Scaling up Continuous-Time Markov Chains Helps Resolve Underspecification

Alkis Gotovos   Rebekka Burkholz

John Quackenbush   Stefanie Jegelka

CSAIL

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

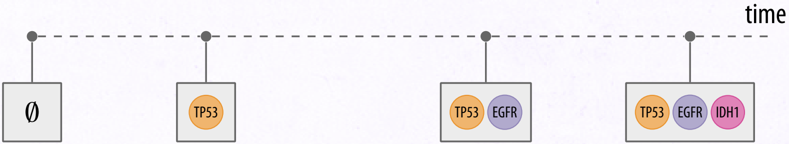
Introduction

Goal: Model the time evolution of discrete sets of items with a continuous-time MC

Introduction

Goal: Model the time evolution of discrete sets of items with a continuous-time MC

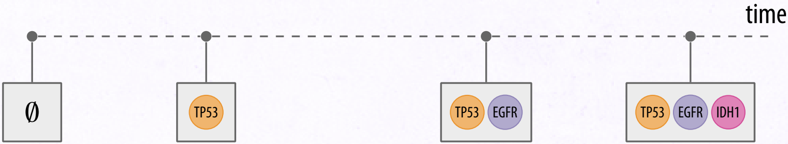
Example: Accumulation of DNA mutations in cancer genomics



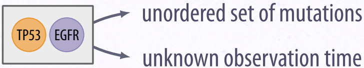
Introduction

Goal: Model the time evolution of discrete sets of items with a continuous-time MC

Example: Accumulation of DNA mutations in cancer genomics



Challenge: Available data are cross-sectional



Introduction

State of the art (Schill et al., '19)

- Constrain analysis to $n \approx 20$ important mutations
- Run $\mathcal{O}(2^n)$ max. likelihood

Introduction

State of the art (Schill et al., ’19)

- Constrain analysis to $n \approx 20$ important mutations
- Run $\mathcal{O}(2^n)$ max. likelihood

Our contributions

- Show that “unimportant” mutations are valuable to resolve underspecification
- Propose approximate max. likelihood scalable to hundreds of mutations
- Evaluate our method on synthetic and real cancer data

Problem setup

Ground set $V = \{1, \ldots, n\}$

Define continuous-time Markov chain $\{X_t\}_{t \geq 0}$ on state space 2^V

Problem setup

Ground set $V = \{1, \ldots, n\}$

Define continuous-time Markov chain $\{X_t\}_{t \geq 0}$ on state space 2^V

$$Q = \begin{bmatrix} q_{\emptyset \to \emptyset} & \cdots & q_{\emptyset \to R} & \cdots & q_{\emptyset \to V} \\ \vdots & & \vdots & & \vdots \\ q_{S \to \emptyset} & \cdots & q_{S \to R} & \cdots & q_{S \to V} \\ \vdots & & \vdots & & \vdots \\ q_{V \to \emptyset} & \cdots & q_{V \to R} & \cdots & q_{V \to V} \end{bmatrix} \in \mathbb{R}^{2^n \times 2^n}$$

Problem setup

Ground set $V = \{1, \ldots, n\}$

Define continuous-time Markov chain $\{X_t\}_{t \geq 0}$ on state space 2^V

$$Q = \begin{bmatrix} q_{\emptyset \to \emptyset} & \cdots & q_{\emptyset \to R} & \cdots & q_{\emptyset \to V} \\ \vdots & & \vdots & & \vdots \\ q_{S \to \emptyset} & \cdots & q_{S \to R} & \cdots & q_{S \to V} \\ \vdots & & \vdots & & \vdots \\ q_{V \to \emptyset} & \cdots & q_{V \to R} & \cdots & q_{V \to V} \end{bmatrix} \in \mathbb{R}^{2^n \times 2^n}$$

Transition rate from S to R

Problem setup

Constraints and parametrization (Schill et al., 2019)

- $X_0 = \emptyset$
- Only add a single mutation at a time (no removals)

Problem setup

Constraints and parametrization (Schill et al., 2019)

- $X_0 = \emptyset$
- Only add a single mutation at a time (no removals)

$$q_{S \to S \cup \{j\}}(\boldsymbol{\theta}) = \exp\left(\theta_{jj} + \sum_{i \in S} \theta_{ij}\right)$$

individual rate of j

effect of i on j

Problem setup

Constraints and parametrization (Schill et al., 2019)

- $X_0 = \emptyset$
- Only add a single mutation at a time (no removals)

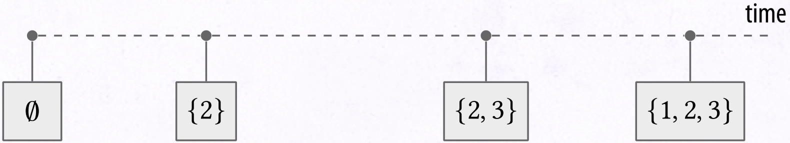
$$q_{S \to S \cup \{j\}}(\boldsymbol{\theta}) = \exp\left(\theta_{jj} + \sum_{i \in S} \theta_{ij}\right)$$

individual rate of j

effect of i on j

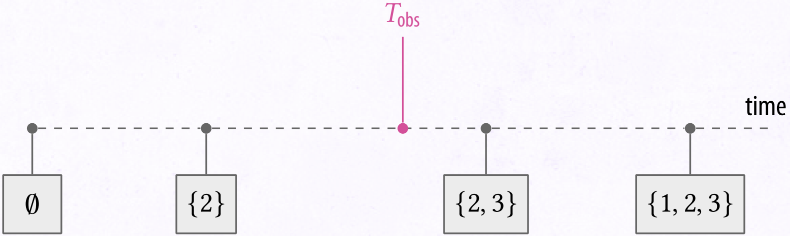
$$\boldsymbol{\Theta} = \begin{bmatrix} \theta_{11} & \dots & \theta_{1n} \\ \vdots & \ddots & \vdots \\ \theta_{n1} & \dots & \theta_{nn} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Problem setup



Problem setup

Draw observation time $T_{\text{obs}} \sim \text{Exp}(1)$



Problem setup

Data $\mathcal{D} = \{S^{(1)}, \ldots, S^{(N)}\}$, $S^{(i)} \subseteq V$

Problem setup

Data $\mathcal{D} = \{S^{(1)}, \ldots, S^{(N)}\}$, $S^{(i)} \subseteq V$

Marginal likelihood

Obs. time

$$p(S^{(i)}; \boldsymbol{\theta}) = \int_0^\infty p(S^{(i)} \mid t; \boldsymbol{\theta}) p(t) dt$$

Markov chain

Problem setup

Data $\mathcal{D} = \{S^{(1)}, \ldots, S^{(N)}\}$, $S^{(i)} \subseteq V$

Marginal likelihood

Obs. time

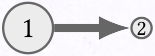
$$p(S^{(i)}; \boldsymbol{\theta}) = \int_0^\infty p(S^{(i)} \mid t; \boldsymbol{\theta}) p(t) dt$$

Markov chain

maximize $\ell(\mathcal{D}; \boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(S^{(i)}; \boldsymbol{\theta})$

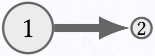
Tackling underspecification

Ground set $V = \{1, 2\}$



Mutation 1 (almost) always occurs before mutation 2

Ground set $V = \{1, 2\}$

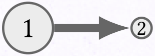


Mutation 1 (almost) always occurs before mutation 2

Data distribution $\boxed{\{1\}}$ $\boxed{\emptyset}$ $\boxed{\{1, 2\}}$ $\boxed{\emptyset}$ $\boxed{\{1\}}$ $\boxed{\{1, 2\}}$ $\boxed{\{1\}}$ $\cdots$

Tackling underspecification

Ground set $V = \{1, 2\}$



Mutation 1 (almost) always occurs before mutation 2

Data distribution $\{1\}$ $\emptyset$ $\{1, 2\}$ $\emptyset$ $\{1\}$ $\{1, 2\}$ $\{1\}$ $\cdots$

Proposition 1 (simplified)

There is a one-dimensional family of models with identical data distribution as above.

Tackling underspecification

Another ground set V_+ containing i.i.d. mutations with no interaction to V

$$\Theta_{\text{full}} = \left(\begin{array}{c|c} \boldsymbol{\Theta} & \mathbf{0} \\ \hline \mathbf{0} & \theta_+ \boldsymbol{I}_m \end{array}\right)$$

Tackling underspecification

Another ground set V_+ containing i.i.d. mutations with no interaction to V

$$\boldsymbol{\Theta}_{\text{full}} = \left(\begin{array}{c|c} \boldsymbol{\Theta} & \mathbf{0} \\ \hline \mathbf{0} & \theta_+ \boldsymbol{I}_m \end{array}\right)$$

Intuition: Mutations in V_+ act like a clock – give us an estimate of T_{obs}

Tackling underspecification

Another ground set V_+ containing i.i.d. mutations with no interaction to V

$$\mathbf{\Theta}_{\text{full}} = \left(\begin{array}{c|c} \mathbf{\Theta} & \mathbf{0} \\ \hline \mathbf{0} & \theta_+ \boldsymbol{I}_m \end{array}\right)$$

Intuition: Mutations in V_+ act like a clock – give us an estimate of T_{obs}

Theorem 1 (simplified)

Let t^* be the true observation time. Then, the mean and variance of the posterior observation time distribution can be bounded as follows:

$$\left|M_{\text{post}} - t^*\right| \approx \sqrt{\frac{\log m}{m}}$$

$$V_{\text{post}} \approx \frac{1}{m}$$

Tackling underspecification

Takeaway: “Unimportant” mutations can be valuable in resolving underspecification

Tackling underspecification

Takeaway: “Unimportant” mutations can be valuable in resolving underspecification

Need SCALABLE likelihood maximization

Scalable approximate likelihood maximization

$$\boxed{\text{Maximize} \quad \ell(\mathcal{D}; \boldsymbol{\theta})}$$

Scalable approximate likelihood maximization

Maximize $\ell(\mathcal{D}; \boldsymbol{\theta})$

First-order optimizer

Scalable approximate likelihood maximization

Maximize $\ell(\mathcal{D}; \boldsymbol{\theta})$

First-order optimizer

Compute $\nabla_{\boldsymbol{\theta}} \log p(S; \boldsymbol{\theta})$ in $\mathcal{O}(2^n)$ (Schill et al., '19)

Scalable approximate likelihood maximization

Maximize $\ell(\mathcal{D}; \boldsymbol{\theta})$

First-order optimizer

Approximate $\nabla_{\boldsymbol{\theta}} \log p(S; \boldsymbol{\theta}) \approx \frac{1}{M} \sum_i \nabla_{\boldsymbol{\theta}} \log p(\sigma^{(i)}; \boldsymbol{\theta})$

Scalable approximate likelihood maximization

Maximize $\ell(\mathcal{D};\boldsymbol{\theta})$

First-order optimizer

Approximate $\nabla_{\boldsymbol{\theta}} \log p(S;\boldsymbol{\theta}) \approx \frac{1}{M}\sum_i \nabla_{\boldsymbol{\theta}} \log p(\sigma^{(i)};\boldsymbol{\theta})$

Compute $\nabla_{\boldsymbol{\theta}} \log p(\sigma^{(i)};\boldsymbol{\theta})$
using modified CTMC

Scalable approximate likelihood maximization

Maximize $\ell(\mathcal{D};\boldsymbol{\theta})$

First-order optimizer

Approximate $\nabla_{\boldsymbol{\theta}} \log p(S;\boldsymbol{\theta}) \approx \frac{1}{M}\sum_i \nabla_{\boldsymbol{\theta}} \log p(\sigma^{(i)};\boldsymbol{\theta})$

Compute $\nabla_{\boldsymbol{\theta}} \log p(\sigma^{(i)};\boldsymbol{\theta})$
using modified CTMC

Sample $\sigma^{(1)},\ldots,\sigma^{(M)}$
using custom M-H sampler

Real data experiments

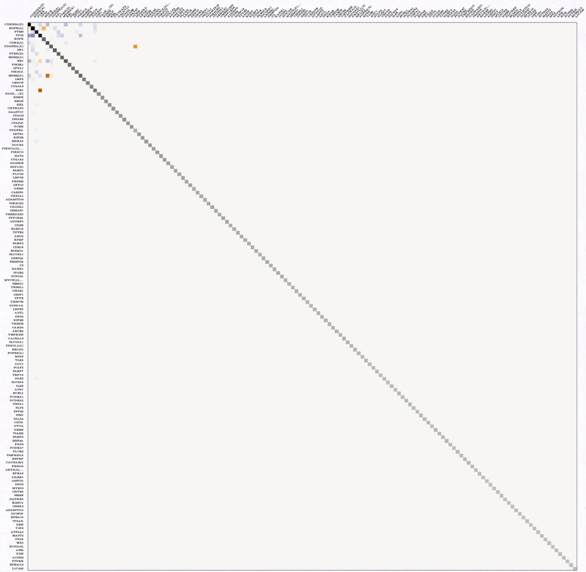
- TCGA glioblastoma data
- $|V| = 410$ mutations, amplifications, and deletions

Real data experiments

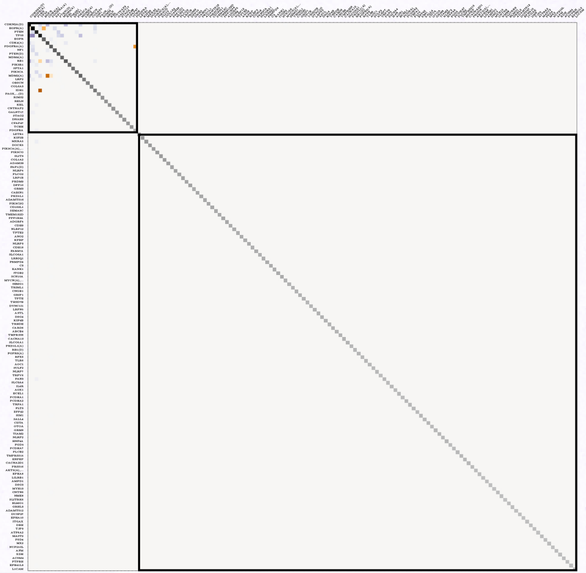
- TCGA glioblastoma data
- $|V| = 410$ mutations, amplifications, and deletions

Method	$n = 20$	$n = 100$
(Schill et al., 2019)	121 m	–
Ours	8 s	33 m 43 s

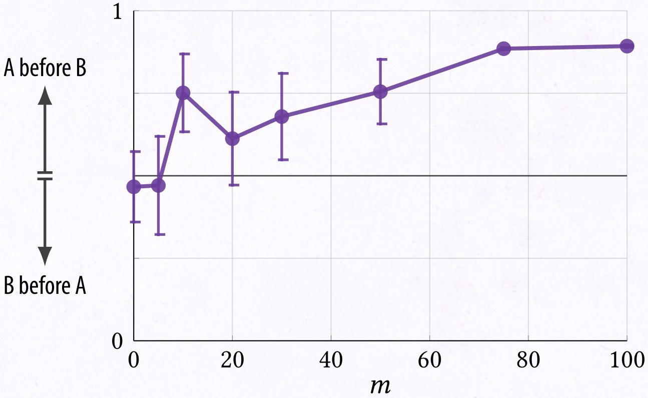
Real data experiments



Real data experiments

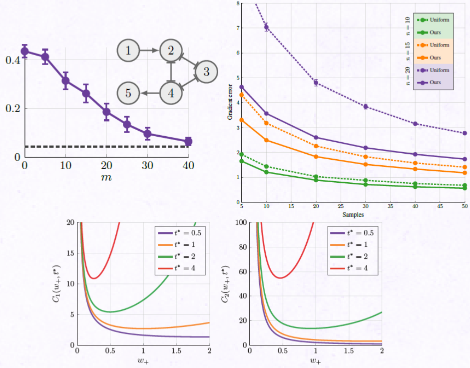


Real data experiments



(A: PDGFRA(A), B: PDGFRA)

Further resources



Paper: `https://arxiv.org/abs/2107.02911/`

Code: `https://github.com/3lectrologos/time/`