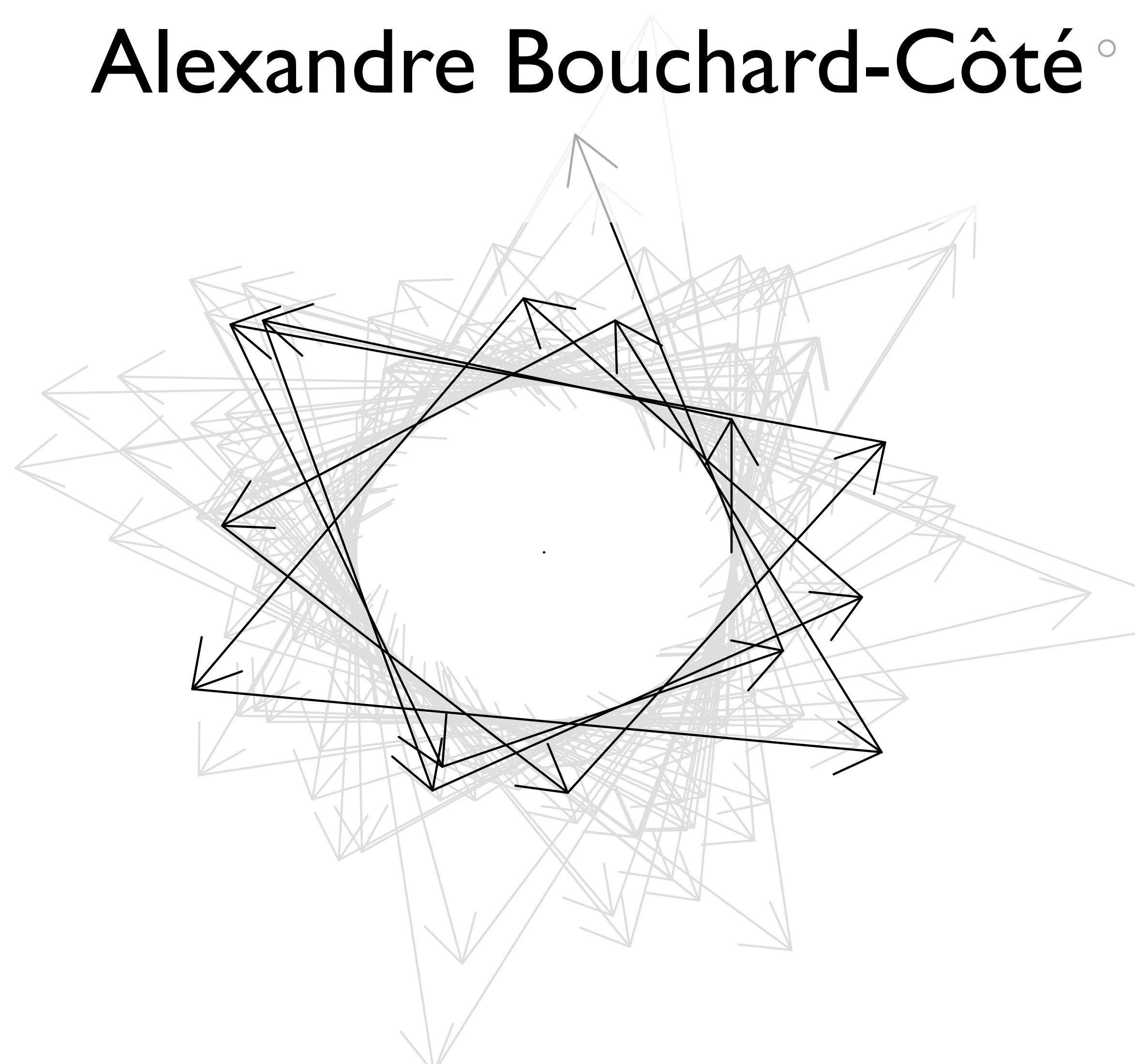
Monte Carlo methods

Alexandre Bouchard-Côté



Def. 1

Computing difficult integrals using the law of large numbers

Goal: Compute $I = \int_{\mathbb{X}} \phi(x)\, \pi(x)\, dx$

Monte Carlo Methods: Truncate a *Law of Large Numbers* (LLN) converging to I.

Example (simple Monte Carlo):

- Simulate independent $X_1, ..., X_n$ from π.
- Return $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i)$.

Goals of this course

- Motivate why we need Monte Carlo (MC) methods.
- For the *users* of MC:
  - Using probabilistic programming languages.
  - Overview of the modern MC toolbox.
    We will iterate:
    - Problem we can't solve with what we know so far.
    - Methodology that solves this problem.
- How theory helps you write **correct** + efficient MC.
- For the *developers* of new MC methods:
  - implementation and analysis of state-of-the-art MC algorithms.
  - Interested in what you work on to guide choice of topics.

Logistics

- Course webpage: http://tinyurl.com/starferret
    - Check the syllabus for references, assessment, info on final project and more
- My office hour: after lecture?
- TA: Sohrab Salehi. Office hour TBA.

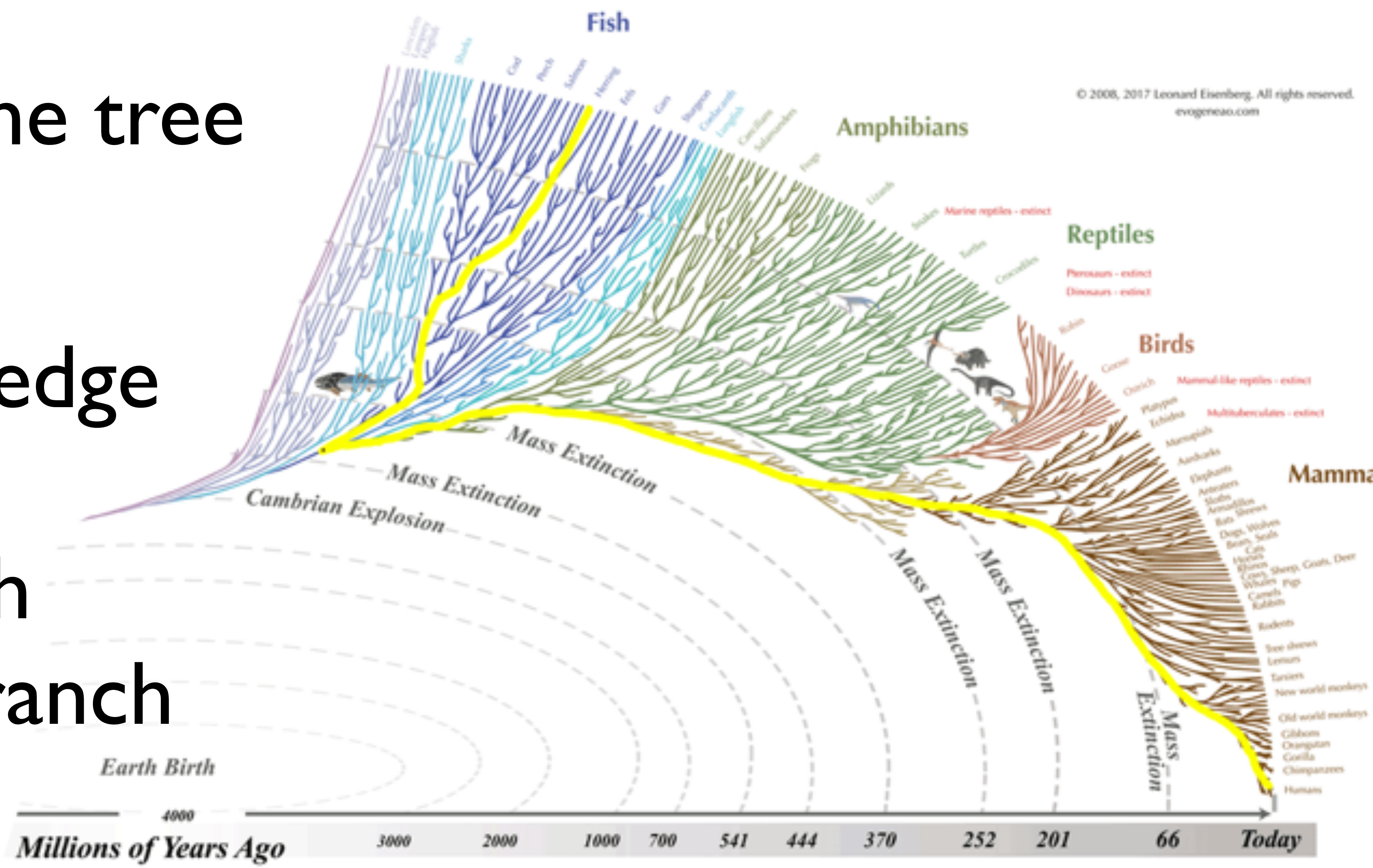
Logistics

- BRING LAPTOP IN CLASS! (Ask about windows users)
- Languages supported: Xtend/Java and Python
- Feel free to use other ones but strongly recommend learning a compiled language if serious about development of MC methods
- Install (use Piazza if encounter problems):
  - git (see https://software-carpentry.org/)
  - R
  - Oracle Java 8
  - PPLs:
    - Stan: http://mc-stan.org/
    - Blang: https://www.stat.ubc.ca/~bouchard/blang/Blang_IDE.html

Motivation: phylogenetic

Ex. 2

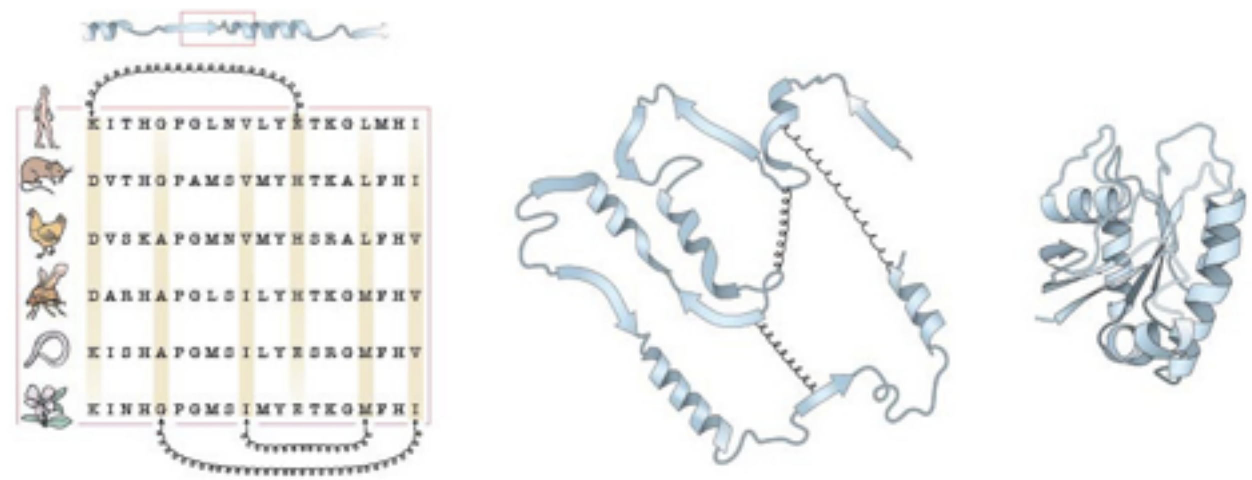
- *Topology:* shape of the tree (discrete)
- *Branch length:* each edge has a parameter encoding how much evolution on the branch (continuous)



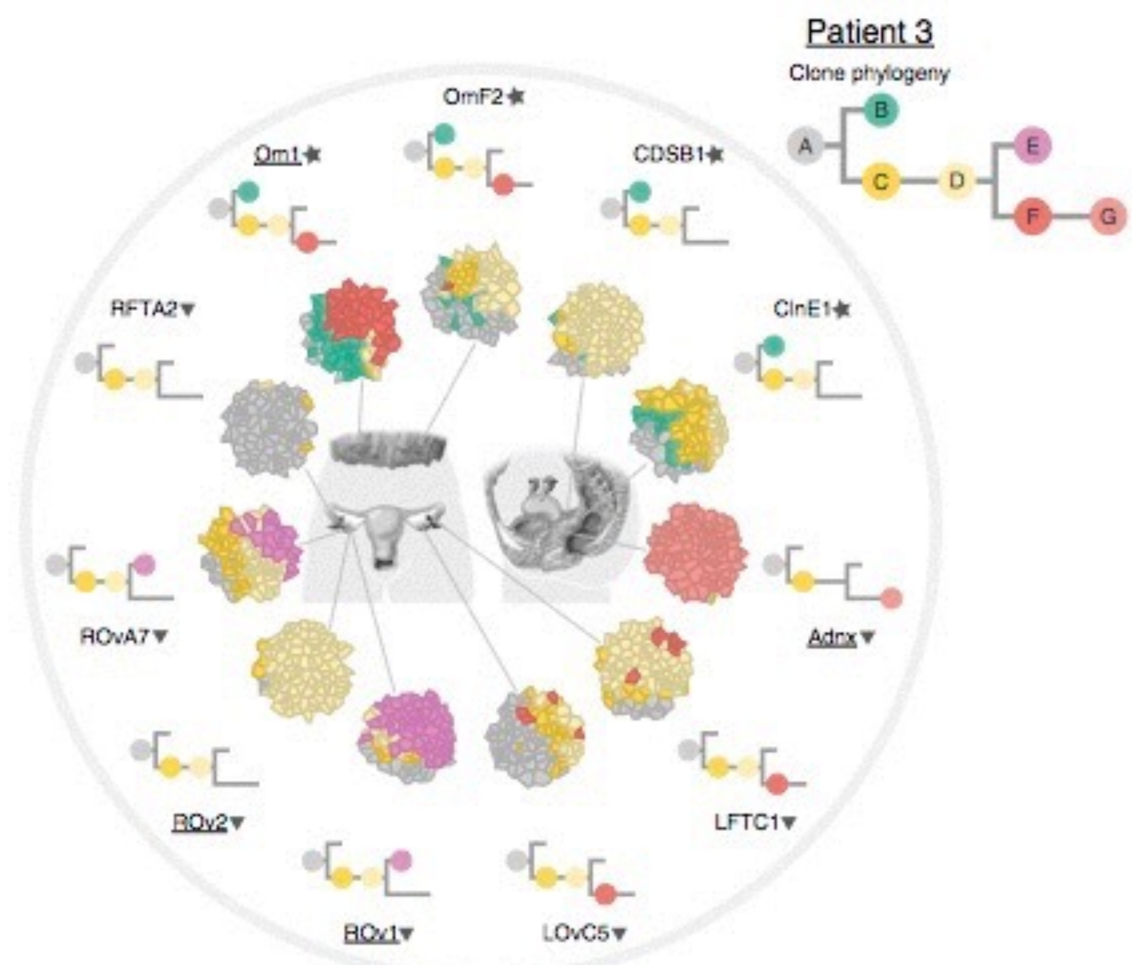
- Data: DNA at leaves (for example)
- Goal: reconstruct the tree (and more! ancestral sequences, models, ...)

Why phylo is important

- Signature left by evolution key to understand function

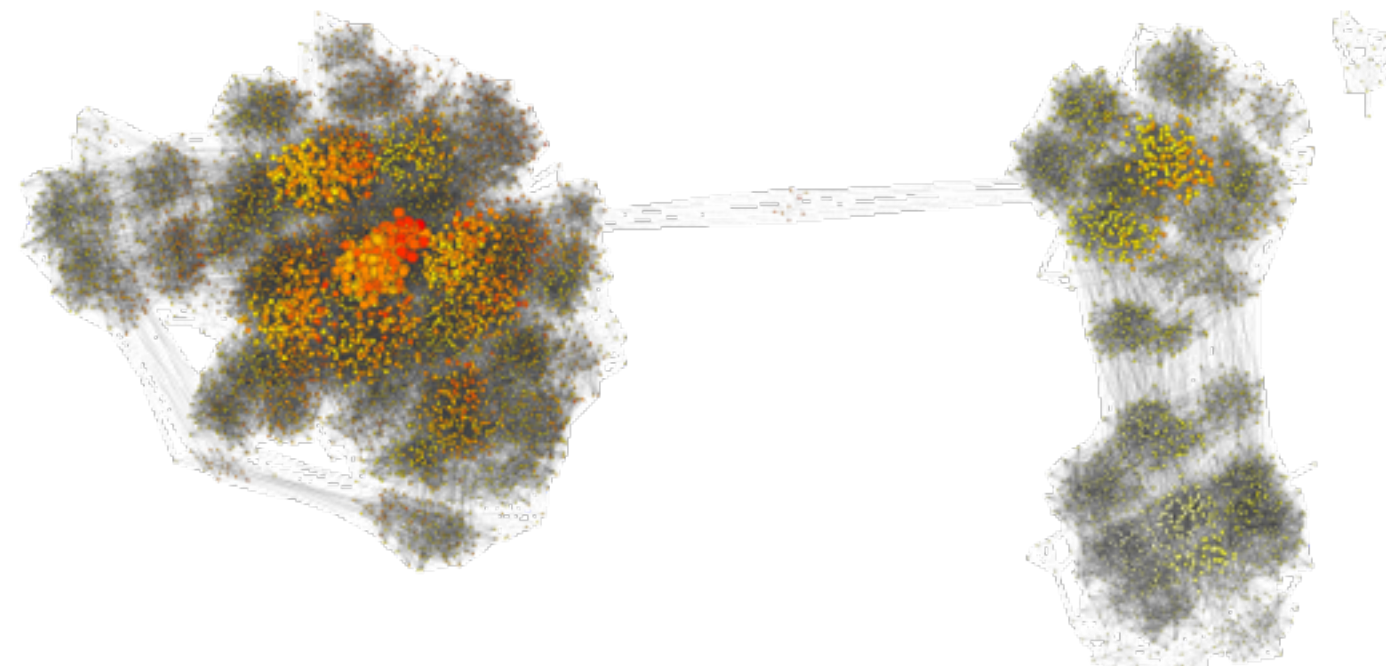


- Where do we come from?
- Beyond the tree of life:
  - cancer phylogenetics
  - trees of languages



Challenges

- Trees space:
  - high-dimensional,
  - combinatorial,
  - non-convex



- Uncertainty does not vanish in asymptotic regime of interest
  (sequence length does *not* go to infinity)
  - In many scientific applications, it is critical to quantify this uncertainty

Optimization and simulation

x : unknown
y : data

Cases where this is advantageous...

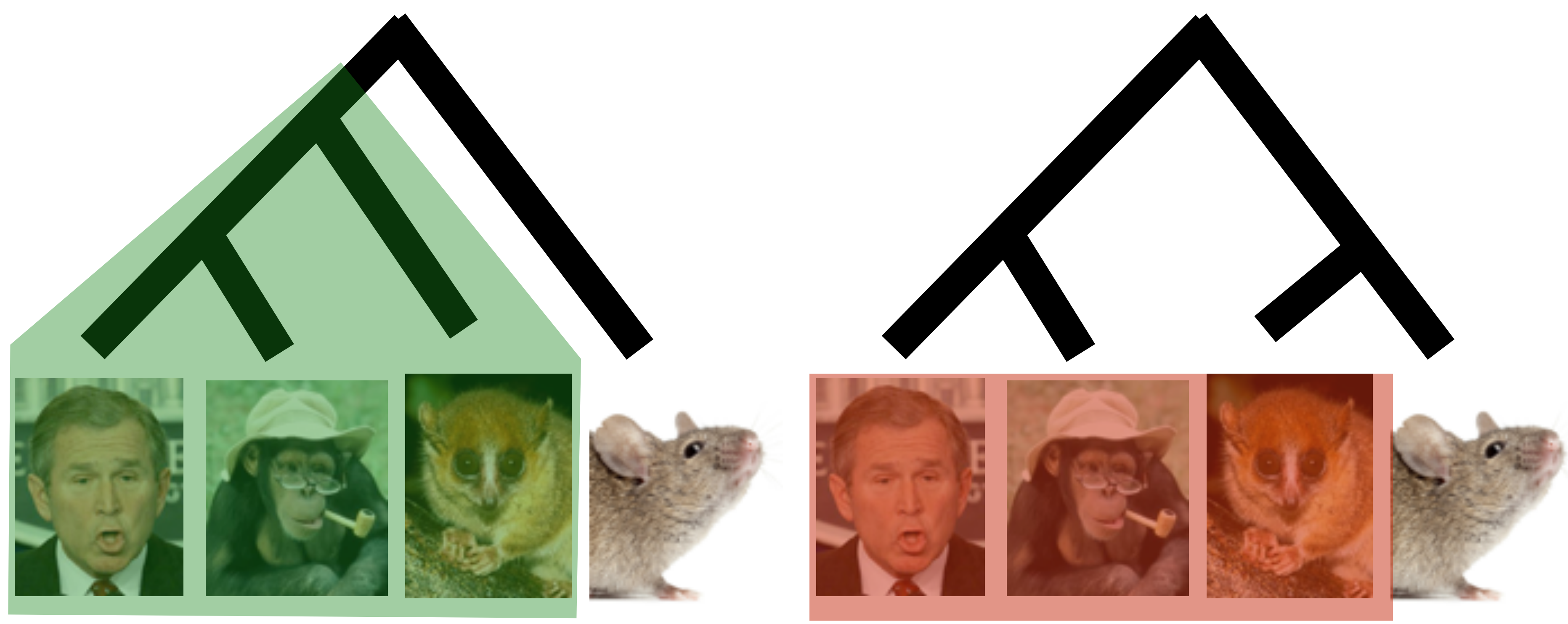
$$x^\star \in \operatorname{argmax}\, p(y|x)$$

$$x^{(i)} \sim \frac{1}{Z} p(x) p(y|x)$$

1. Uncertainty over latent combinatorial structures

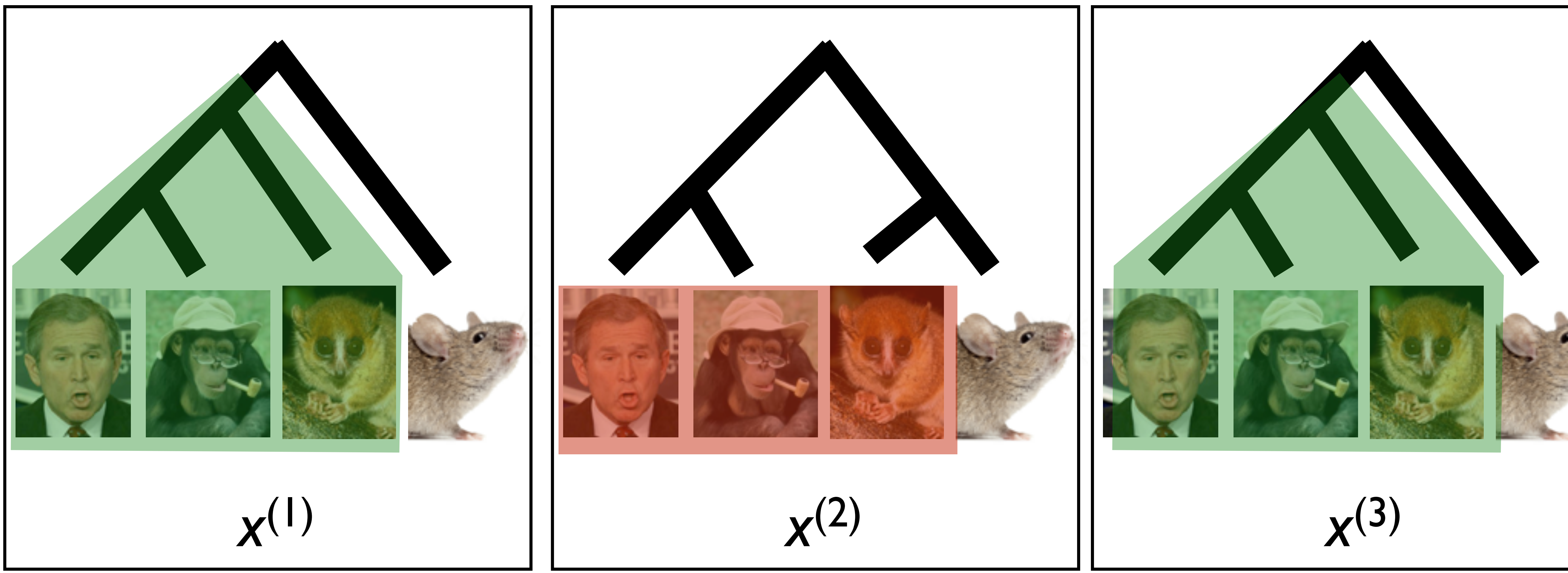
x = unknown **phylogenetic tree**

Question: is {human, monkey, lemur} a **clade**?



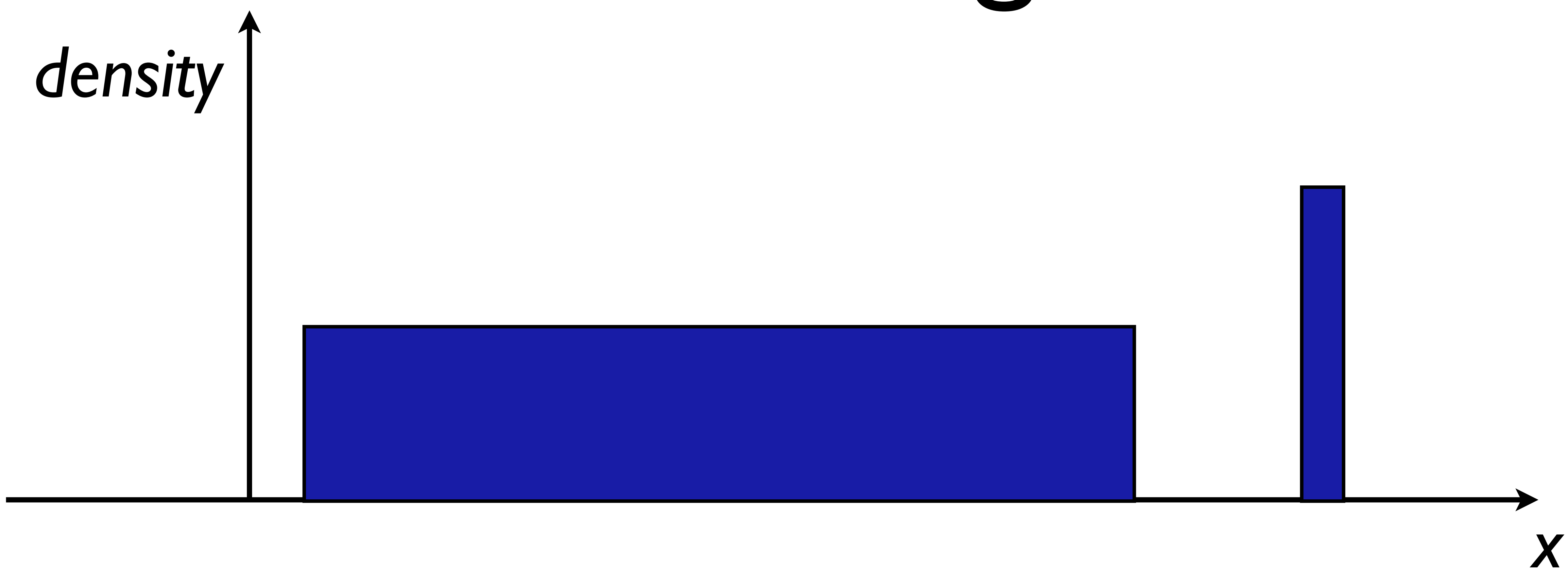
Uncertainty over latent combinatorial structures

x = **phylogenetic tree**



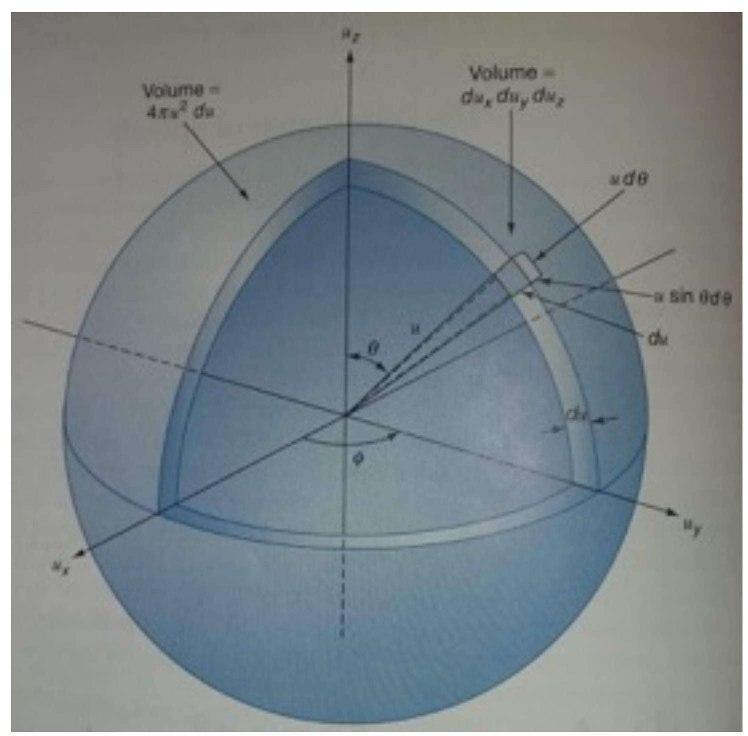
Clade posterior probability: 2/3

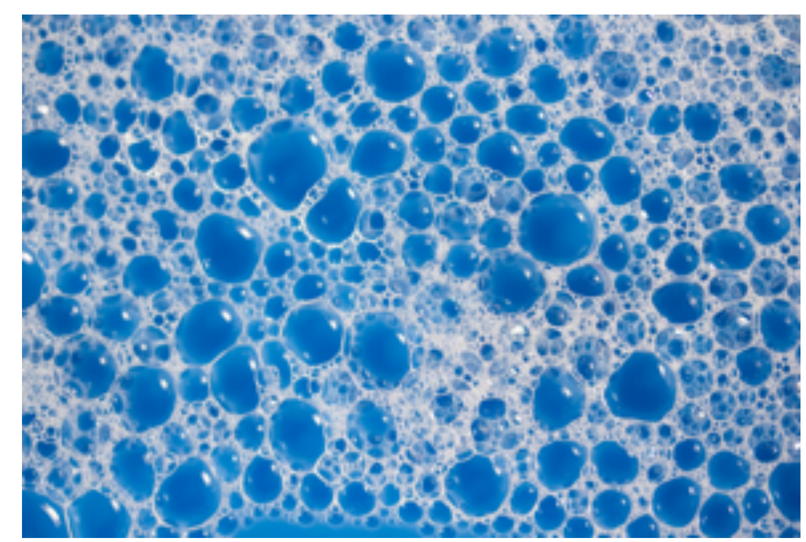
2. Maximization can be misleading



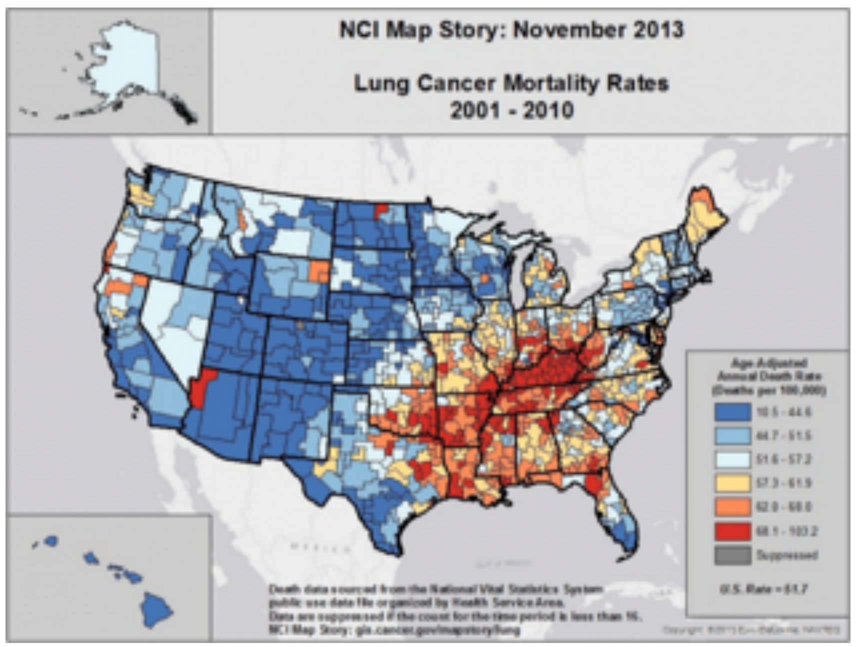
- These cases arise in various contexts:
  - stochastic processes/Bayesian non-parametrics,
  - high-dimensional models,
  - partially identifiable models, ...

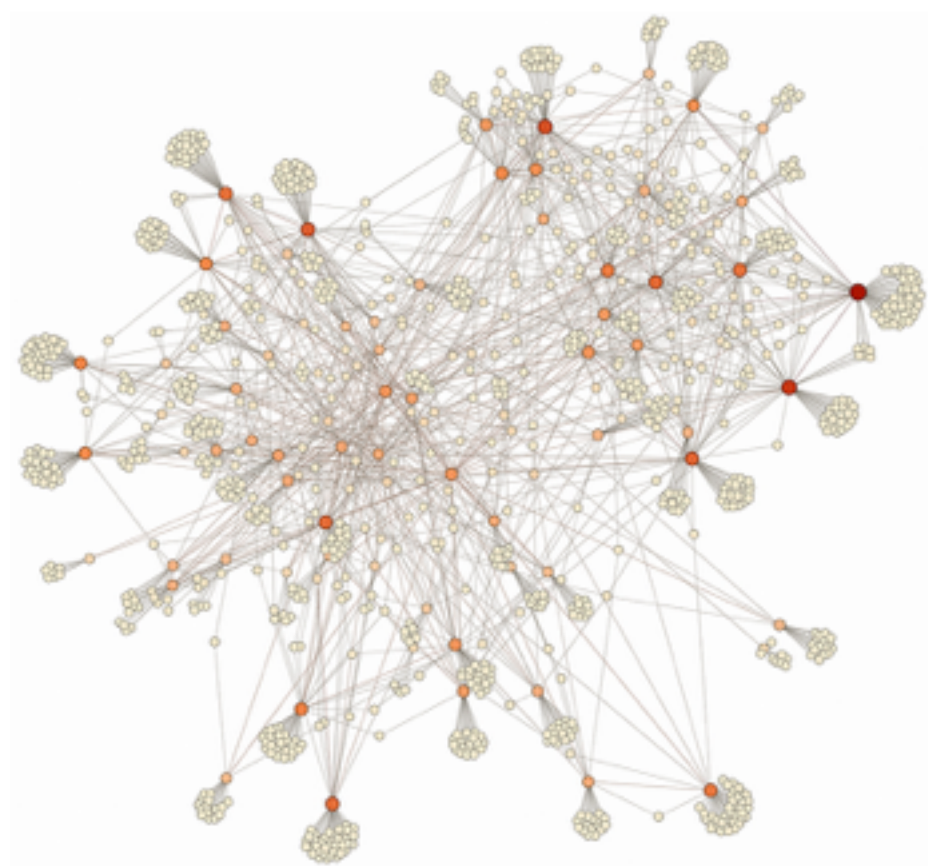
In high dimension, optimization and sampling are profoundly different

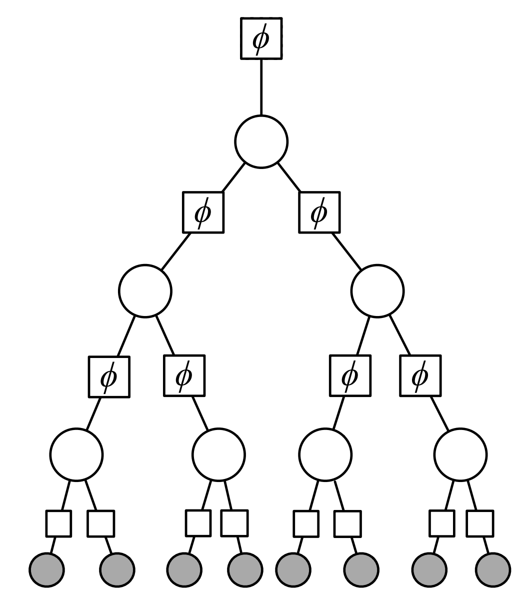




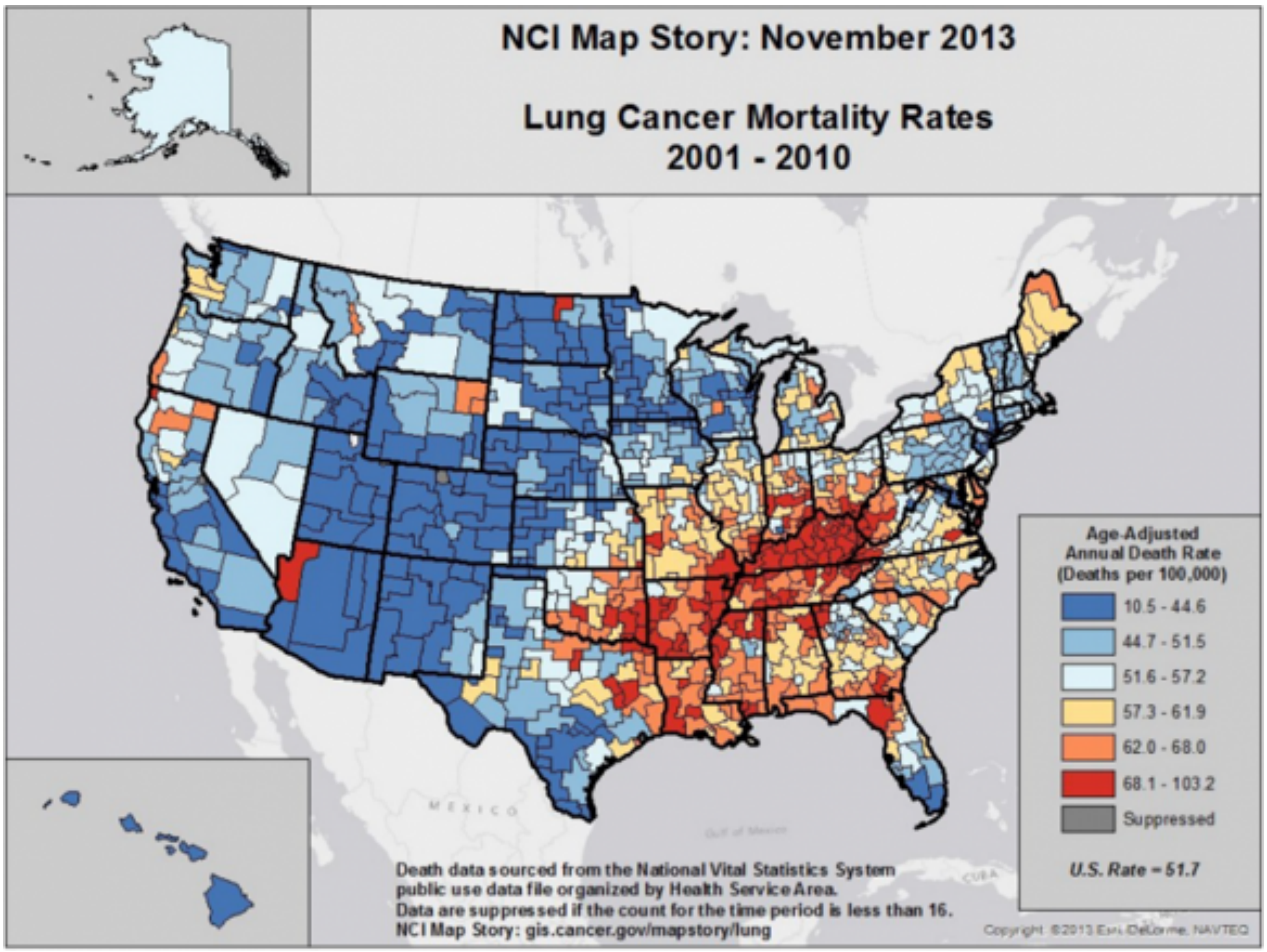
Motivations, continued







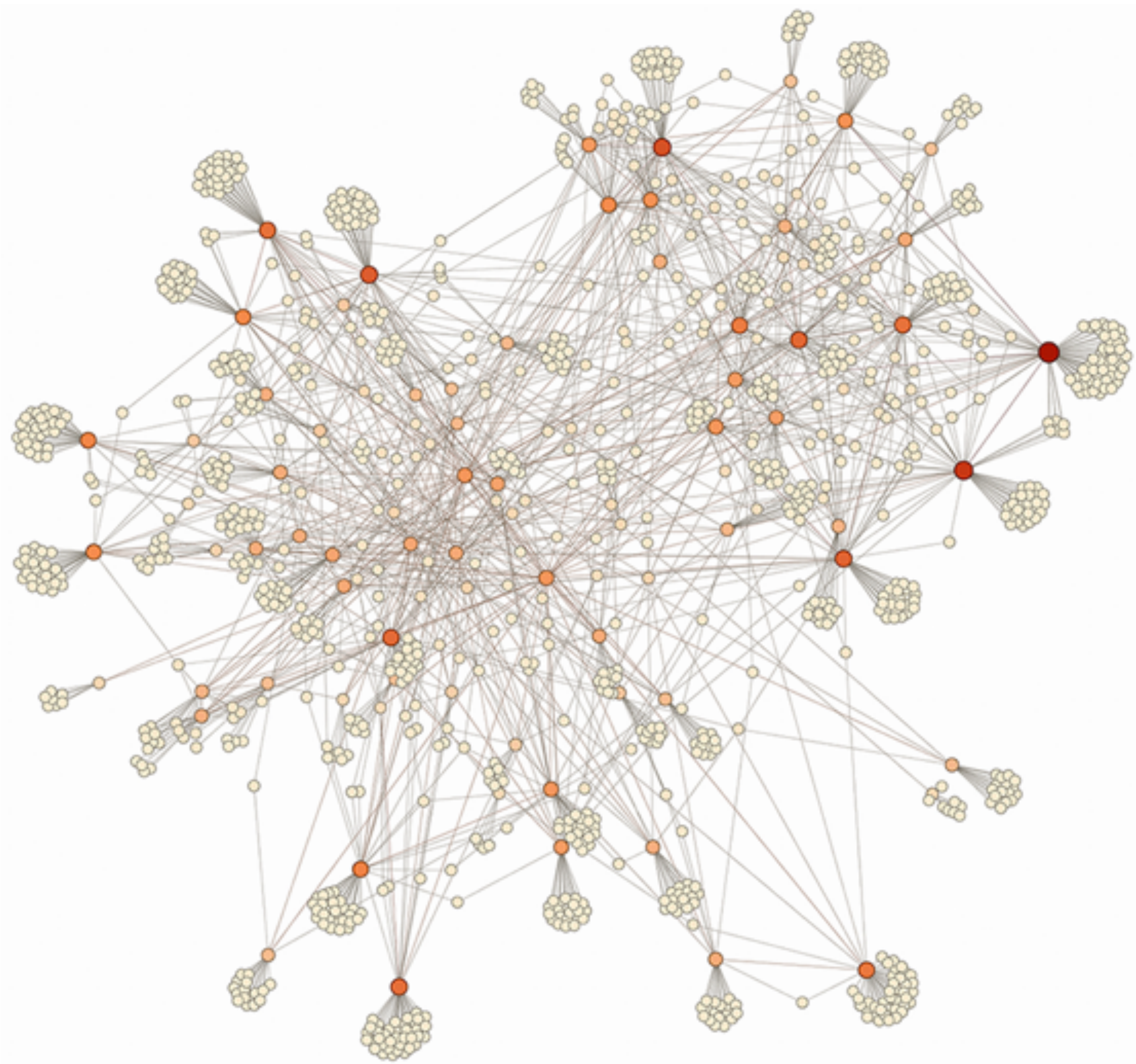
Large scale random effect models



- e.g.: spatial and/or temporal data
- Models: Gaussian Markov random fields, cox processes, etc.

National Cancer Institute

Motivation, continued



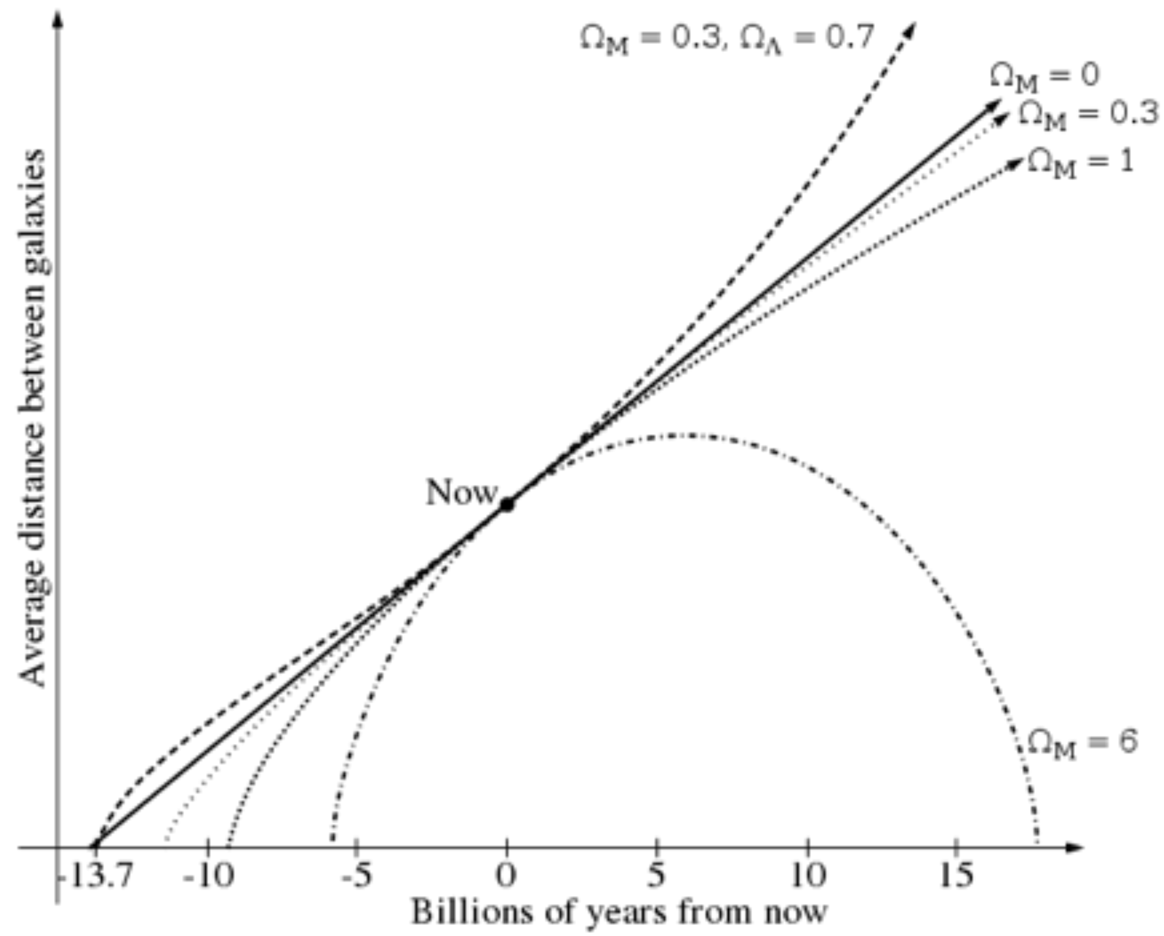
- large networks
  - www, e.g. bitcoin transactions, link analysis, etc
  - biological processes
- sparse exchangeable graph models (e.g. Caron and Fox, 2017)

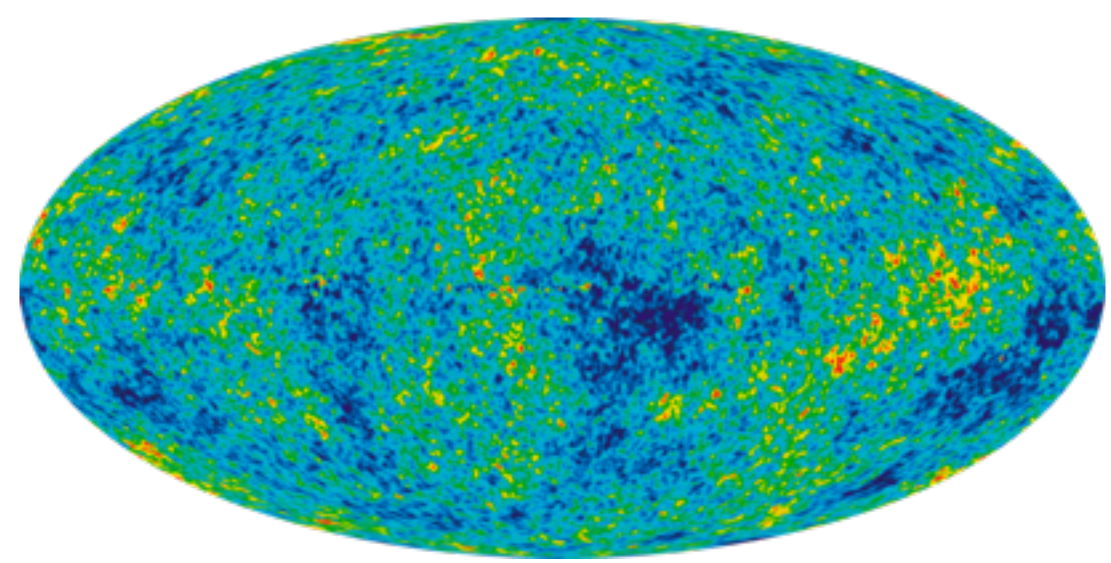
Image credit: Ryan Rossi

Astrophysics: Estimating the age and fate of the Universe

Ex. 3c

- **Goals**: finding the Universe's
  - age
  - density (=> faith)
- **Data**: Cosmic Microwave Background (CMB): remnants of Big Bang
  - Detailed map from the Planck satellite
- Age, Physical constants => known *distribution* on CMP
- Invert using Bayes' rule





More example:

- Computing volumes
- Physics
- Time series

Setup: large scale Bayesian or random effects models

- We are given a density known up to a normalization constant

$$\pi(x) = \frac{\gamma(x)}{Z}$$

- Example: $$\pi(x) = \frac{\text{joint}(x, y)}{\text{evidence}(y)}$$

x : unknown
y : data

- We want a law of large number

$$\frac{1}{N}\sum_{i=1}^{N} \varphi(X^{(i)}) \to \int \varphi(x)\pi(\mathrm{d}x) \text{ a.s.}$$

'test function'

Note: *We almost never care about the samples themselves!*

Overview of the literature

Note 5

Computationally expensive → Computationally efficient

Few assumptions ← More assumptions

$y \sim p(\cdot \| x)$	$\gamma(x) = p(x)p(y \| x)$	$\nabla \log \gamma(x)$
• Approximate Bayesian Computation (ABC) • 'Plug-and-play' Sequential Monte Carlo (SMC)	• Random walk Metropolis • Gibbs sampling	• Hamiltonian Monte Carlo (HMC) • Langevin

Overview of the literature

'Naive' MCMC

- Deterministic start
- Apply kernel *ad nauseam*
- Burn-in, etc

Modern methods

- Sequential change of measure-based
- Replica-based methods

Basics: Simple Monte Carlo

Simple Monte Carlo:

- Simulate independent $X_1, ..., X_n$ from π.
- Return $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i)$.

Simple example

- Consider d-dimensional iid random standard normal vectors
- What is the mean of the distance to the origin?
- Write as:

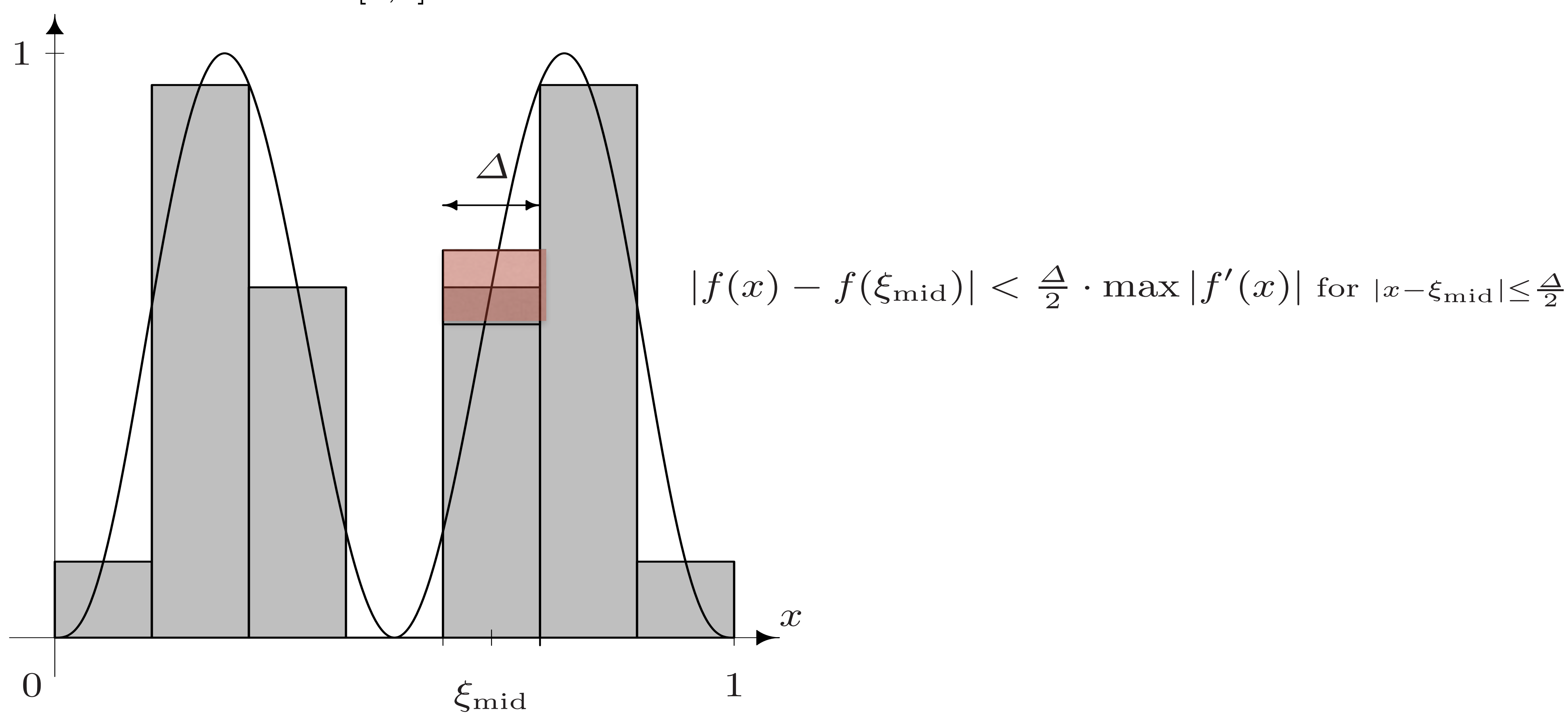
$$I = \int_{\mathbb{X}} f\left(x\right) dx$$

Computational motivation

Integrals: 1D case $I = \int_{\mathbb{X}} f(x)\, dx$

$$\widehat{I}_n = \frac{1}{n} \sum_{i=0}^{n-1} f\left((i+1/2)/n\right).$$

f is differentiable and $\sup_{x \in [0,1]} |f'(x)| < M < \infty$ then the approximation error is $\mathcal{O}\left(n^{-1}\right)$



Computational motivation

Integrals: 2D case $I^{(2)} = \int_{\mathbb{X}} f(x)\, dx$

$$\widehat{I}_n^{(2)} = \frac{1}{n} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} f\left(\left(i + 1/2\right)/n, \left(j + 1/2\right)/n\right) \qquad m = \sqrt{n}$$

approximation error is $\mathcal{O}\left(n^{-1/2}\right)$

Integrals: *d*D case $I^{(d)} = \int_{\mathbb{X}} f(x)\, dx$

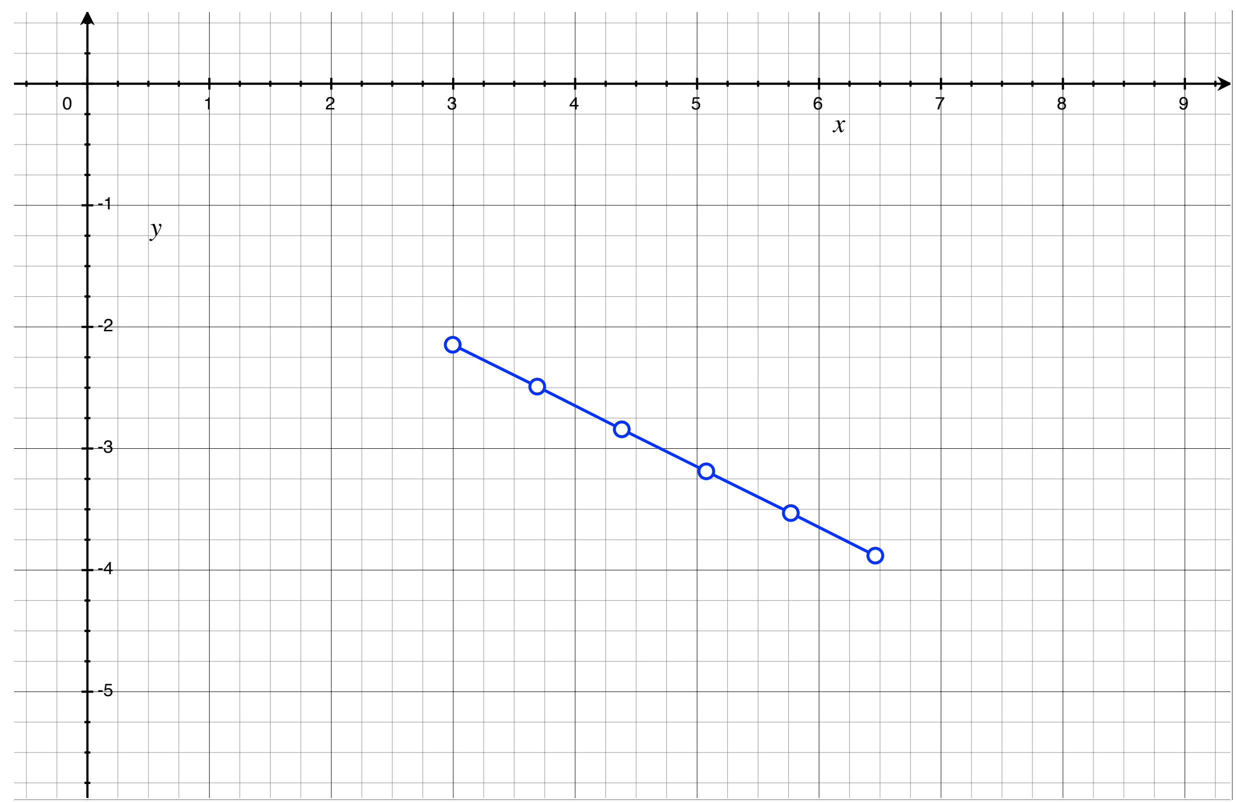
error in $\mathcal{O}\left(n^{-1/d}\right)$

One instance of the 'curse of dimensionality'

Interpretation

- Say found the first two decimal for the integral `0.45???` using a naive numerical integration...
  - in 1d, to get one more decimal correct, need 10x more work
  - in 2d, to get one more decimal correct, need 100x more work
  - in 3d, to get one more decimal correct, need 1000x more work
  - ...

Enters Simple Monte Carlo...



- Exercise:
  - Use this plot to empirically derive the running time of Simple Monte Carlo for a given tolerence *tol*
  - Create the plot for Example 7