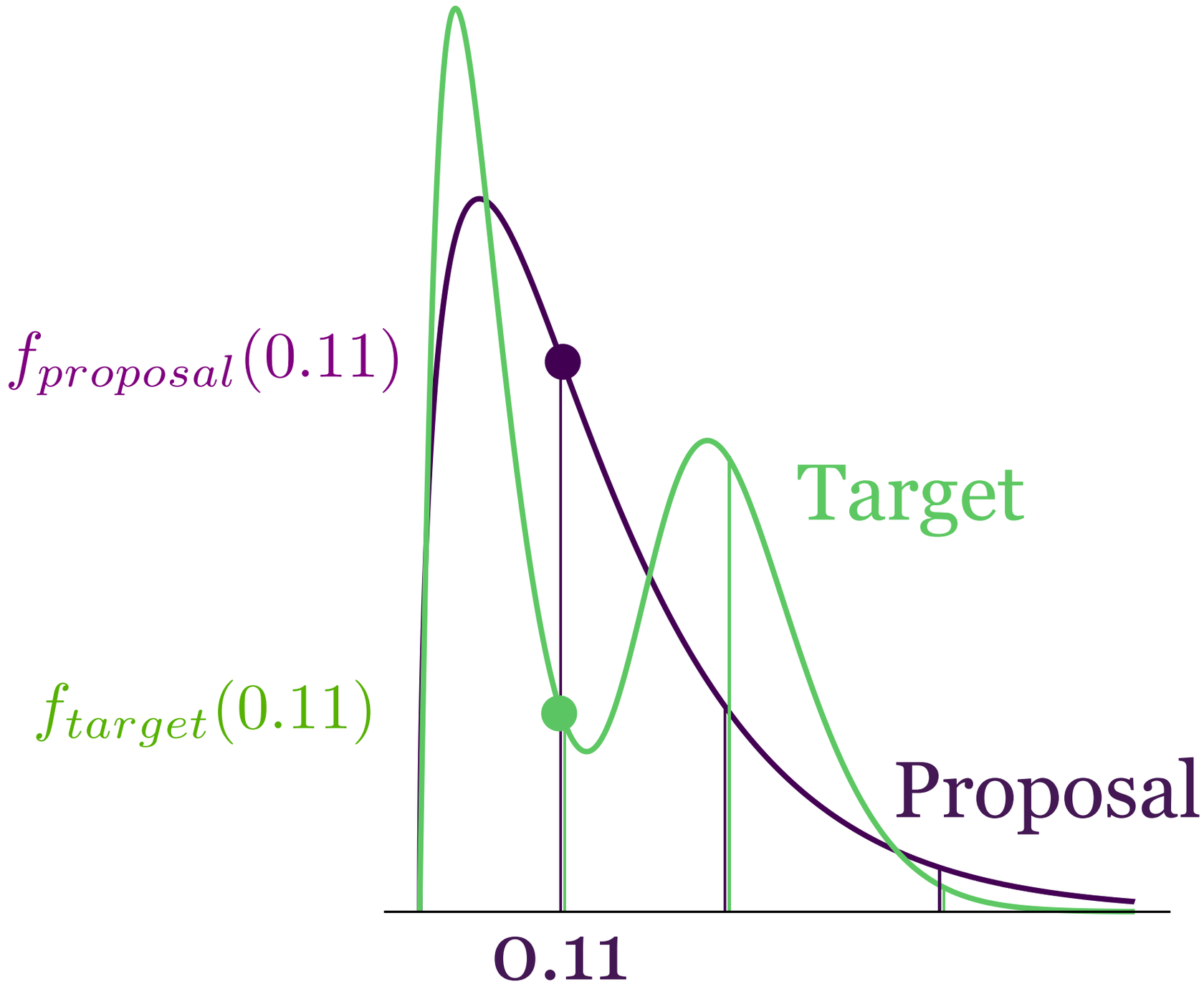
Importance sampling



Importance sampling

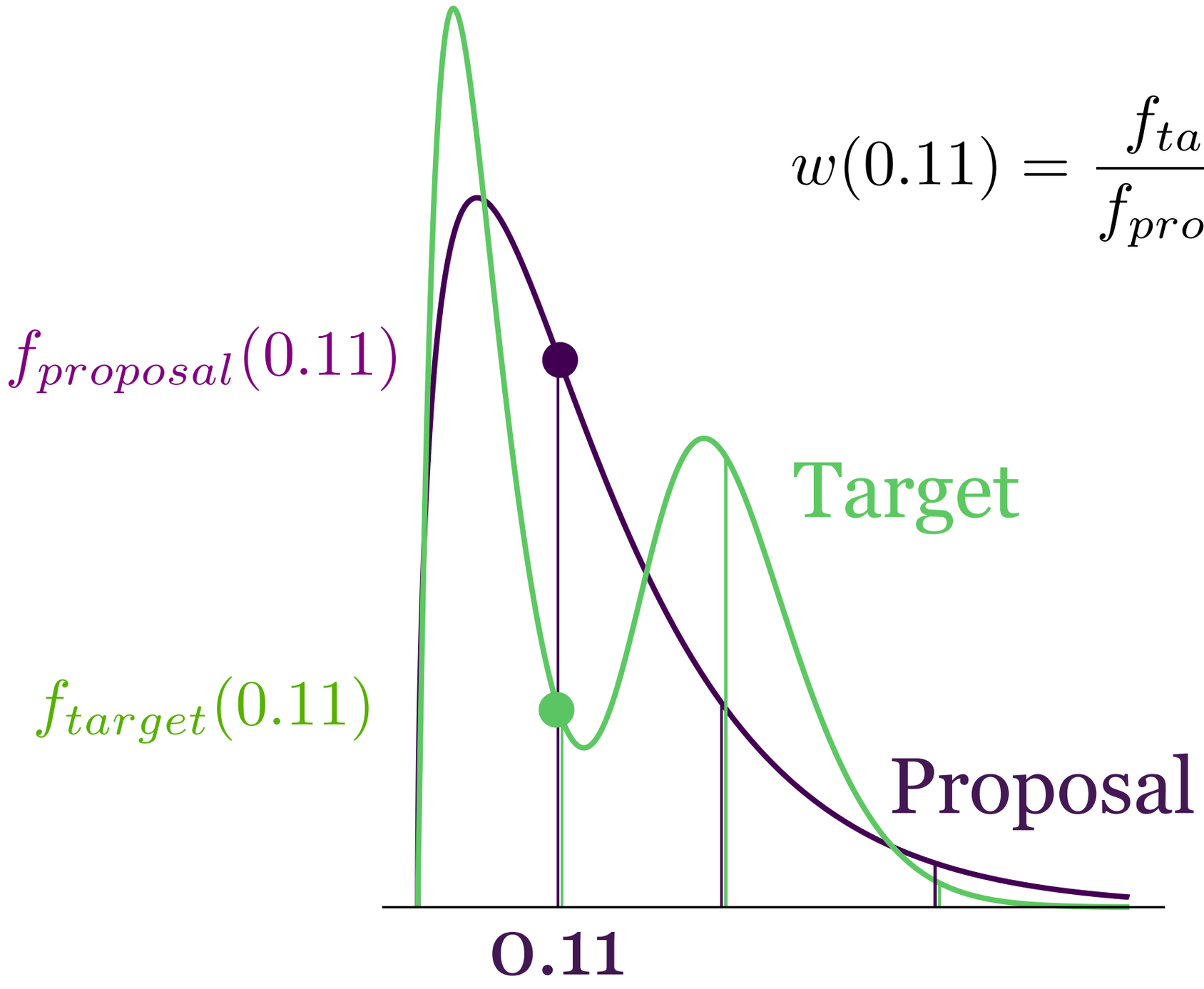


Importance sampling

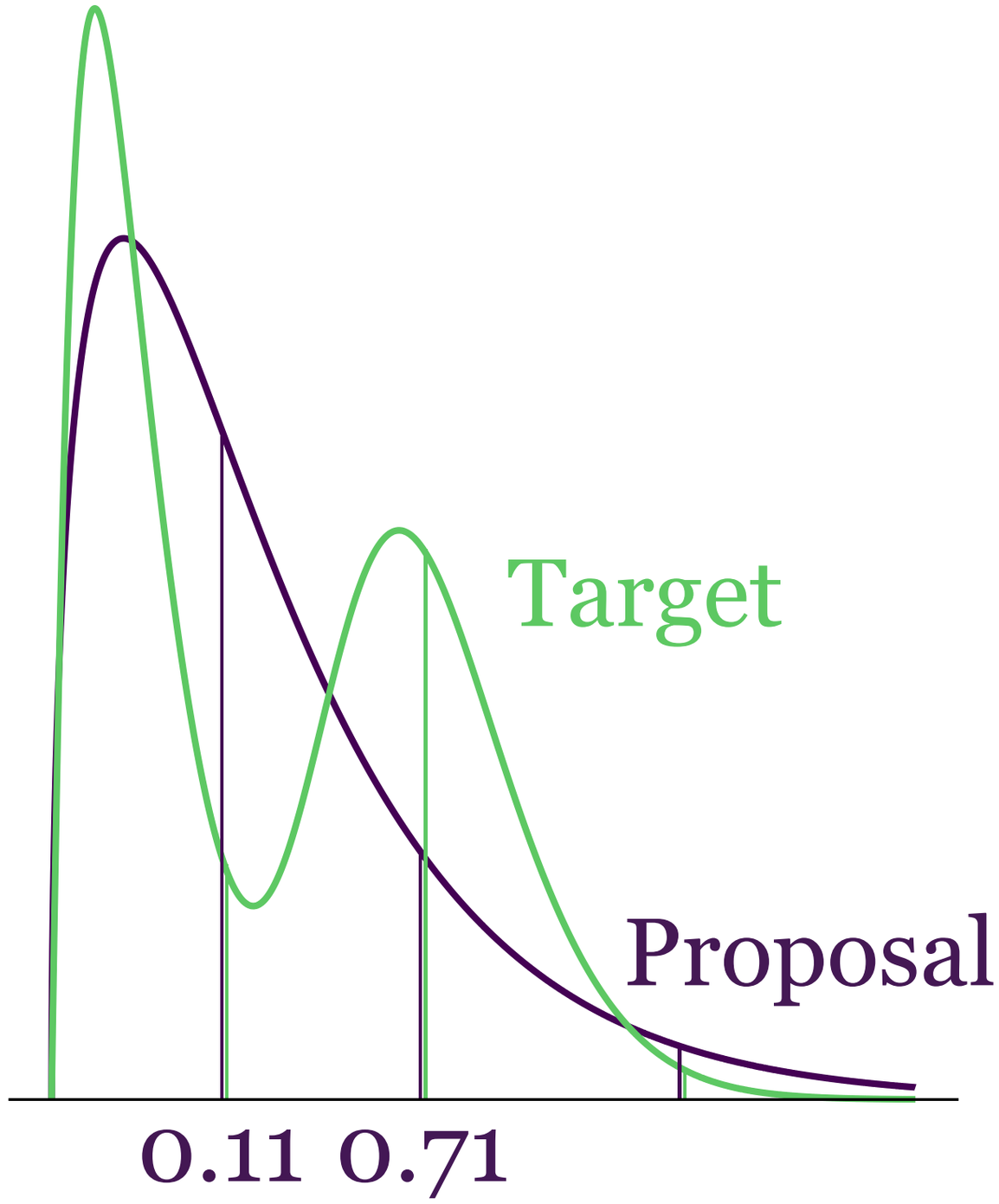


Importance sampling

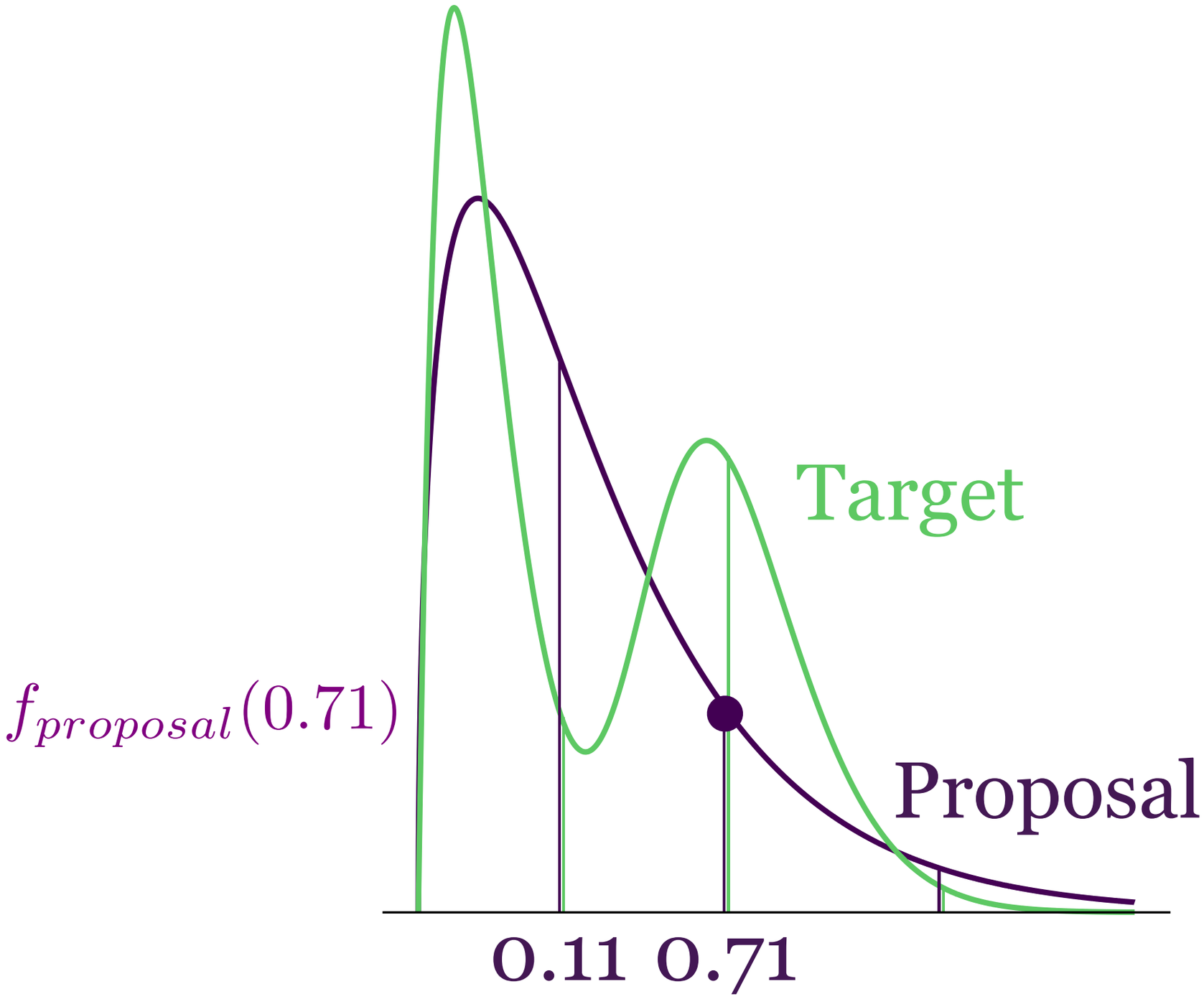
$$w(0.11) = \frac{f_{target}(0.11)}{f_{proposal}(0.11)}$$



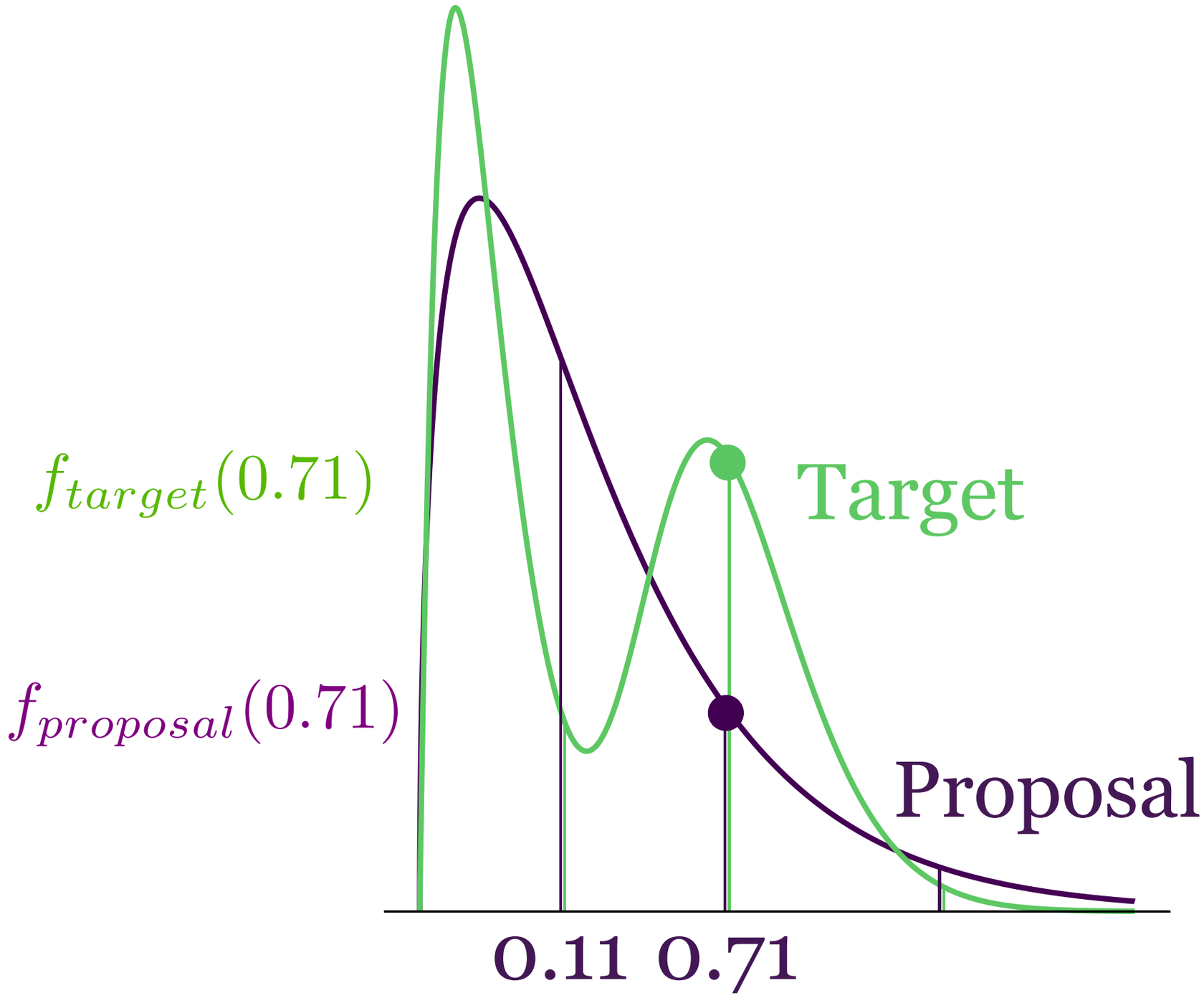
Importance sampling



Importance sampling

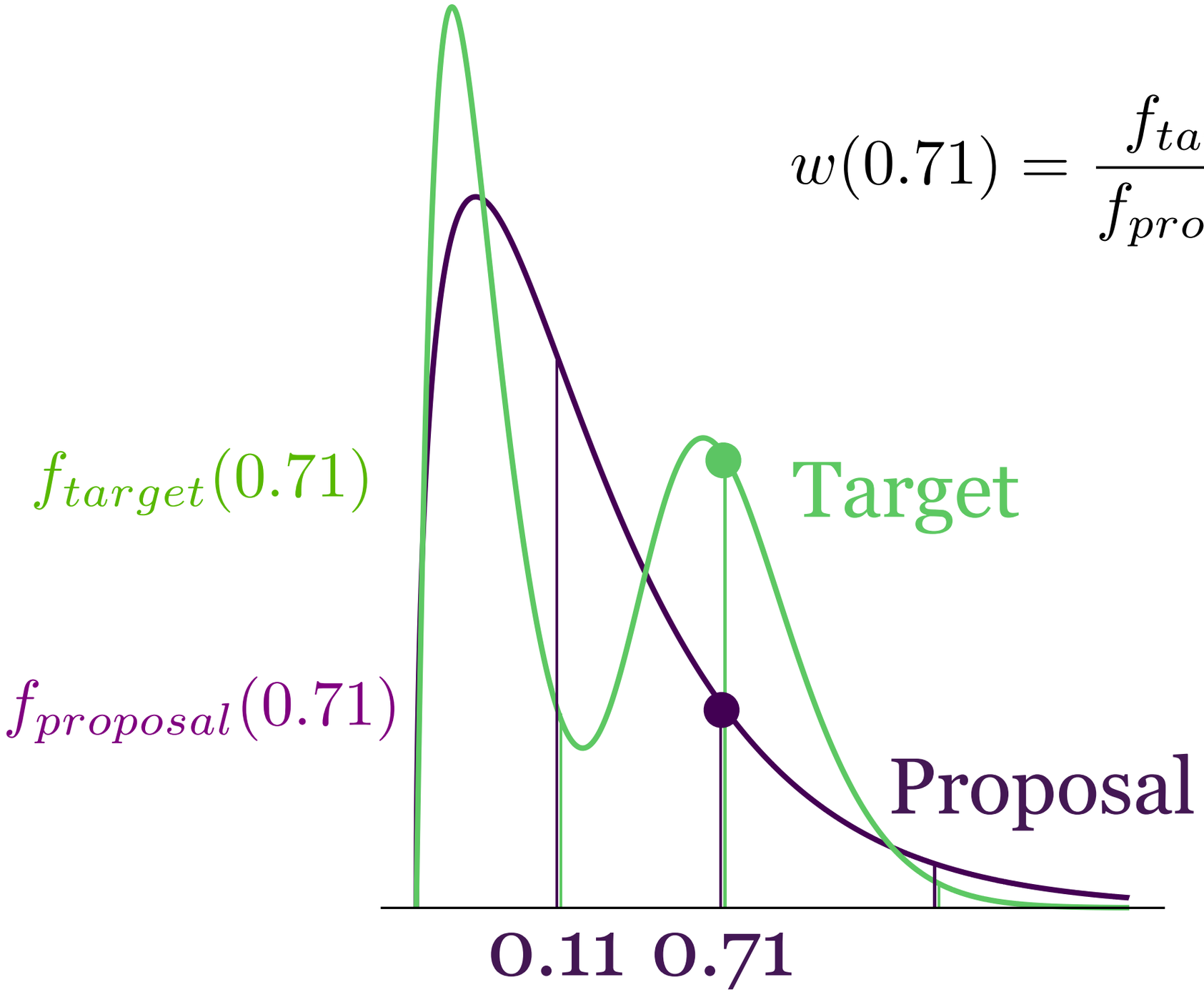


Importance sampling



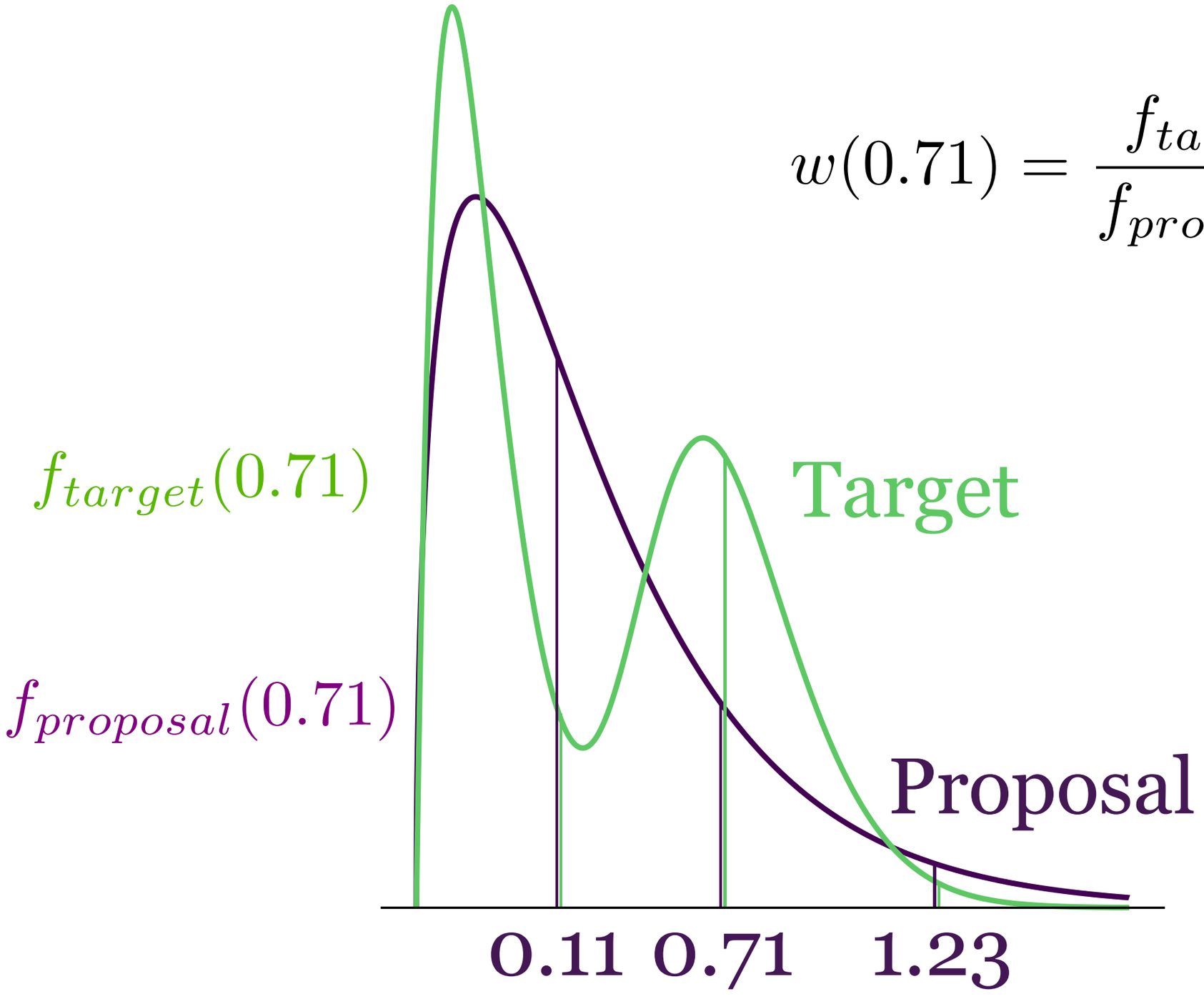
Importance sampling

$$w(0.71) = \frac{f_{target}(0.71)}{f_{proposal}(0.71)}$$

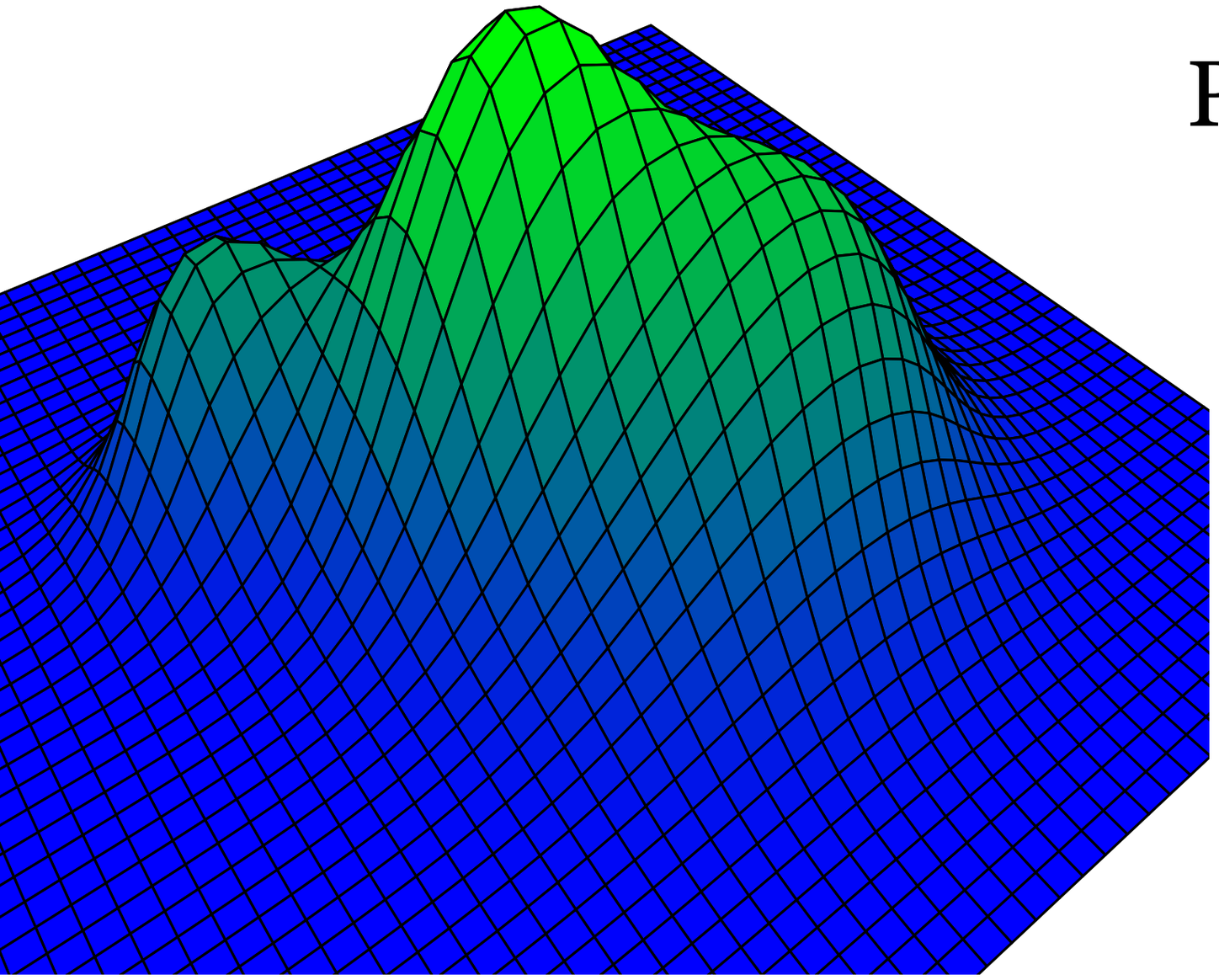


Importance sampling

$$w(0.71) = \frac{f_{target}(0.71)}{f_{proposal}(0.71)}$$



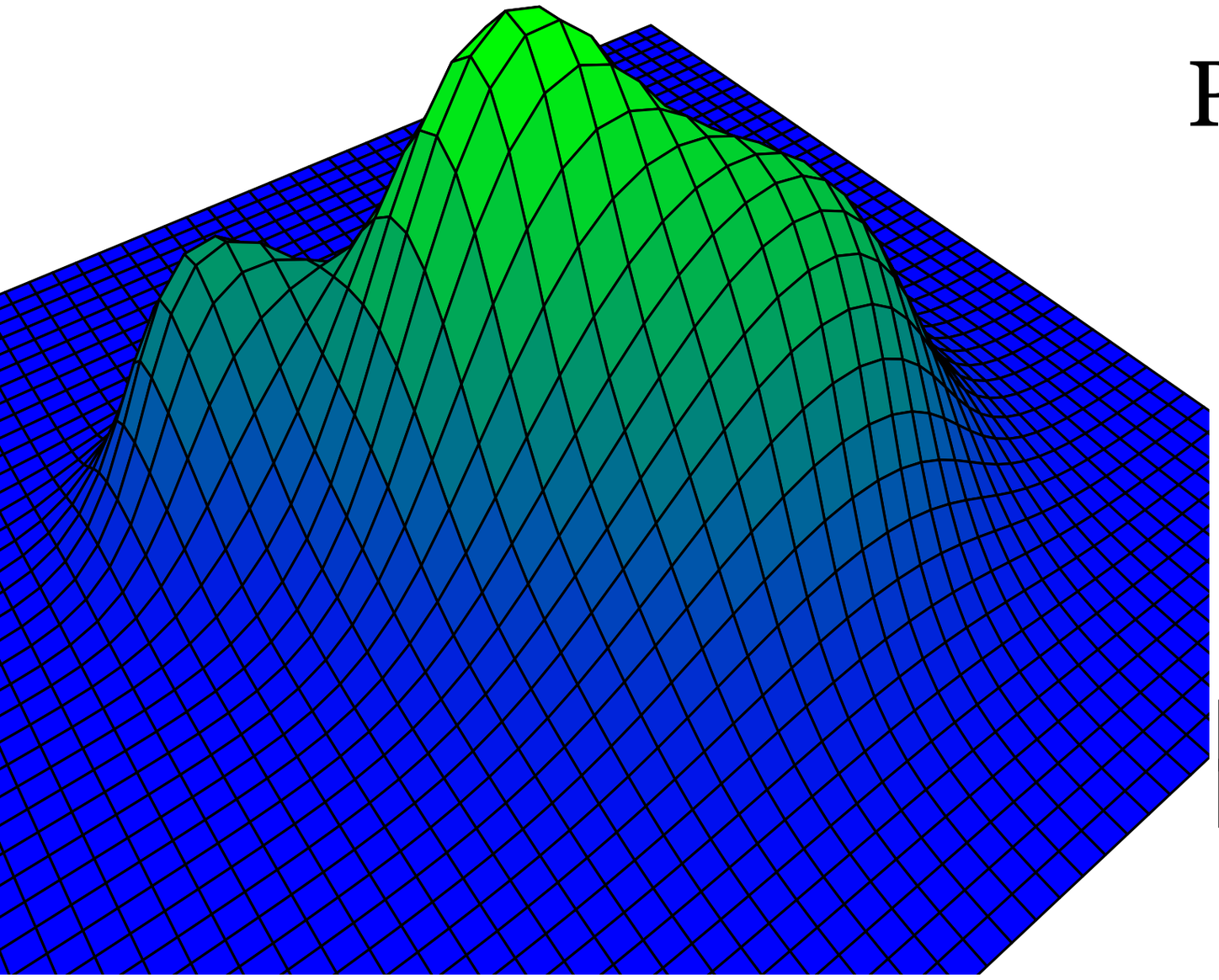
Importance sampling in phylogenetics



Target:
Posterior distribution

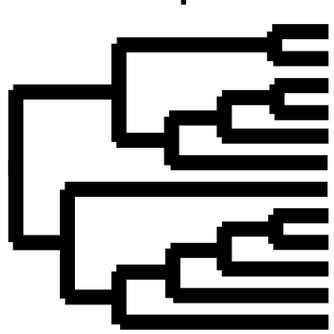
$$(T, t, Q)$$

Importance sampling in phylogenetics

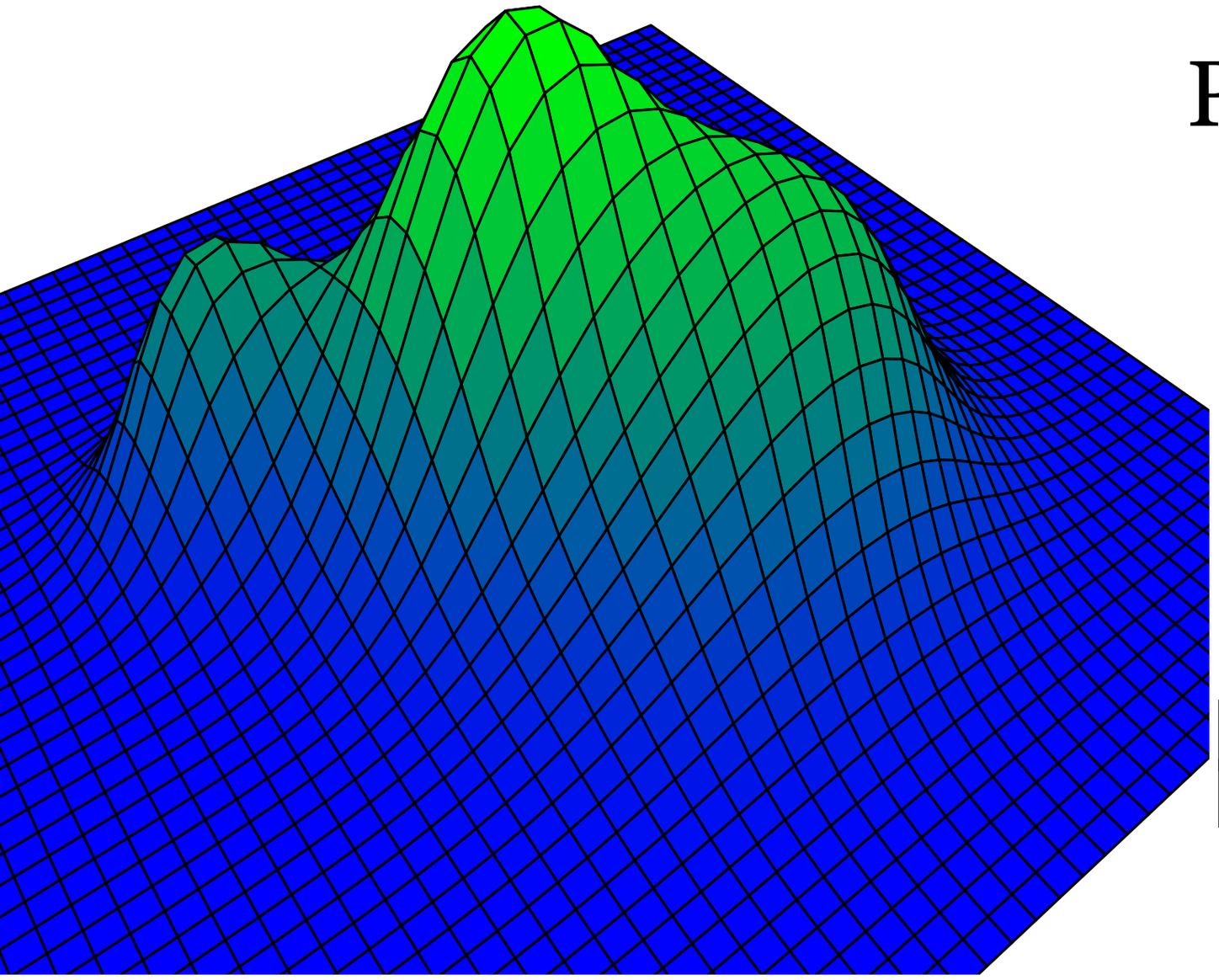


Target:
Posterior distribution

$$(T, t, Q)$$

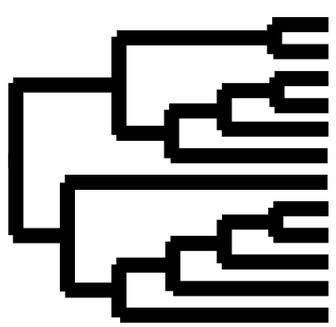


Importance sampling in phylogenetics



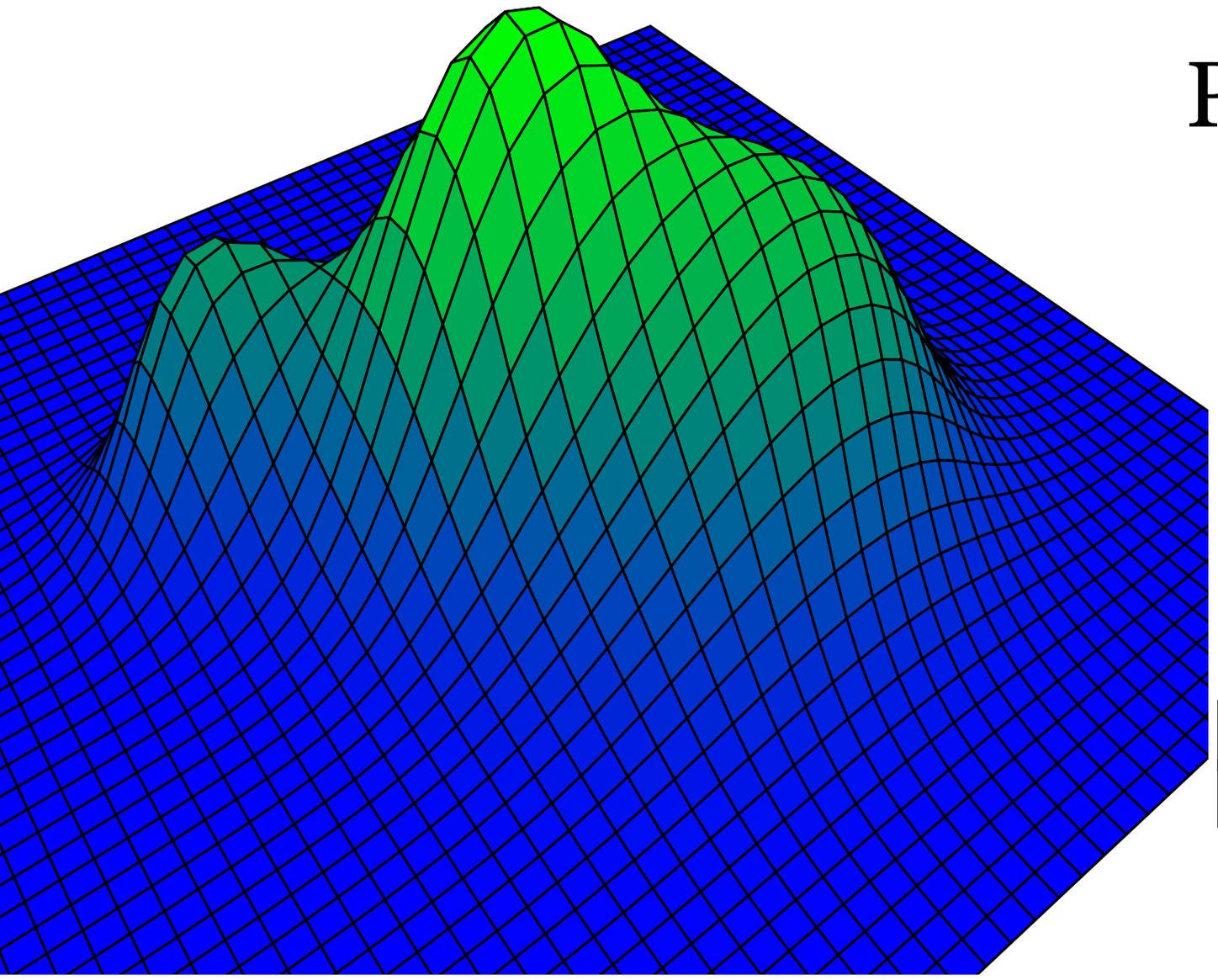
Target:
Posterior distribution

$$(T, t, Q)$$



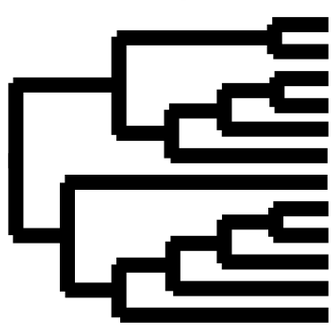
Branch lengths

Importance sampling in phylogenetics



Target:
Posterior distribution

$$(T, t, Q)$$



Branch lengths

Rate matrix

Importance sampling ingredients

Parameters: $\theta = (T, t, Q(\pi, s))$

Data: $X =$ DNA sequences

Target distribution: $p(\theta|X) =$ Posterior

$L(\theta|X) =$ Likelihood

Proposal distribution: $g(\theta|X) = g_1(Q)g_2(T|Q)g_3(t|T, Q)$

Importance sampling ingredients

Parameters: $\theta = (T, t, Q(\pi, s))$

Data: $X =$ DNA sequences

Target distribution: $p(\theta|X) =$ Posterior

$L(\theta|X) =$ Likelihood

Proposal distribution: $g(\theta|X) = g_1(Q)g_2(T|Q)g_3(t|T,Q)$

Proposal density for Q: $g_1(Q)$

Traditionally,

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim Dirichlet(a_1, a_2, a_3, a_4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim Dirichlet(b_1, b_2, b_3, b_4, b_5, b_6)$$

$$q_{ij} = \frac{s_{ij}}{2\pi_i}, q_{ii} = -\sum_j q_{ij}$$

Proposal density for Q: $g_1(Q)$

Traditionally,

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim Dirichlet(a_1, a_2, a_3, a_4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim Dirichlet(b_1, b_2, b_3, b_4, b_5, b_6)$$

$$q_{ij} = \frac{s_{ij}}{2\pi_i}, q_{ii} = -\sum_j q_{ij}$$

We propose a symmetric generalized Dirichlet:

(Craiu, 1969)

Proposal density for Q: $g_1(Q)$

Traditionally,

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim Dirichlet(a_1, a_2, a_3, a_4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim Dirichlet(b_1, b_2, b_3, b_4, b_5, b_6)$$

$$q_{ij} = \frac{s_{ij}}{2\pi_i}, q_{ii} = -\sum_j q_{ij}$$

We propose a symmetric generalized Dirichlet:

(Craiu, 1969)

$$(X_1, \cdots, X_k) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^k, \{\lambda_i\}_{i=1}^k)$$

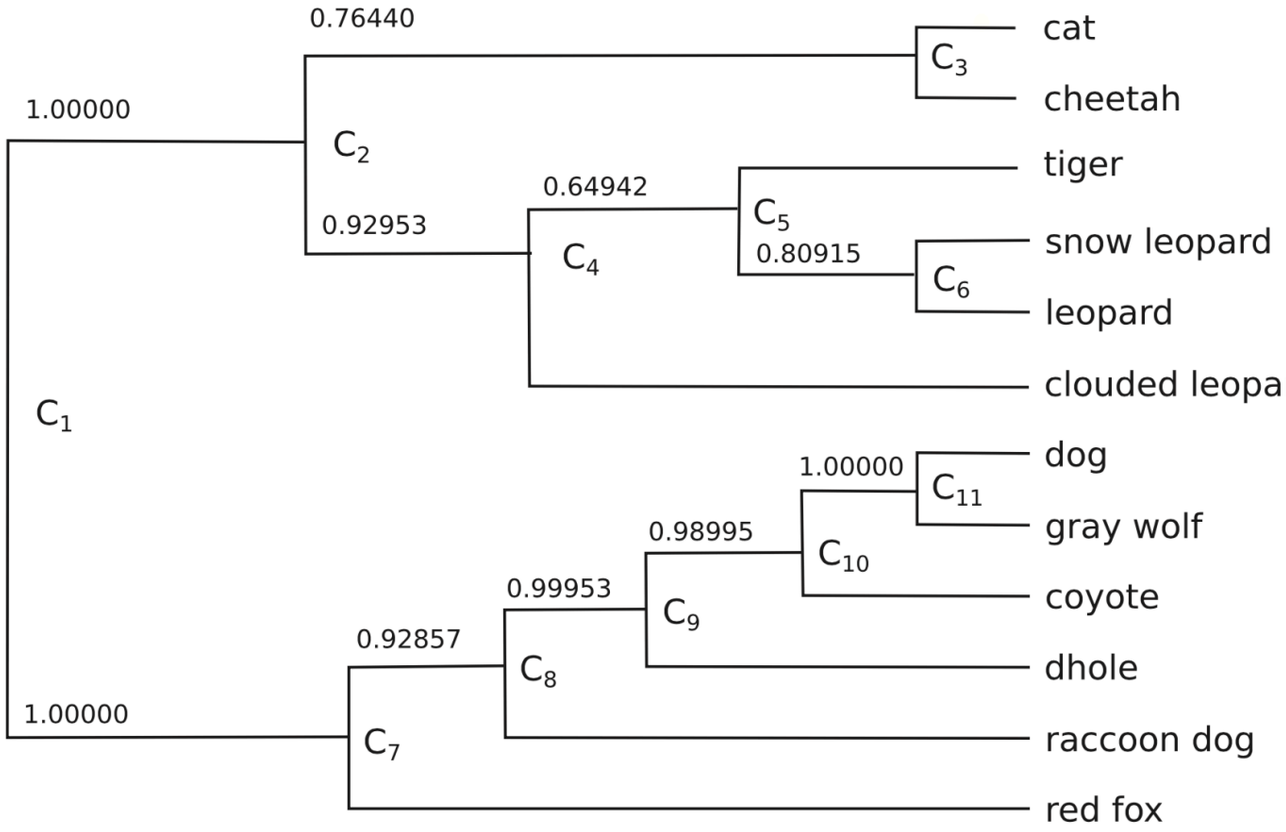
$$X_i = \frac{Y_i}{\sum_{j=1}^k Y_j}, Y_j \sim Gamma(\alpha_j, \lambda_j)$$

$$\text{with constraint } \sum_{j=1}^k \lambda_j = k$$

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

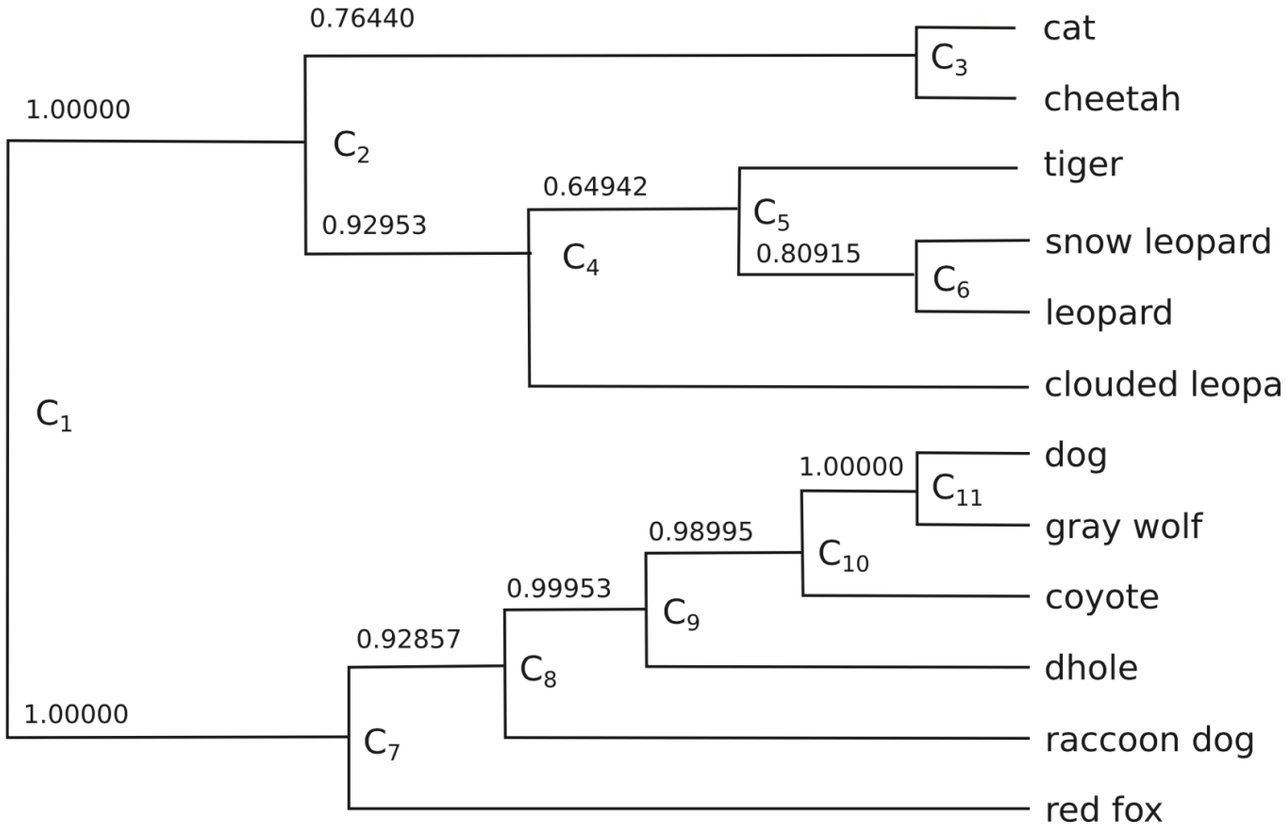
(Höhna, Drummond, 2012; Larget, 2013)



Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

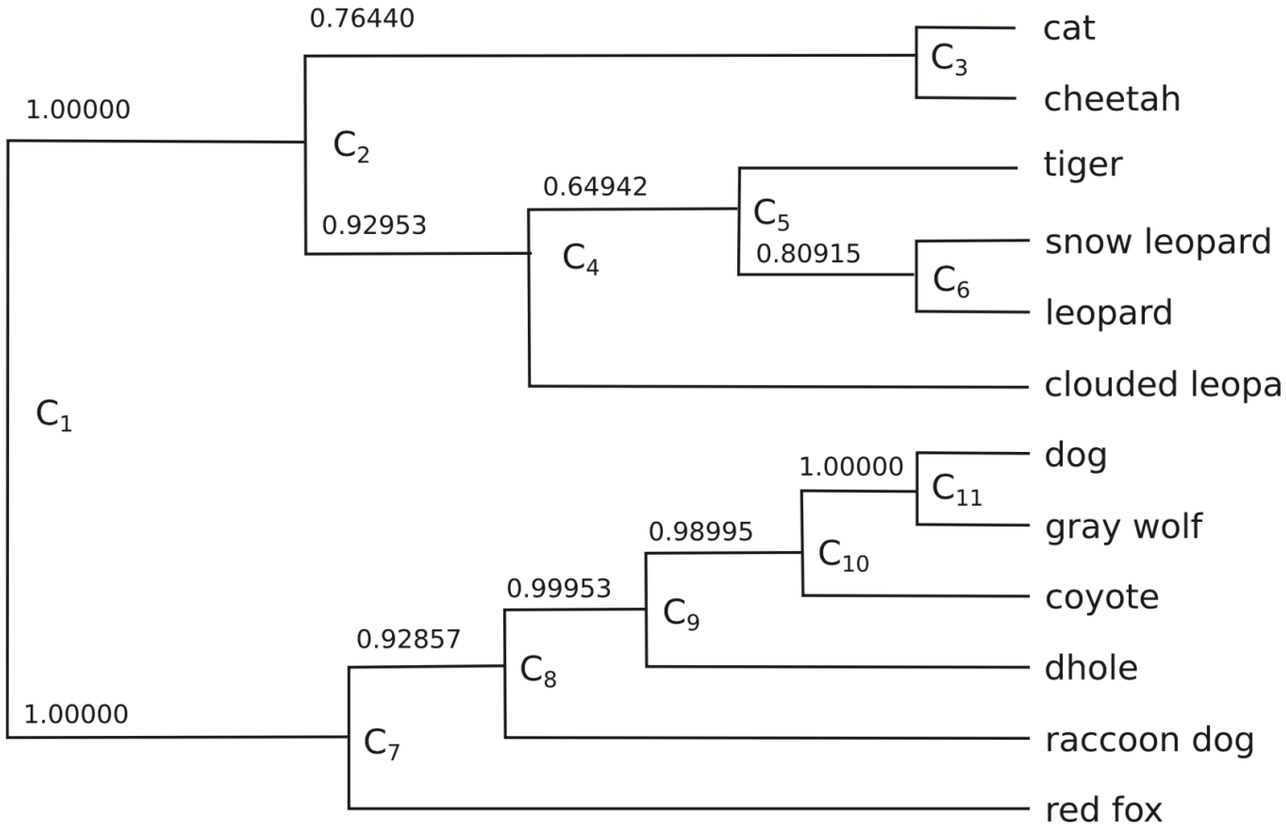


$$\mathrm{P}(T_1) = \mathrm{P}(C_2 \cap C_3 \cap \cdots \cap C_{11})$$

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

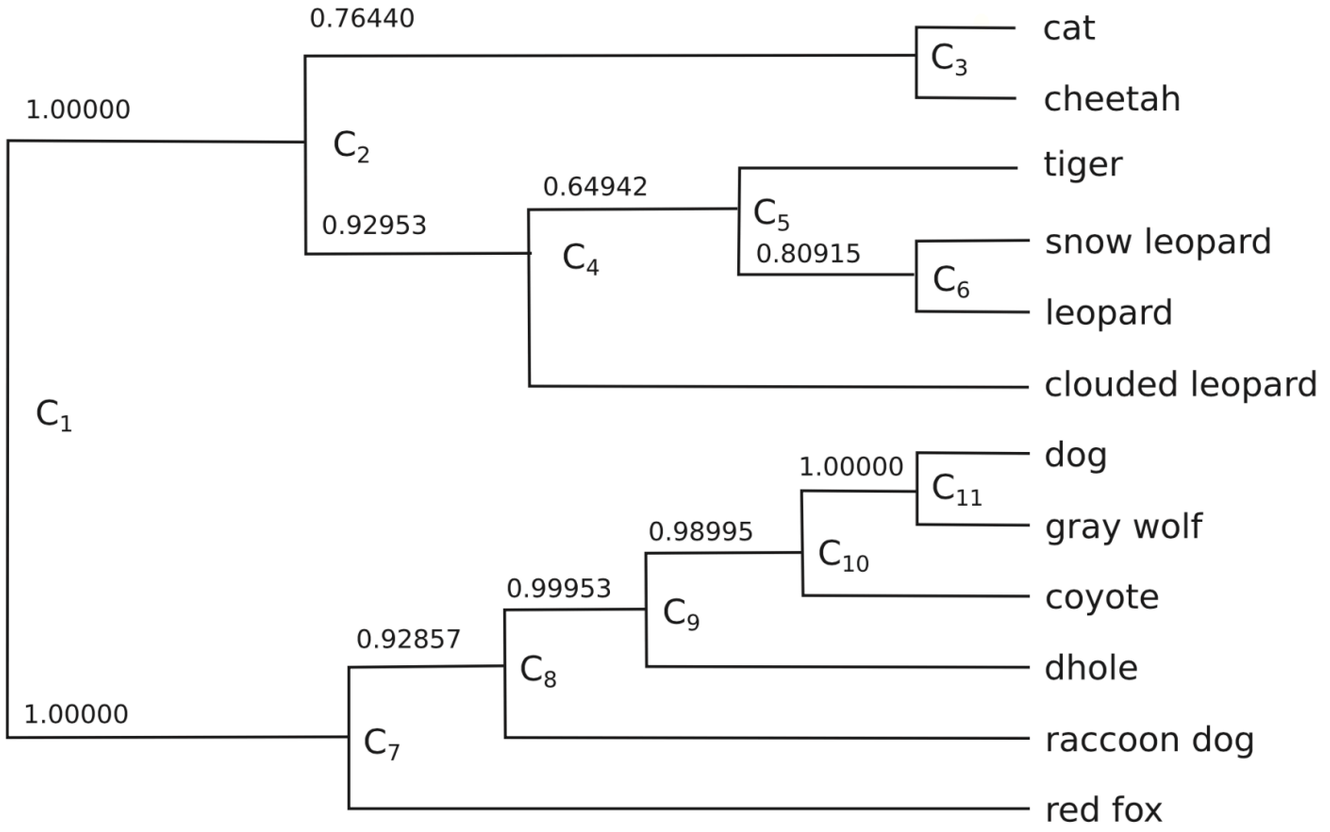


$$\begin{aligned}\mathsf{P}(T_1) = \mathsf{P}(C_2 \cap C_3 \cap \cdots \cap C_{11}) = & \mathsf{P}(C_2 \cap C_7) \\ & \times \mathsf{P}(C_3 \cap C_4 | C_2 \cap C_7) \\ & \times \mathsf{P}(C_5 | C_2 \cap C_3 \cap C_4 \cap C_7) \\ & \times \mathsf{P}(C_6 | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_7) \\ & \times \mathsf{P}(C_8 | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_6 \cap C_7) \\ & \times \mathsf{P}(C_9 | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_6 \cap C_7 \cap C_8) \\ & \times \mathsf{P}(C_{10} | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_6 \cap C_7 \cap C_8 \cap C_9) \\ & \times \mathsf{P}(C_{11} | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_6 \cap C_7 \cap C_8 \cap C_9 \cap C_{10}).\end{aligned}$$

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)



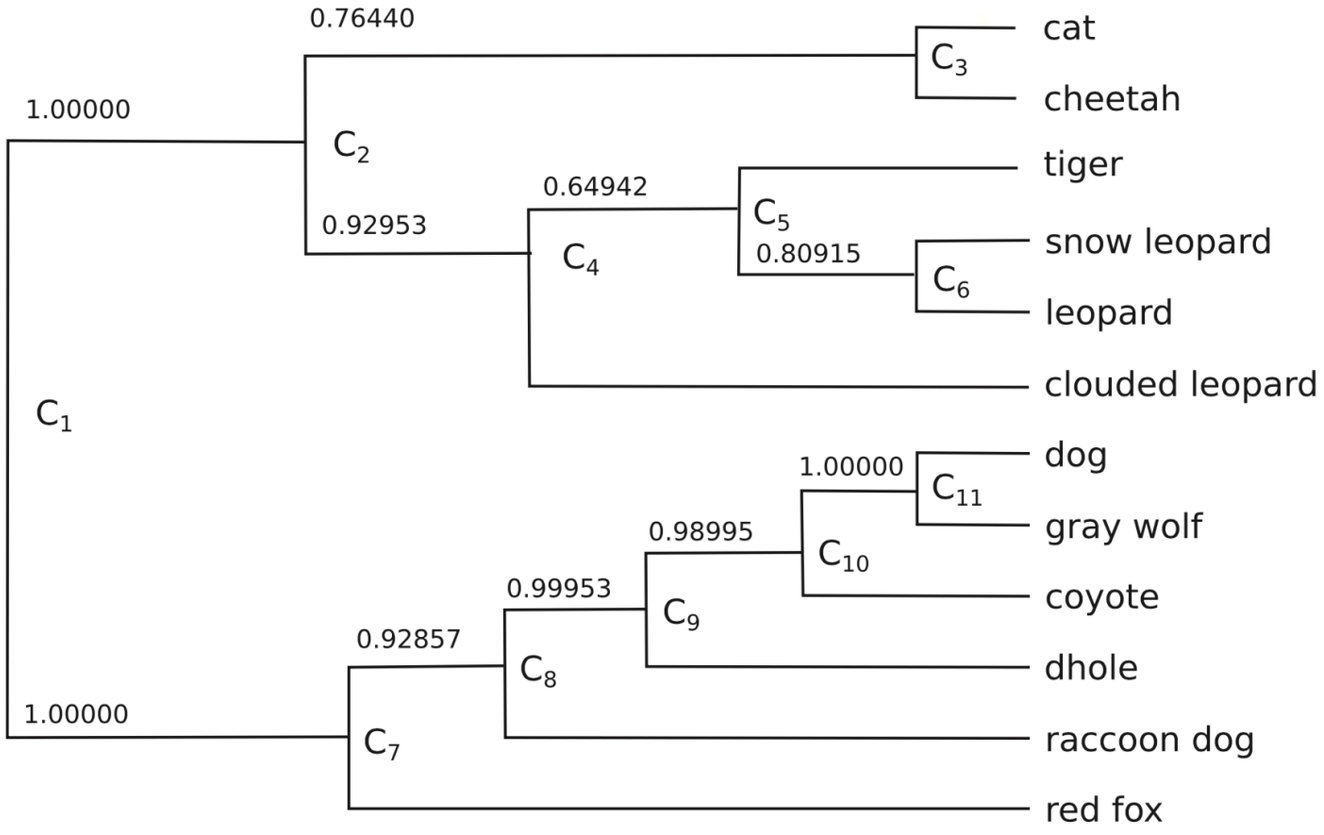
$$P(T_1) = P(C_2 \cap C_3 \cap \cdots \cap C_{11})$$

$$P(C_6 | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_7) \approx P(C_6 | C_5)$$

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)



$$\mathsf{P}(T_1) = \mathsf{P}(C_2 \cap C_3 \cap \cdots \cap C_{11})$$

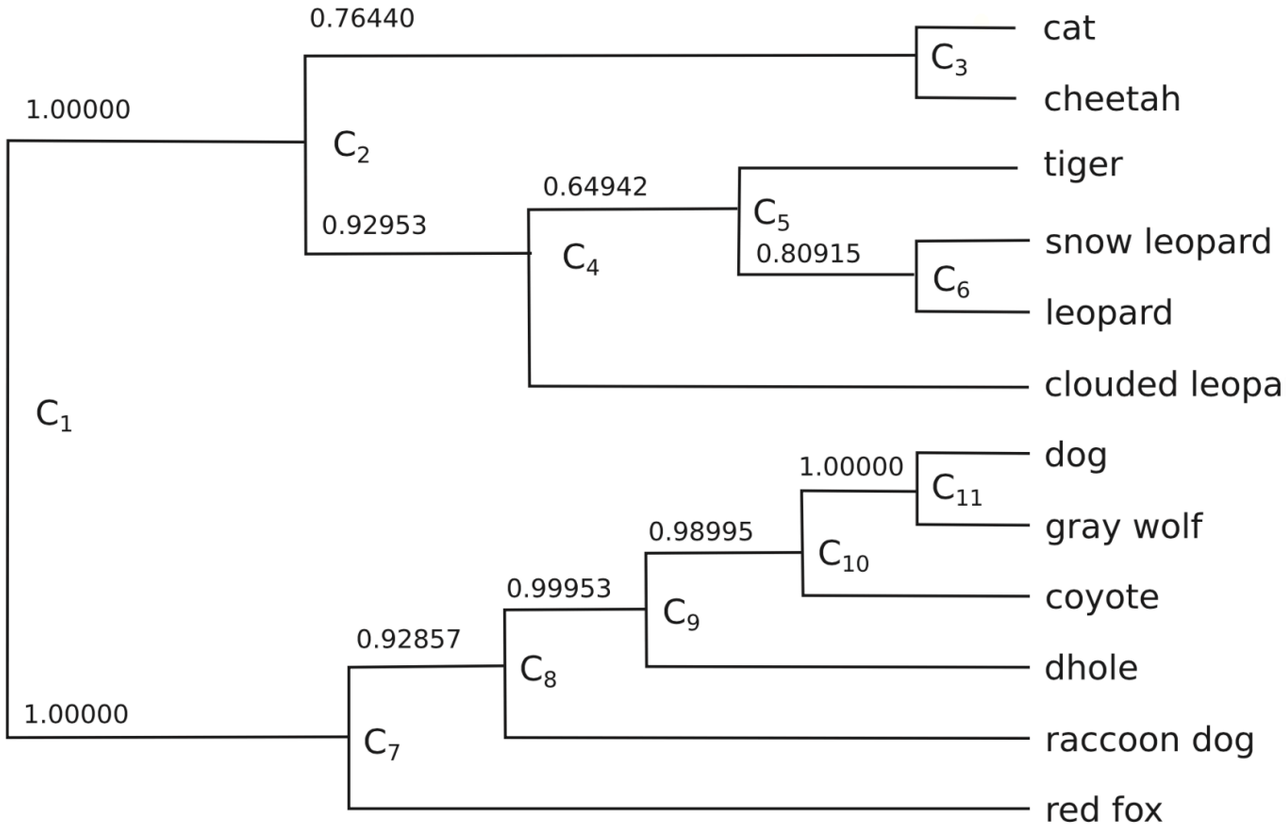
$$\mathsf{P}(C_6 | C_2 \cap C_3 \cap C_4 \cap C_5 \cap C_7) \approx \mathsf{P}(C_6 | C_5)$$

$$\mathsf{P}(T_1) \approx \mathsf{P}(C_2 \cap C_7)\mathsf{P}(C_3 \cap C_4 | C_2)\mathsf{P}(C_5 | C_4)\mathsf{P}(C_6 | C_5)$$
$$\times \mathsf{P}(C_8 | C_7)\mathsf{P}(C_9 | C_8)\mathsf{P}(C_{10} | C_9)\mathsf{P}(C_{11} | C_{10}).$$

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)



$$\mathsf{P}(T) \approx \prod_{\substack{C \in \text{ all clades of } T \\ |C|>1}} \mathsf{P}\left(L(C,T) \cap R(C,T) \middle| C\right).$$

Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

Proposal density for T: $g_2(T|Q)$

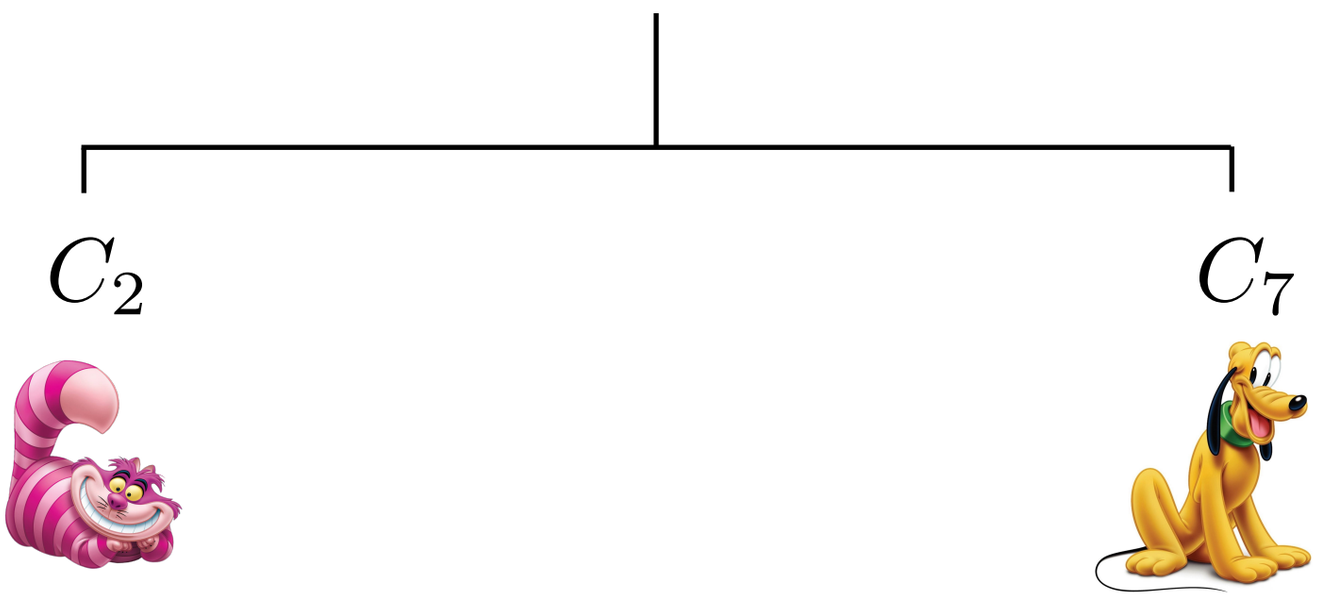
How to sample a tree from CCD?

$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$

Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

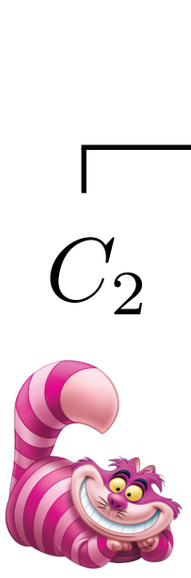
$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$



Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$

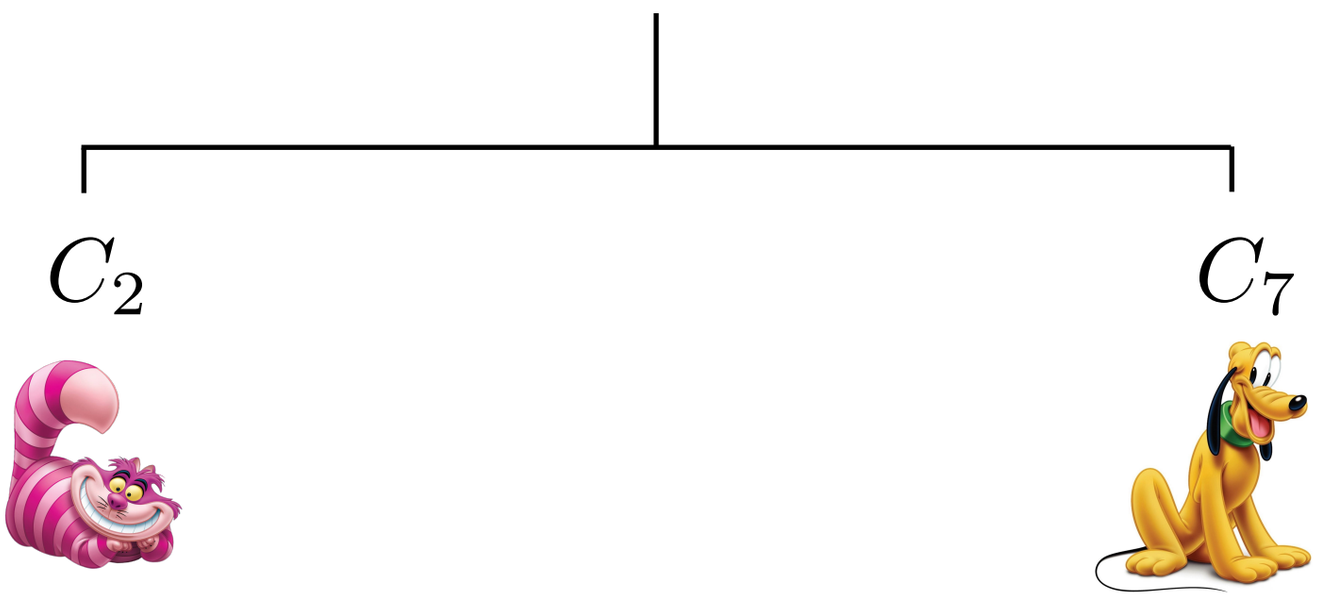


$$\mathsf{P}(\mathsf{C}_2 \cap \mathsf{C}_7) \approx 1$$

Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

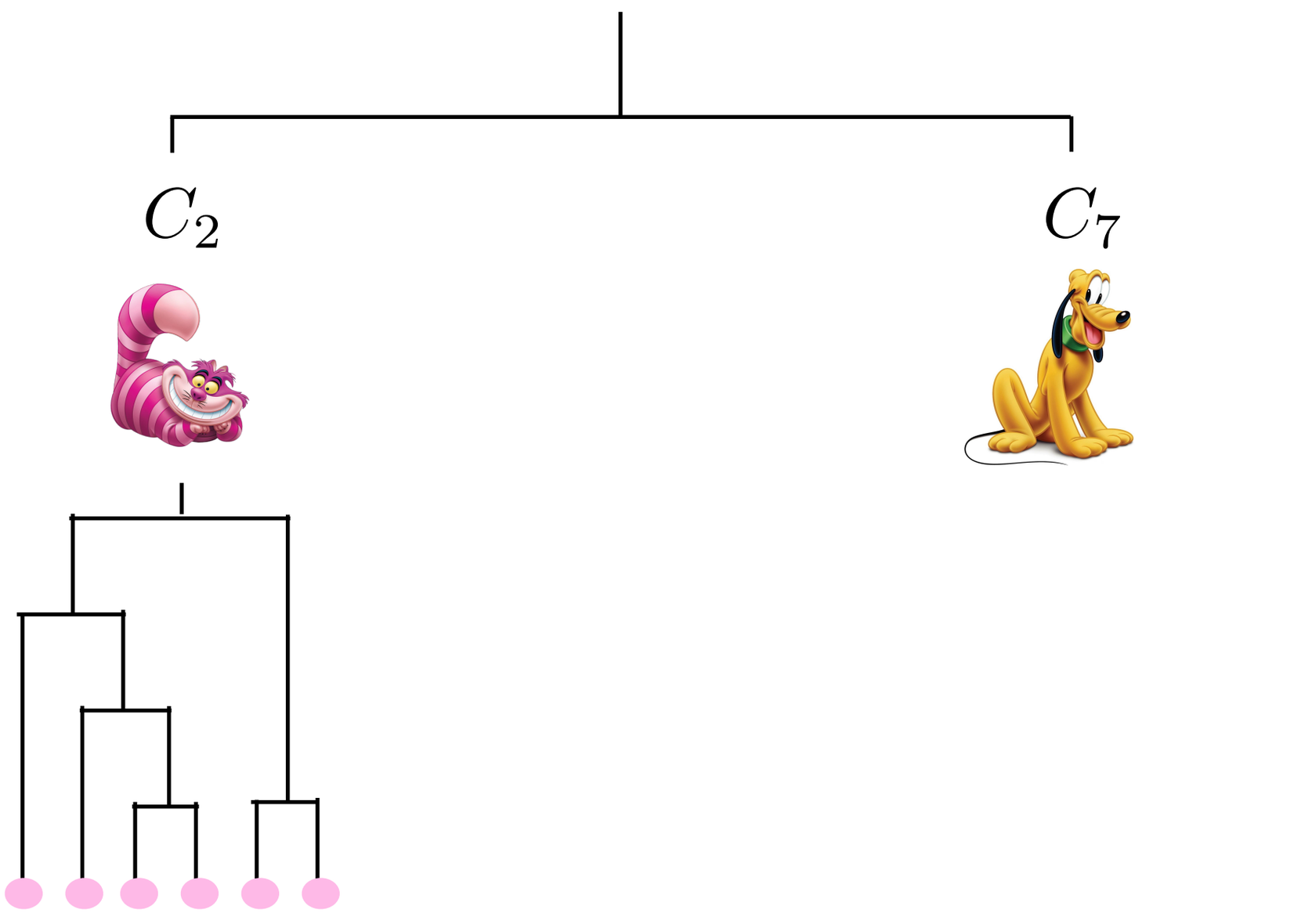
$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$



Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$

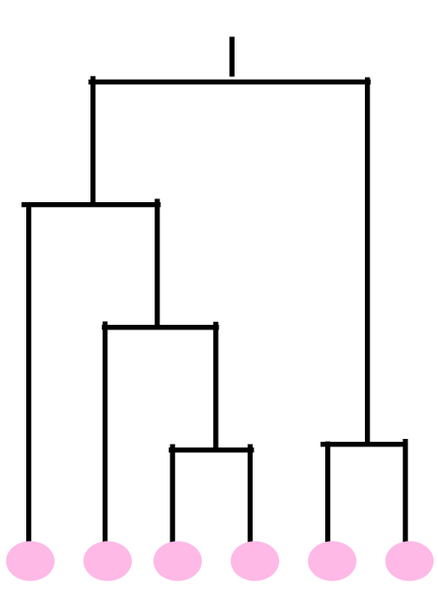


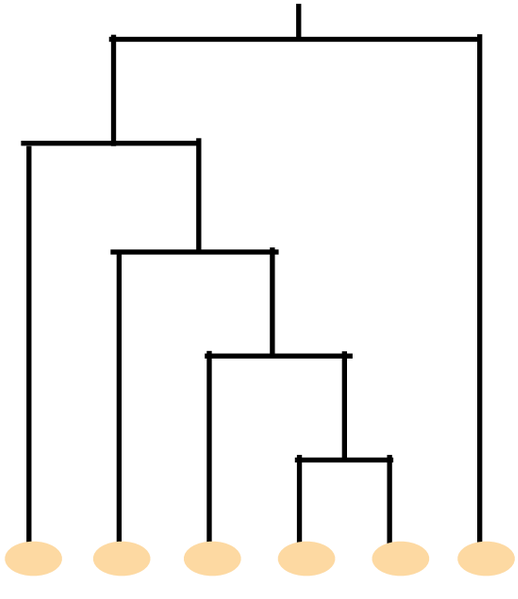
Proposal density for T: $g_2(T|Q)$

How to sample a tree from CCD?

$$C_1 = \{cat, cheetah, tiger, \cdots, dog, coyote\}$$

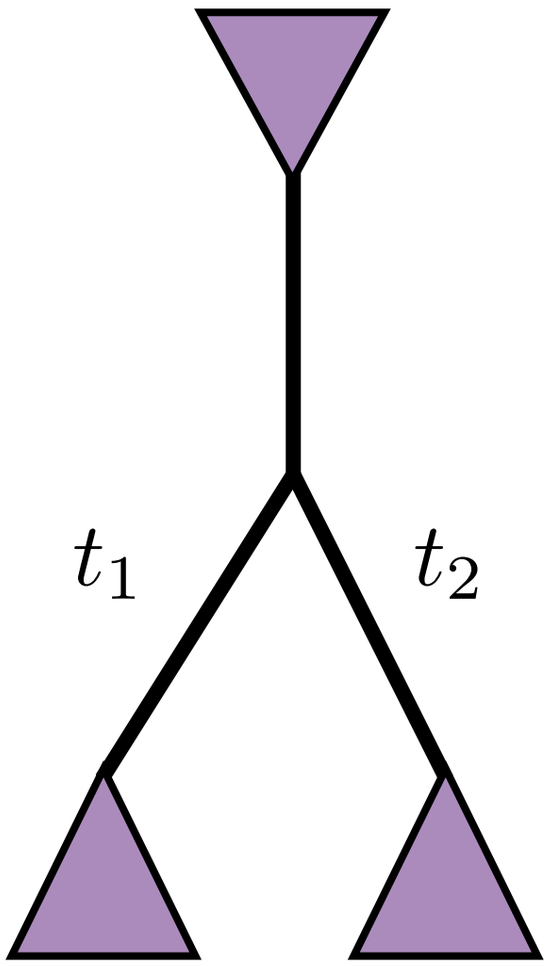
C_2 C_7





Proposal density for t: $g_3(t|T,Q)$

Correlation of sister edges



$$(t_1, t_2) \sim Gamma$$

$$\mu = MLE$$

$$\Sigma = I^{-1}$$

Importance sampling in phylogenetics: Bistro

- Sample Q from Symmetric Generalized Dirichlet
- Sample a topology (T) from clade distribution
- Sample branch lengths from Gamma
- Compute the likelihood of topology with branch lengths. Weight = target/proposal
- Repeat
- Do inference on weighted sample

Work in progress with B. Larget

Importance sampling in phylogenetics: Bistro

- Sample Q from Symmetric Generalized Dirichlet
- Sample a topology (T) from clade distribution
- Sample branch lengths from Gamma
- Compute the likelihood of topology with branch lengths. Weight = target/proposal
- Repeat
- Do inference on weighted sample

😱 Independent!

Work in progress with B. Larget

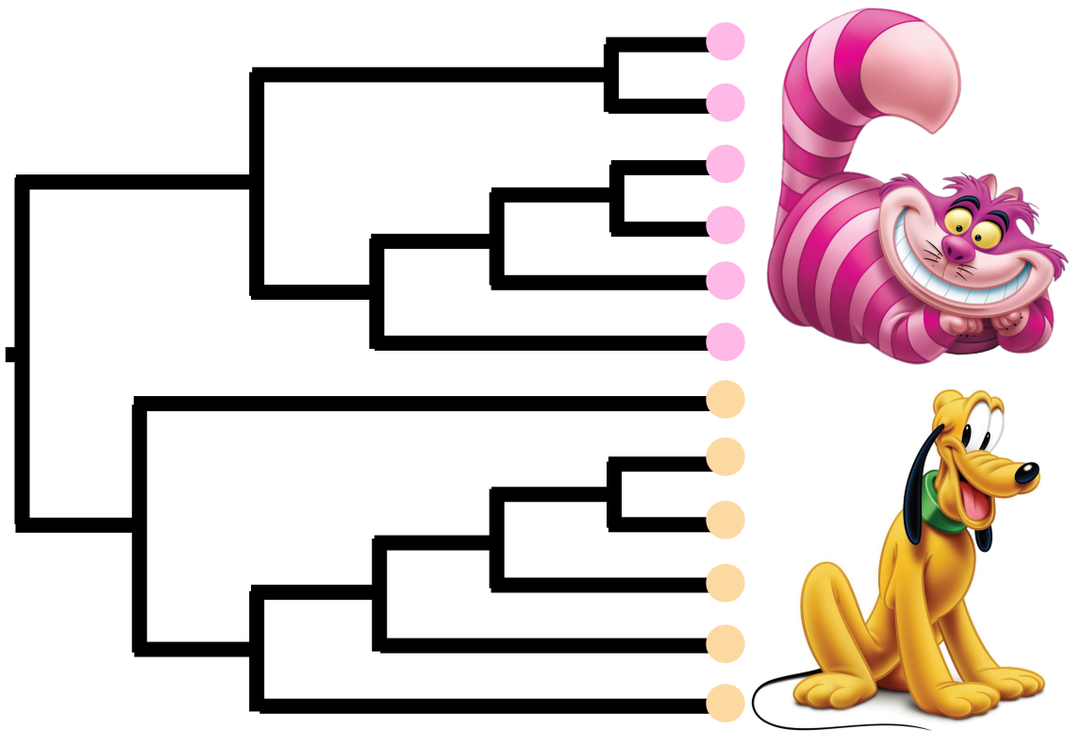
Importance sampling in phylogenetics: Bistro

- Sample Q from Symmetric Generalized Dirichlet
- Sample a topology (T) from clade distribution
- Sample branch lengths from Gamma
- Compute the likelihood of topology with branch lengths. Weight = target/proposal
- Repeat
- Do inference on weighted sample

😱 ~ Independent!

Work in progress with B. Larget

Results



	MCMC	Bistro
#Trees	1,000,000	1,000
ESS	250	116
Efficiency	0.025%	**12%**

Work in progress with B. Larget

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

Observed frequencies and pairwise counts

The devil is in the details…

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

✘ ~~Observed frequencies and pairwise counts~~

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

✘ ~~Observed frequencies and pairwise counts~~

MCMC sample on fixed tree

The devil is in the details…

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

- ~~Observed frequencies and pairwise counts~~
- ~~MCMC sample on fixed tree~~

Proposal density for Q: $g_1(Q)$

Symmetric generalized Dirichlet:

(Craiu, 1969)

$$(\pi_1, \pi_2, \pi_3, \pi_4) \sim S.G.Dirichlet(\{\alpha_i\}_{i=1}^4, \{\lambda_i\}_{i=1}^4)$$

$$(s_1, s_2, s_3, s_4, s_5, s_6) \sim S.G.Dirichlet(\{\tilde{\alpha}_i\}_{i=1}^6, \{\tilde{\lambda}_i\}_{i=1}^6)$$

need unbiased estimates

- ~~Observed frequencies and pairwise counts~~
- ~~MCMC sample on fixed tree~~
- Golden run MCMC

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

Bootstrap sample of NJ trees

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

weighted by parsimony

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

✗ ~~weighted by parsimony~~

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

✗ ~~weighted by parsimony~~

weighted by BHV distance to Frechet mean tree

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

✗ ~~weighted by parsimony~~

✗ ~~weighted by BHV distance to Frechet mean tree~~

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

✗ ~~Bootstrap sample of NJ trees~~

✗ ~~weighted by parsimony~~

✗ ~~weighted by BHV distance to Frechet mean tree~~

✓ Golden run MCMC

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

The devil is in the details…

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

will it be always biased?

The devil is in the details...

Proposal density for T: $g_2(T|Q)$

Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

need to estimate it

will it be always biased?

Quantifying MCMC Exploration of Phylogenetic Tree Space

CHRIS WHIDDEN* AND FREDERICK A. MATSEN IV

The devil is in the details...

Proposal density for T: $g_2(T|Q)$

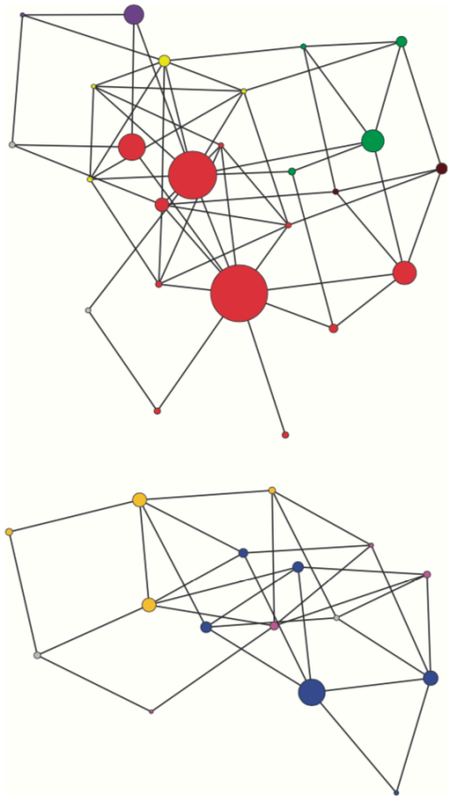
Conditional Clade Distribution (CCD)

(Höhna, Drummond, 2012; Larget, 2013)

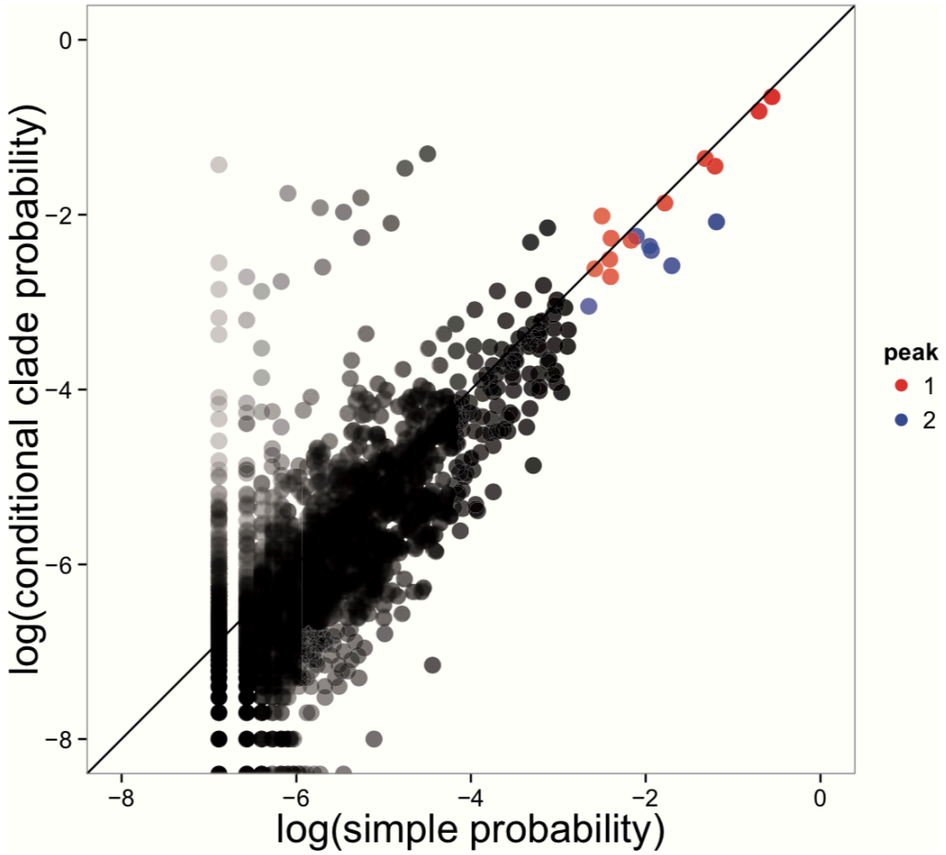
need to estimate it → will it be always biased?

Quantifying MCMC Exploration of Phylogenetic Tree Space

CHRIS WHIDDEN* AND FREDERICK A. MATSEN IV



DS1



a) DS1

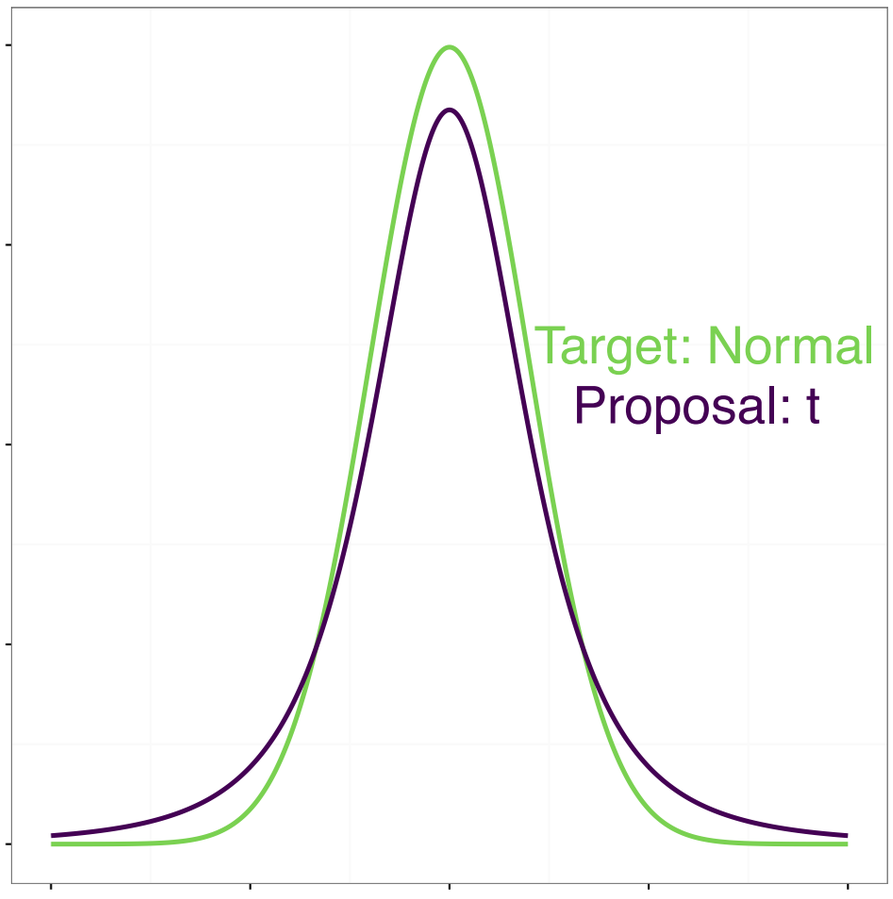
The devil is in the details...

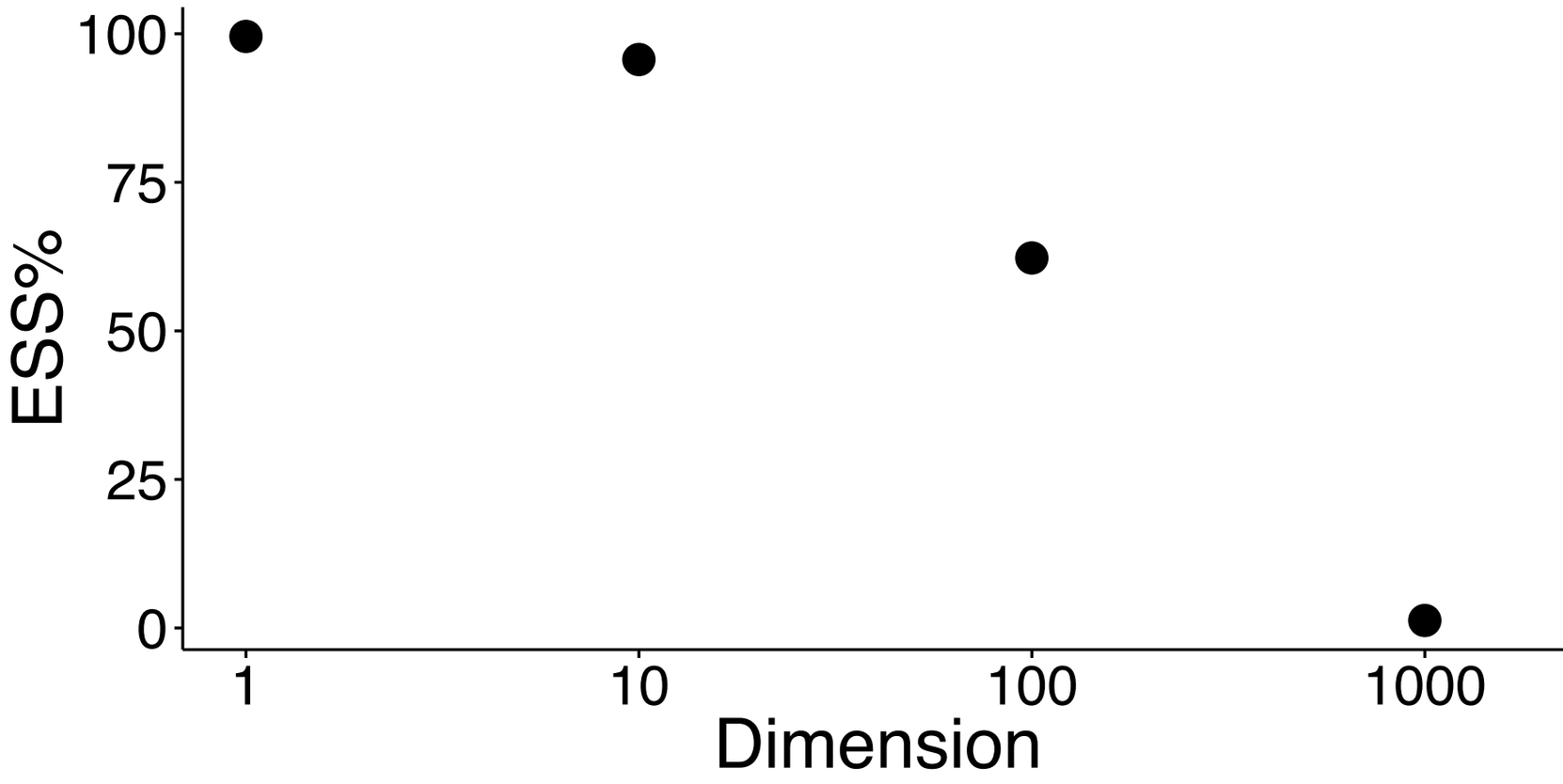
Proposal density for t: $g_3(t|T,Q)$

The devil is in the details...

Proposal density for t: $g_3(t|T,Q)$

Curse of dimensionality





Extra variance 10%

Summary: Bistro

- Optimistic results with importance sampling: **efficiency gains**
- **Serious challenges remain**: unbiased estimates for proposal distributions

Work in progress with B. Larget

Why is MCMC so slow?

1) Huge tree space size
2) Low acceptance of moves unless small neighborhood
3) Small neighborhood implies very dependent sample, which means small effective sample size

We need a gigantic chain because the space is huge and we are making tiny moves

Why is MCMC so slow?

Strengths of Bayesian inference:

Ability to combine information from multiple sources

More encompassing account of uncertainty

(Gelman et al, 2004)

We need a gigantic chain because the space is huge and we are making tiny moves

1) Huge tree space size
2) Low acceptance of moves unless small neighborhood
3) Small neighborhood implies very dependent sample, which means small effective sample size

Acknowledgements

Cécile Ané (UW)
Paul Bastide (KU-Leuven)
Bret Larget (UW)
Douglas Bates (UW)
David Baum (UW)
Sarah Friedrich (UW)
Michael Epstein (Emory)

John Malloy
John Spaw
Noah Stenz
Nan Ji
Jordan Vonderwell
Josh McGrath

http://crsl4.github.io/