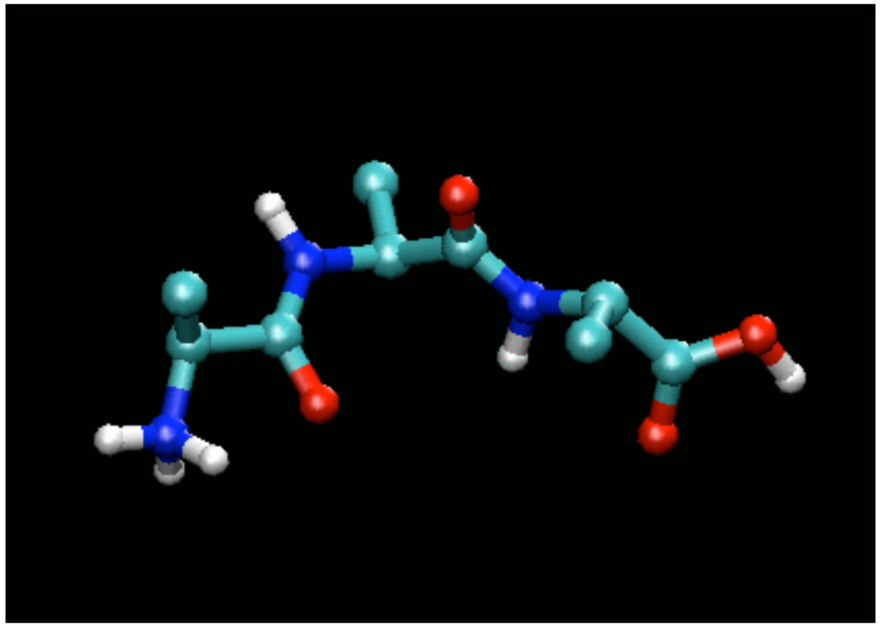
Gibbs Sampling in Factorized Continuous-Time Markov Processes

Tal El-Hay
joint work with
Nir Friedman and Raz Kupferman

The Hebrew University, Jerusalem

Modeling Dynamic Systems

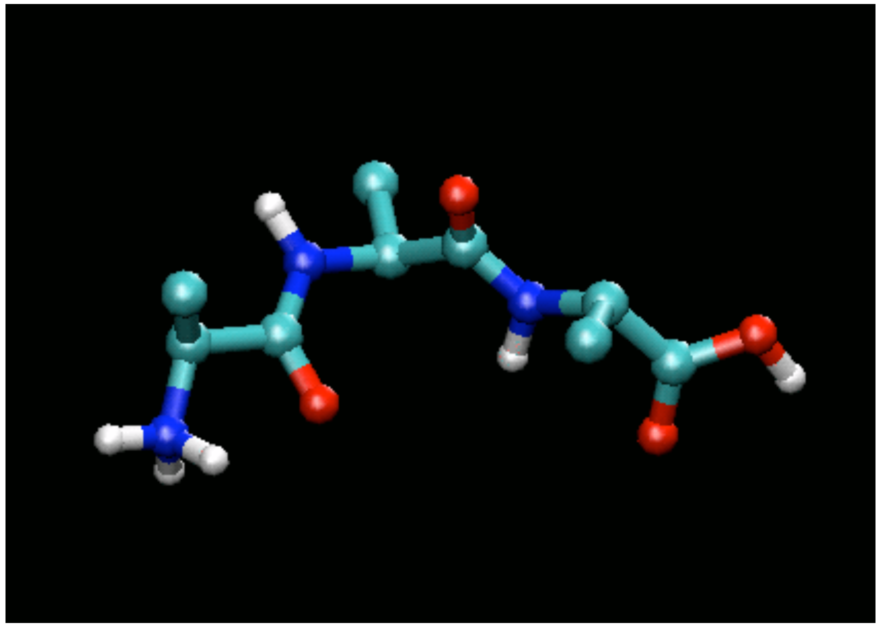
- Continuous time



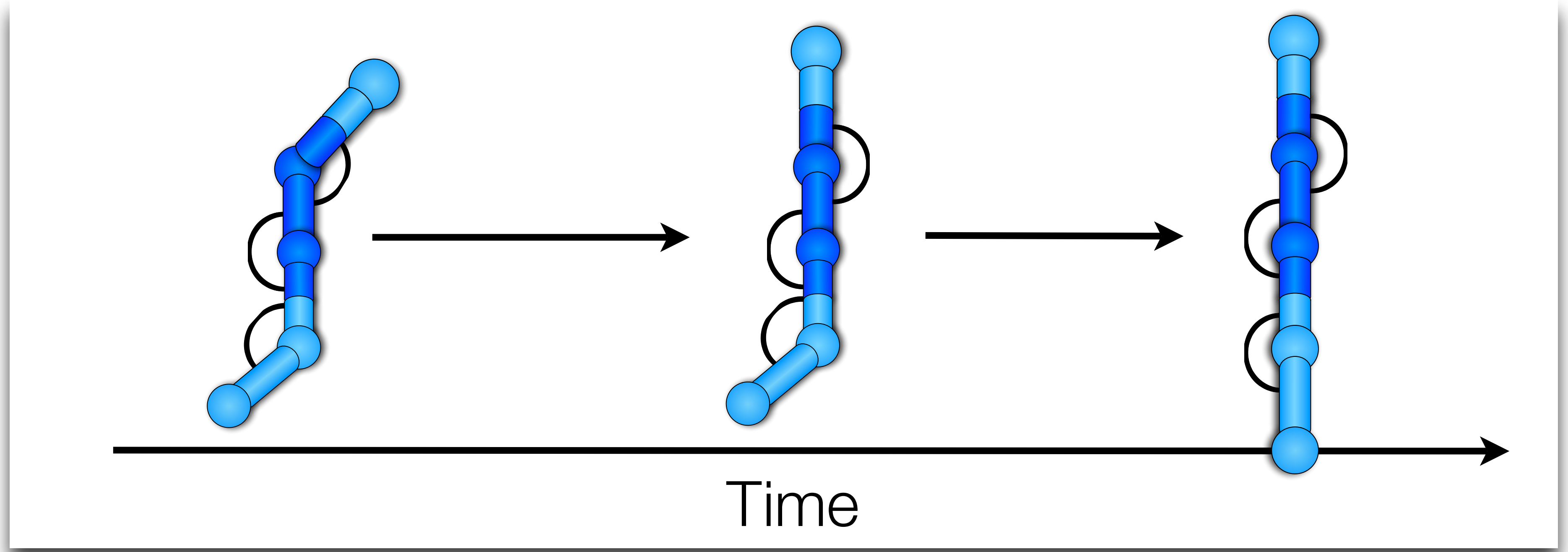
http://www.ntu.edu.sg/home/ygmu/

Modeling Dynamic Systems

- Continuous time
- Multi-component

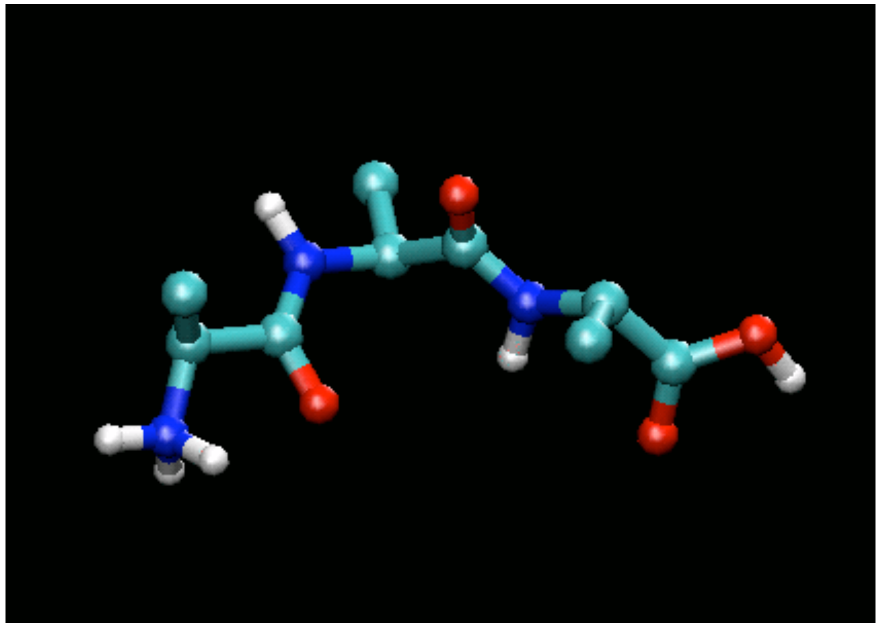


http://www.ntu.edu.sg/home/ygmu/

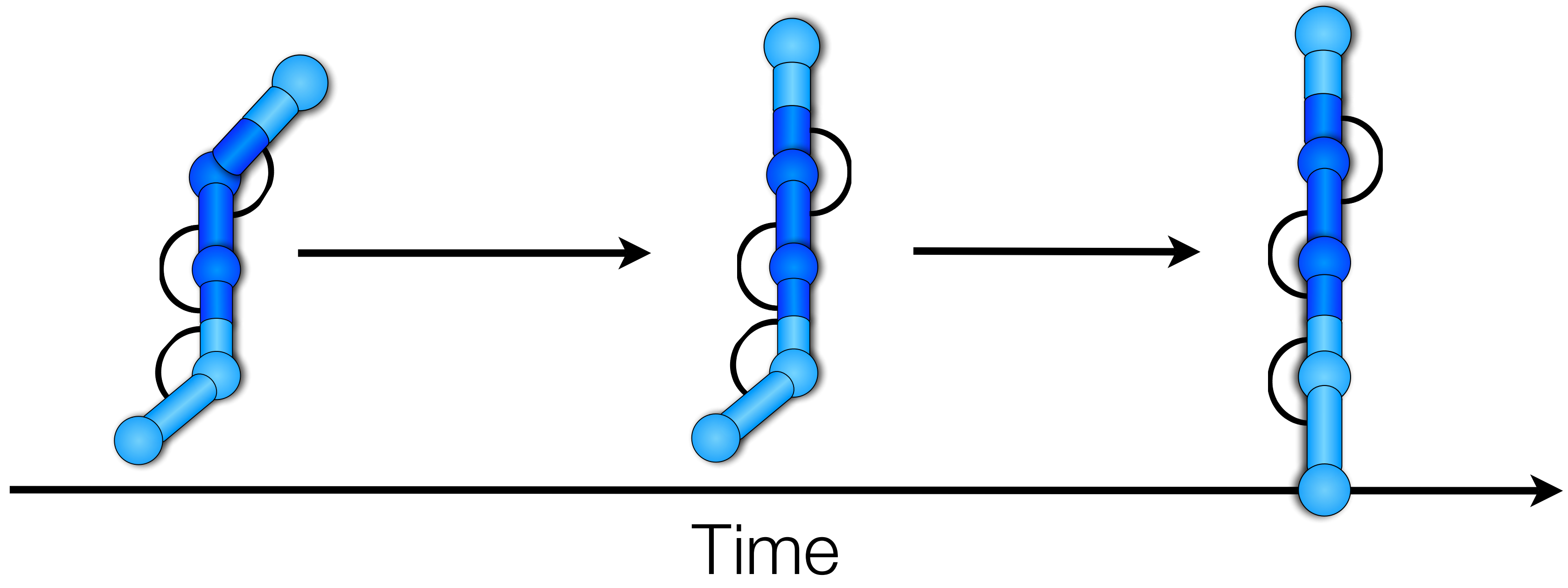


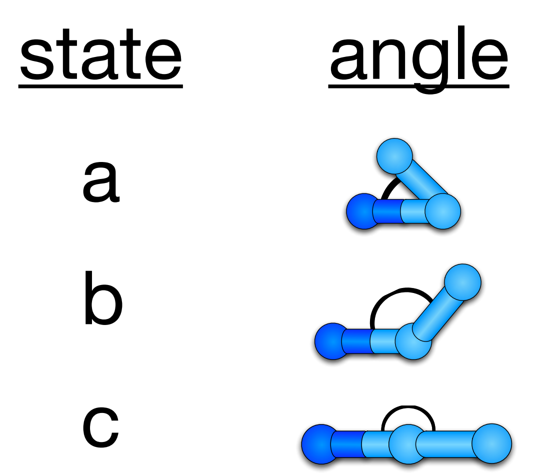
Modeling Dynamic Systems

- Continuous time
- Multi-component
- Discrete states



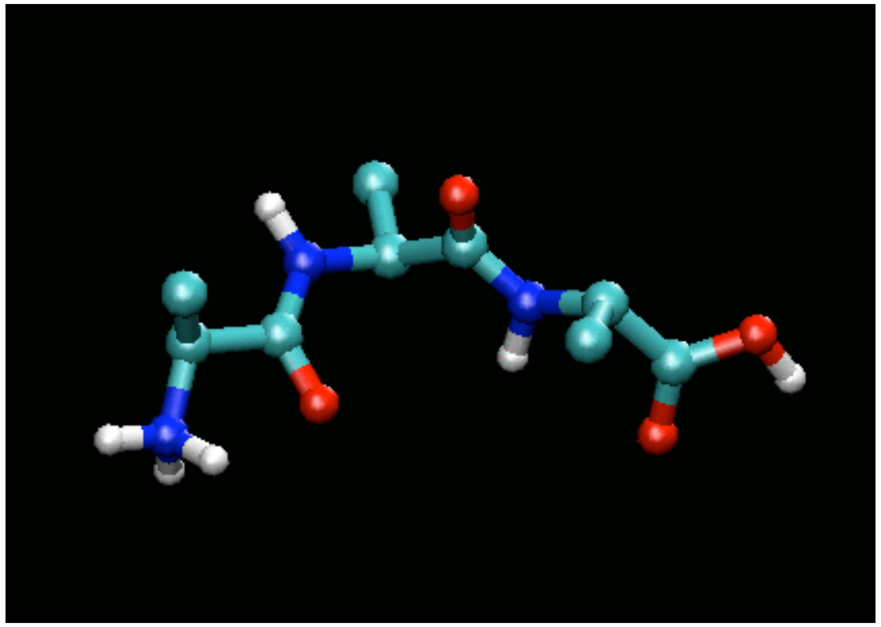
http://www.ntu.edu.sg/home/ygmu/



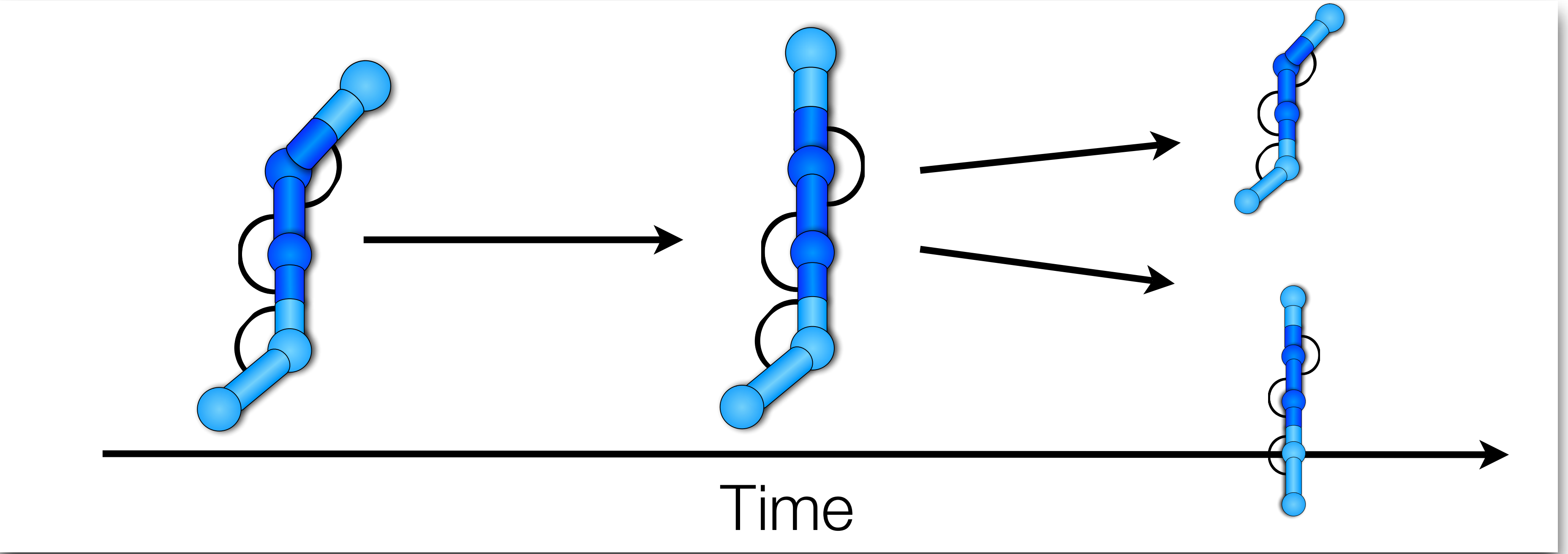


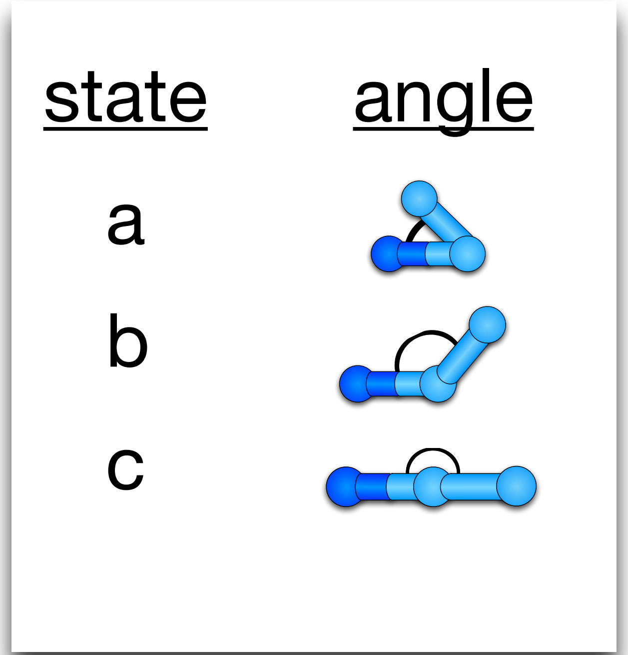
Modeling Dynamic Systems

- Continuous time
- Multi-component
- Discrete states
- Stochastic Dynamics



http://www.ntu.edu.sg/home/ygmu/



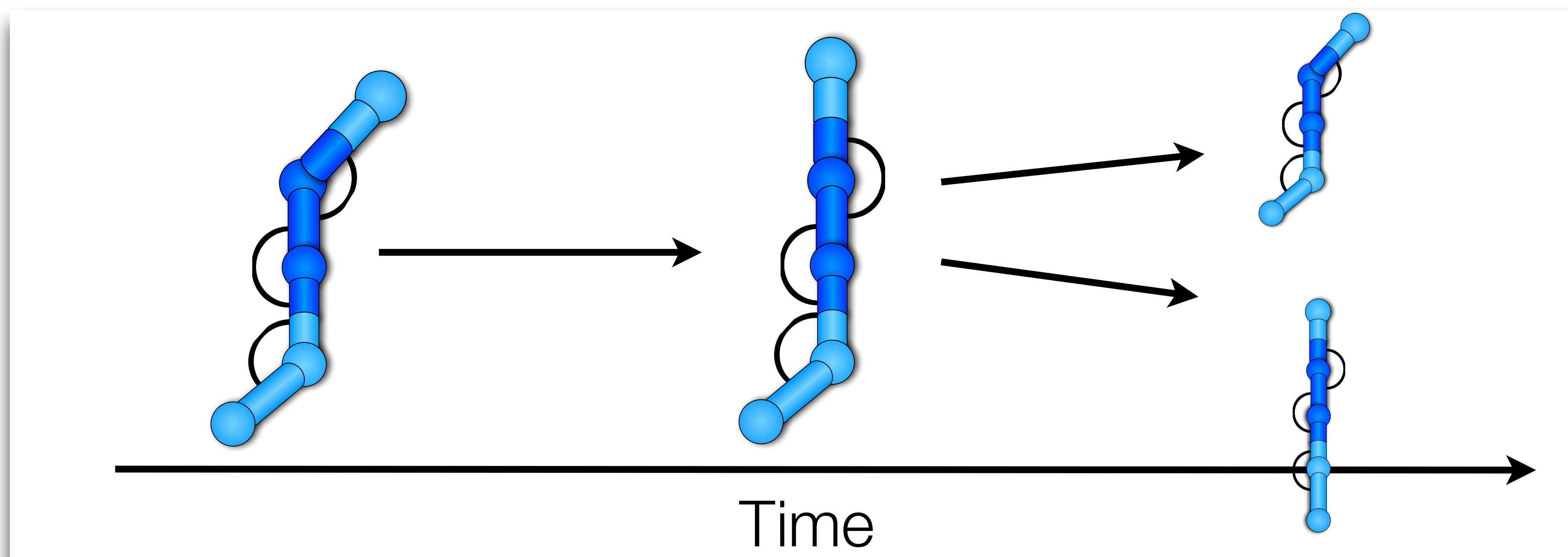


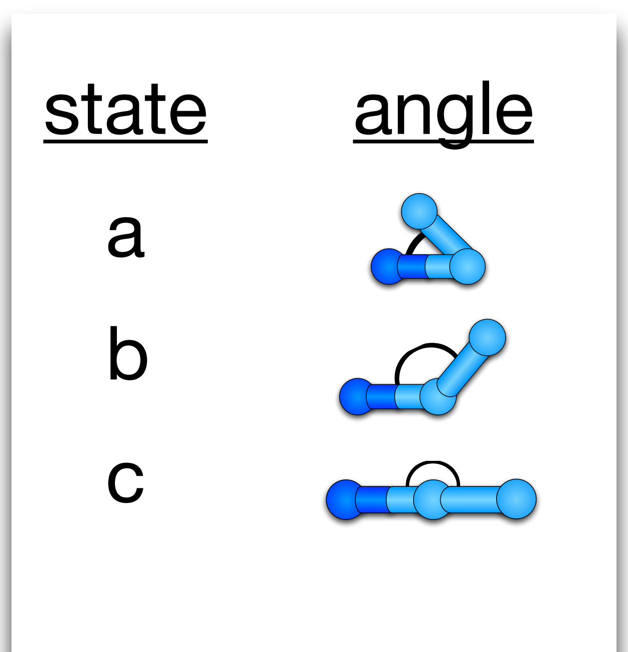
Modeling Dynamic Systems

- Continuous time
- Multi-component
- Discrete states
- Stochastic Dynamics

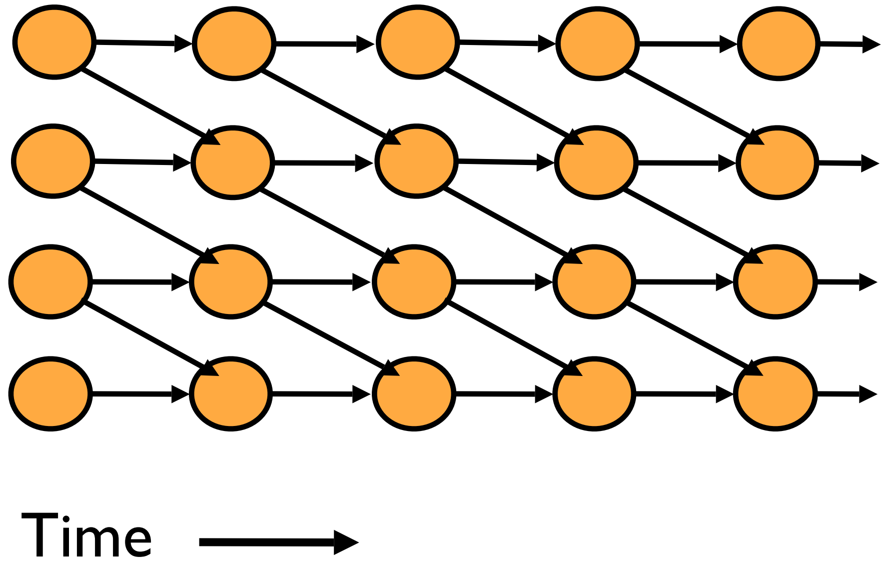
<u>Some Applications</u>

- Molecular Biology
- Evolution
- Robot monitoring
- Computer networks

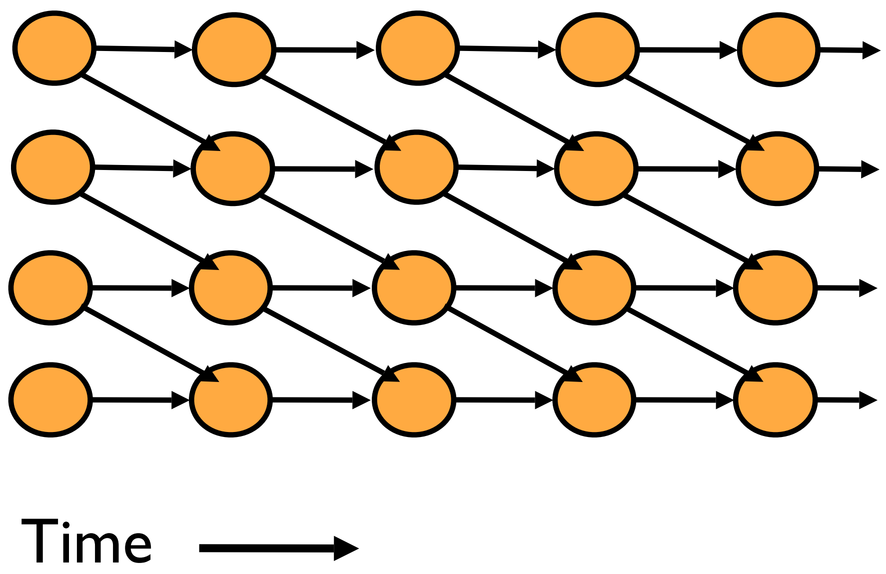




Continuous Time versus Discrete Time Modeling



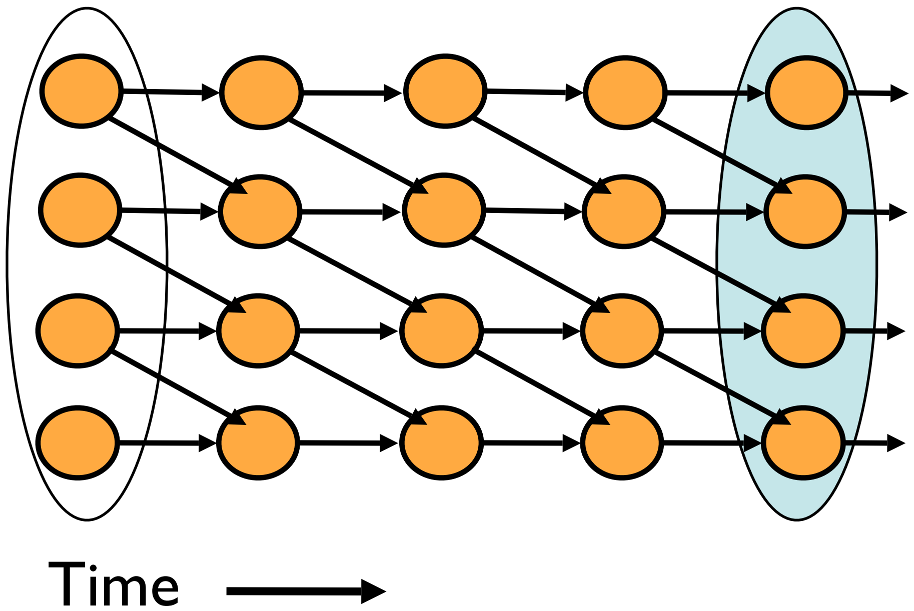
Continuous Time versus Discrete Time Modeling



Discretization:

- Too fine - high computational overhead

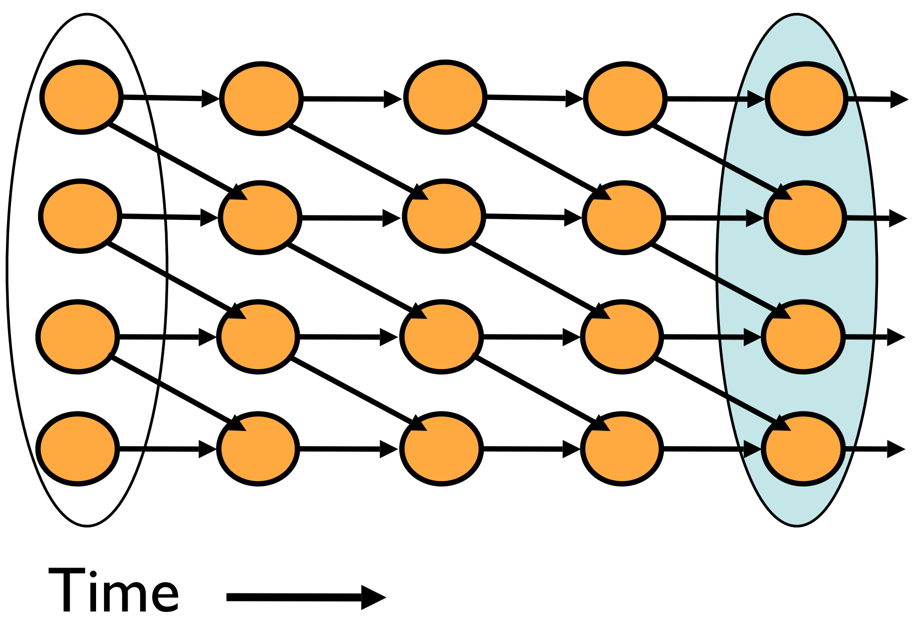
Continuous Time versus Discrete Time Modeling



Discretization:

- Too fine - high computational overhead
- Too coarse - entanglement

Continuous Time versus Discrete Time Modeling

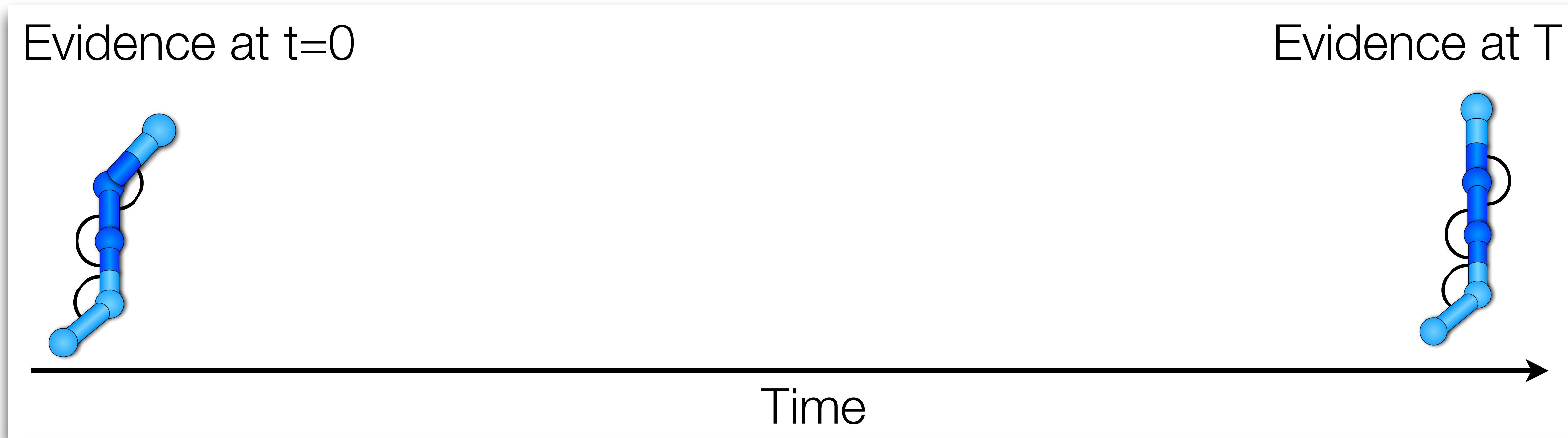


Discretization:

- Too fine - high computational overhead
- Too coarse - entanglement

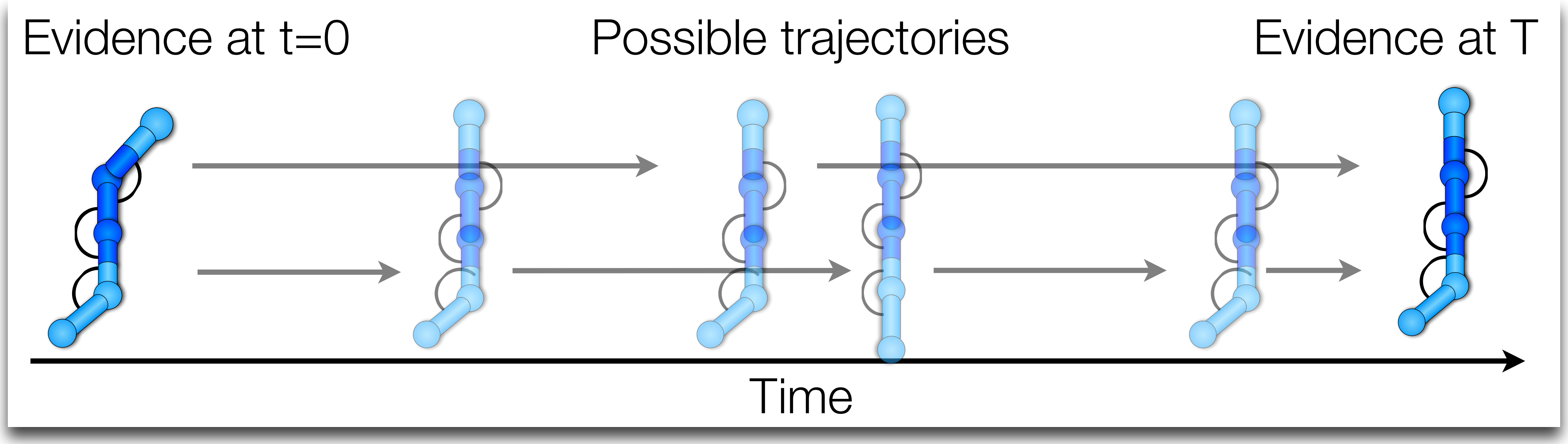
Avoid granularity issues → model continuous time

Inference in Continuous Time Models



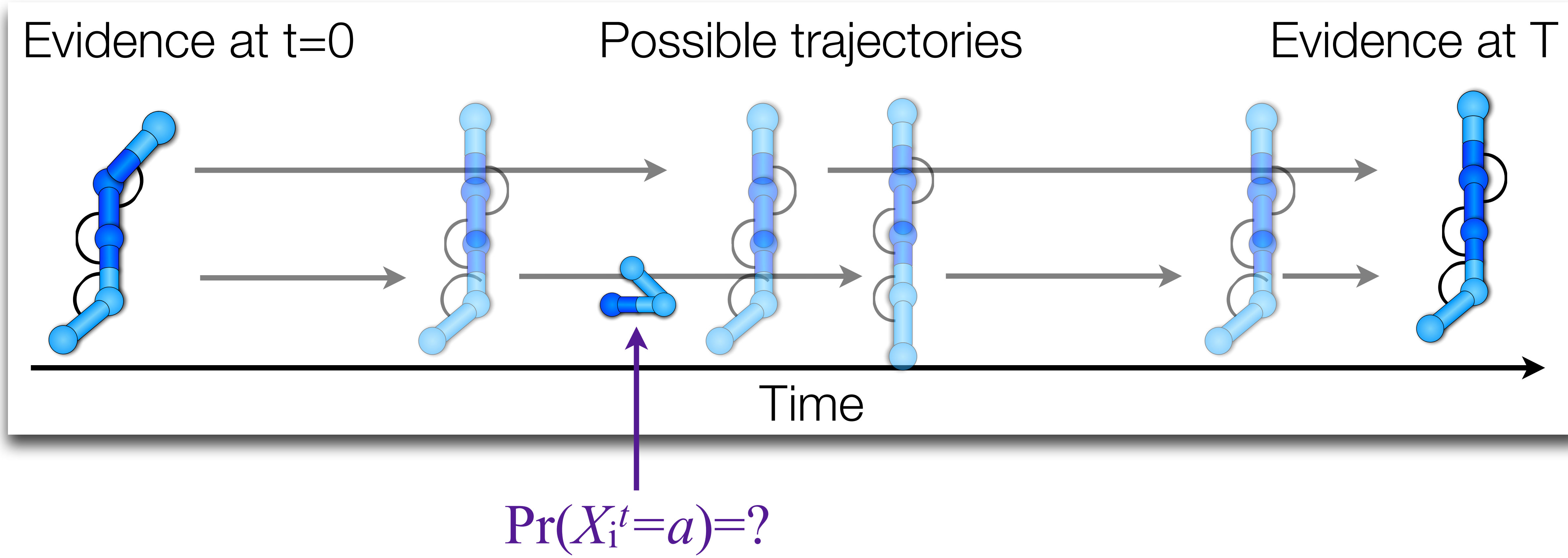
Inference in Continuous Time Models

Uncertainty over uncountably many possible trajectories



Inference in Continuous Time Models

Uncertainty over uncountably many possible trajectories

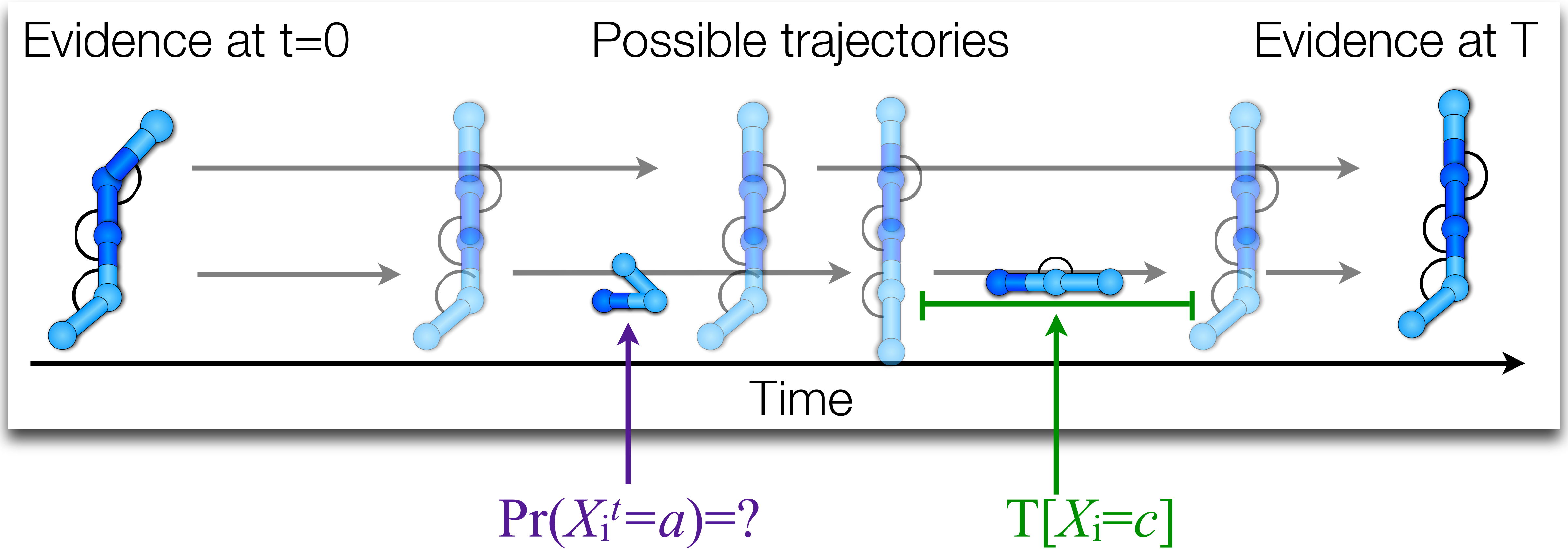


Queries:

1. Marginals at given times

Inference in Continuous Time Models

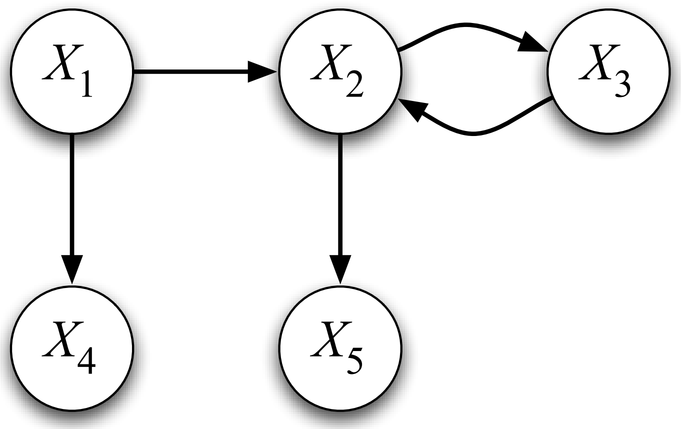
Uncertainty over uncountably many possible trajectories



Queries:

1. Marginals at given times
2. Expectations of statistics - e.g. state durations

Challenges



Representation

CTBN (Nodelman et al. 2002)

CTMN (El-Hay et al. 2006)

Inference

Expectation propagation (Nodelman et al. 2005; Saria et al. 2007)

Sampling/Particle filtering (Ng et al. 2005; Fan and Shelton 2008)

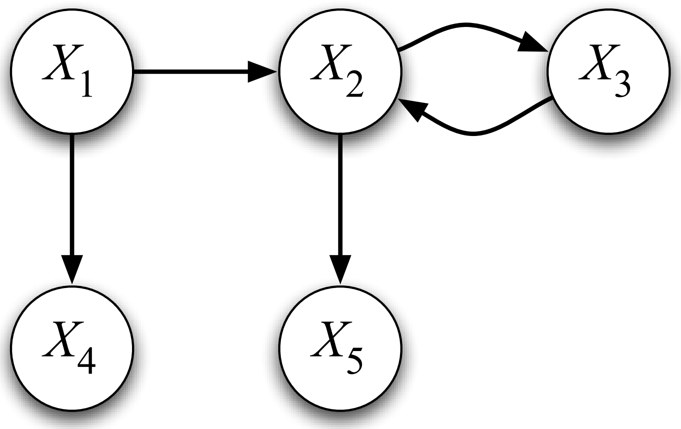
Parameter estimation

(Nodelman et al. 2003, 2005; El-Hay et al. 2006)

Structure learning

(Nodelman et al. 2003)

Challenges



Representation

CTBN (Nodelman et al. 2002)

CTMN (El-Hay et al. 2006)

Inference

Expectation propagation (Nodelman et al. 2005; Saria et al. 2007)

Sampling/Particle filtering (Ng et al. 2005; Fan and Shelton 2008)

Parameter estimation

(Nodelman et al. 2003, 2005; El-Hay et al. 2006)

Structure learning

(Nodelman et al. 2003)

Goal

Unbiased sampling scheme in continuous time

- Handle all types evidence
- Efficient

- Enrich toolbox of inference algorithms
- Allow evaluation of faster but biased schemes

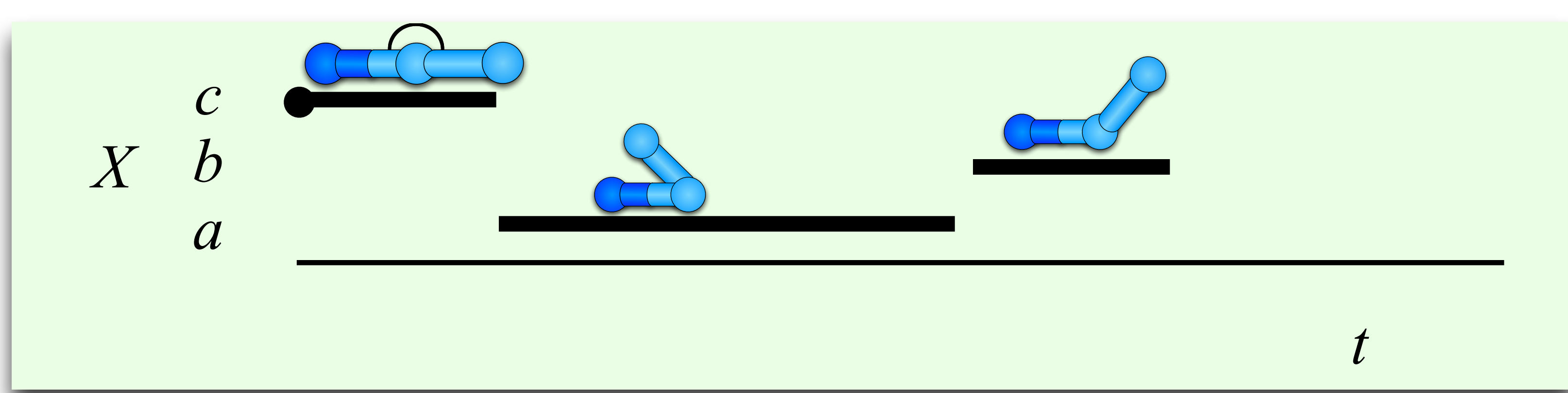
Continuous Time Markov Processes - Definitions

A collection of discrete random variables $X^{(t)}$ where $t \in [0,\infty)$

Satisfies Markov property

$$P(X(t_{k+1})|X(t_k), X(t_{k-1}), \ldots, X(t_0)) = P(X(t_{k+1})|X(t_k))$$

Time Homogeneity



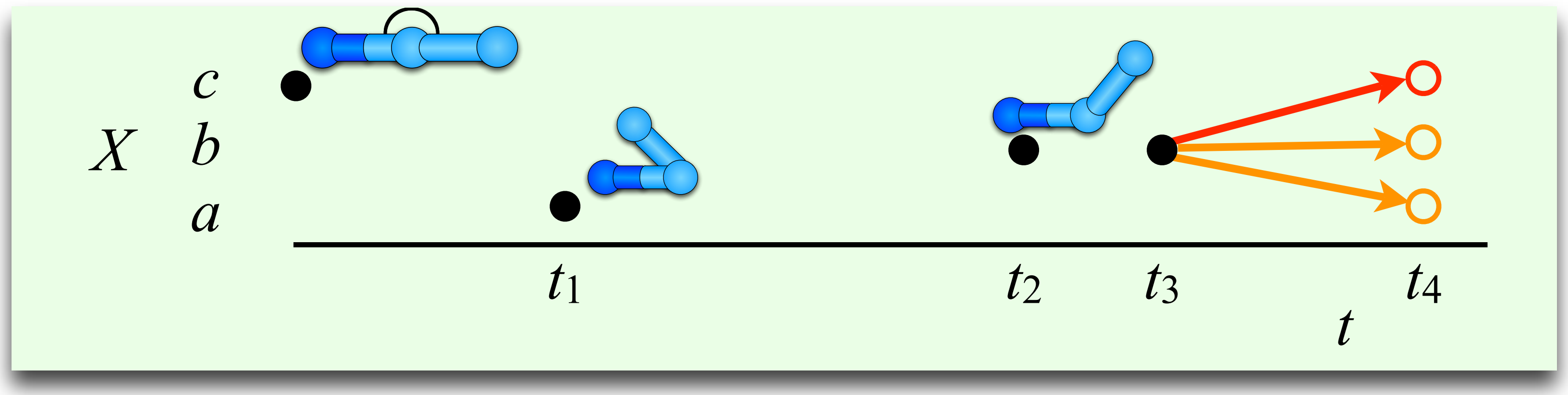
Continuous Time Markov Processes - Definitions

A collection of discrete random variables $X^{(t)}$ where $t \in [0,\infty)$

Satisfies Markov property

$$P(X(t_{k+1})|X(t_k), X(t_{k-1}), \ldots, X(t_0)) = P(X(t_{k+1})|X(t_k))$$

Time Homogeneity



Continuous Time Markov Processes - Definitions

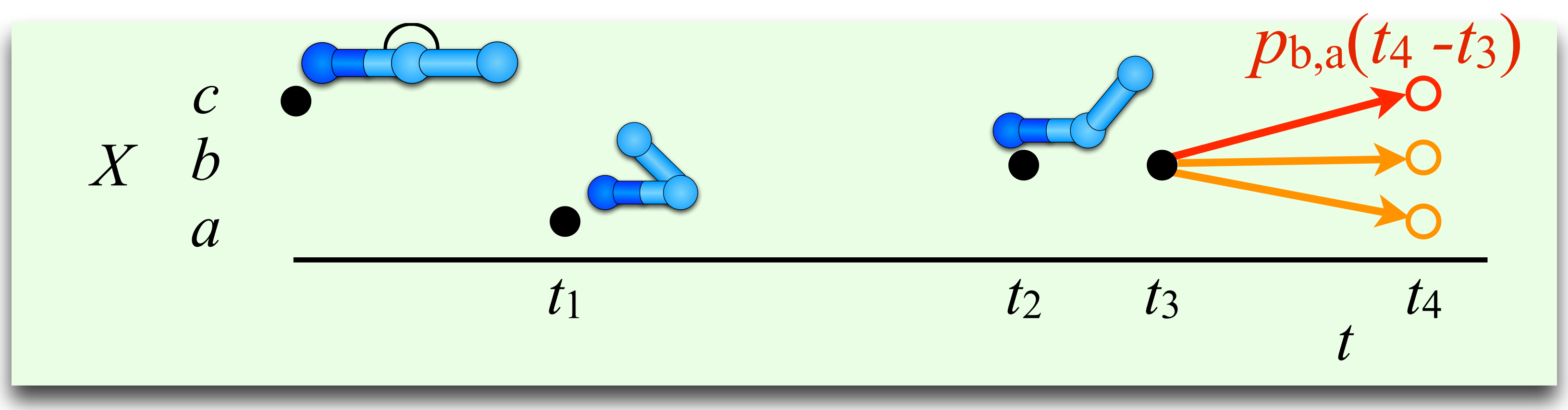
A collection of discrete random variables $X^{(t)}$ where $t \in [0,\infty)$

Satisfies Markov property

$$P(X(t_{k+1})|X(t_k), X(t_{k-1}), \ldots, X(t_0)) = P(X(t_{k+1})|X(t_k))$$

Time Homogeneity

$$p_{a,b}(t) = \Pr(X^{(s+t)} = b | X^{(s)} = a)$$

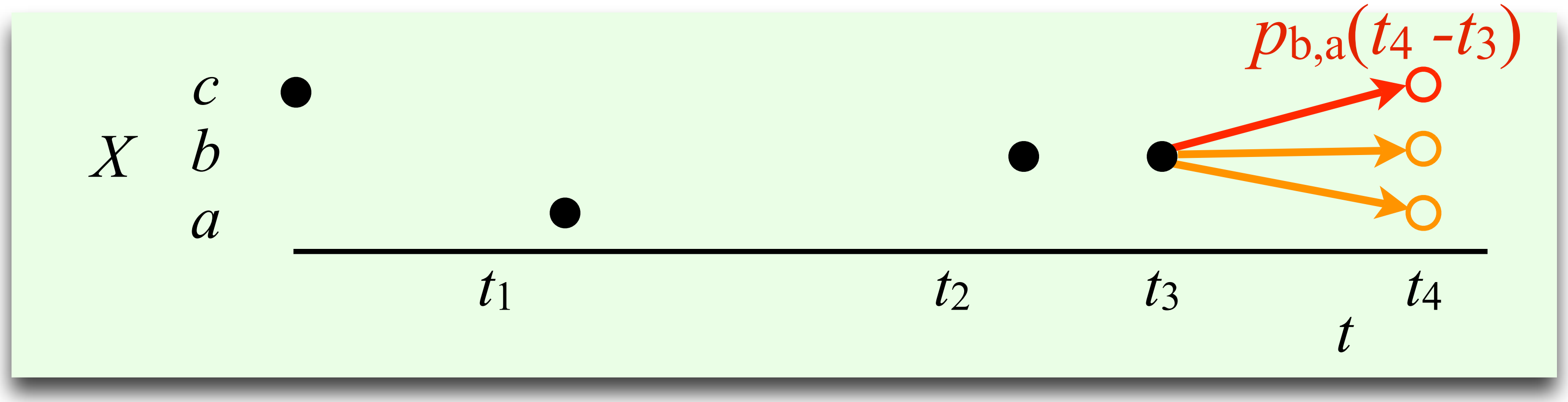


Continuous Time Markov Processes - Parameterization

Parameterized by a *rate matrix* $\mathbb{Q}$:

$$p_{a,b}(\Delta t) \approx q_{a,b}\Delta t, \qquad \Delta t \to 0, a \neq b$$

Transition probabilities satisfy



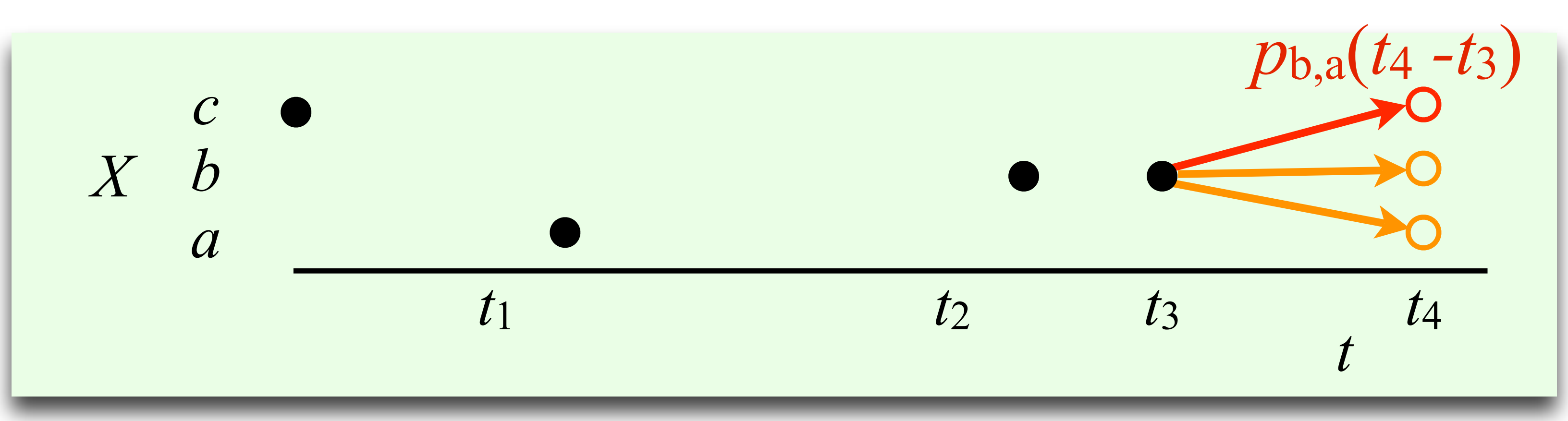
Continuous Time Markov Processes - Parameterization

Parameterized by a *rate matrix* $\mathbb{Q}$:

$$p_{a,b}(\Delta t) \approx q_{a,b}\Delta t, \qquad \Delta t \rightarrow 0, a \neq b$$

transition rates

Transition probabilities satisfy



Continuous Time Markov Processes - Parameterization

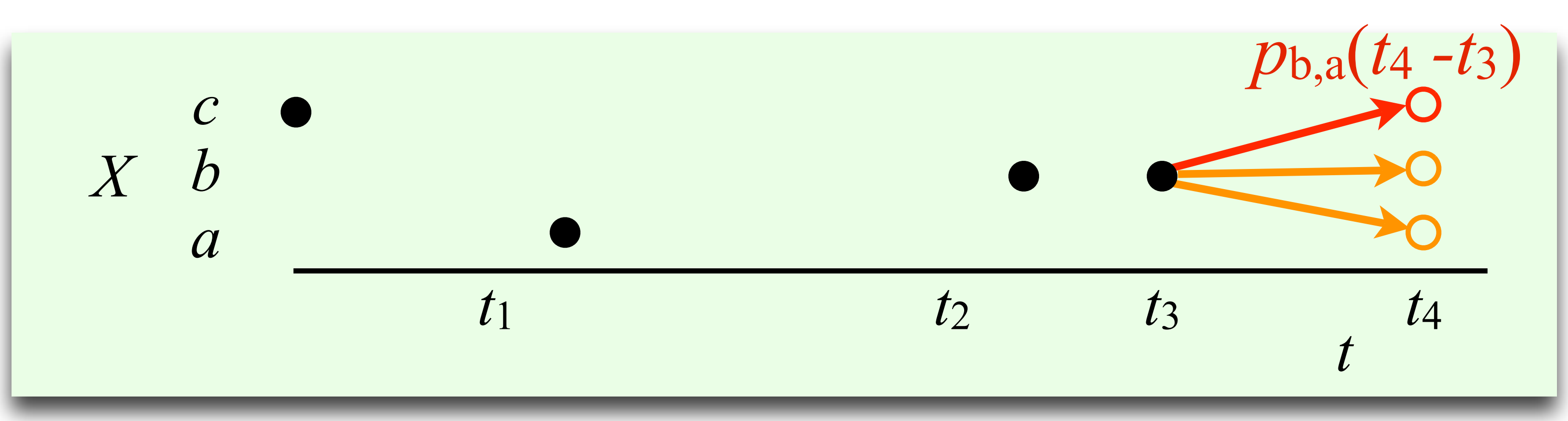
Parameterized by a *rate matrix* $\mathbb{Q}$:

$$p_{a,b}(\Delta t) \approx q_{a,b} \Delta t, \qquad \Delta t \to 0, a \neq b$$

transition rates

Transition probabilities satisfy

$$p_{a,b}(t) = \left[e^{t\mathbb{Q}}\right]_{a,b}$$



Continuous Time Bayesian Networks

Multi component Markov process $X=(X_1, \ldots, X_N)$

Naive representation ➜ size of rate matrix is exponential in number of components

Continuous Time Bayesian Networks

Multi component Markov process $X=(X_1, \ldots, X_N)$

Naive representation ➜ size of rate matrix is exponential in number of components

Compact representation

[Nodelman et al. 2002]

Continuous Time Bayesian Networks

Multi component Markov process $X=(X_1, \ldots, X_N)$

Naive representation ➜ size of rate matrix is exponential in number of components

<u>Compact representation</u>

- Only one component changes its state in a single transition

[Nodelman et al. 2002]

Continuous Time Bayesian Networks

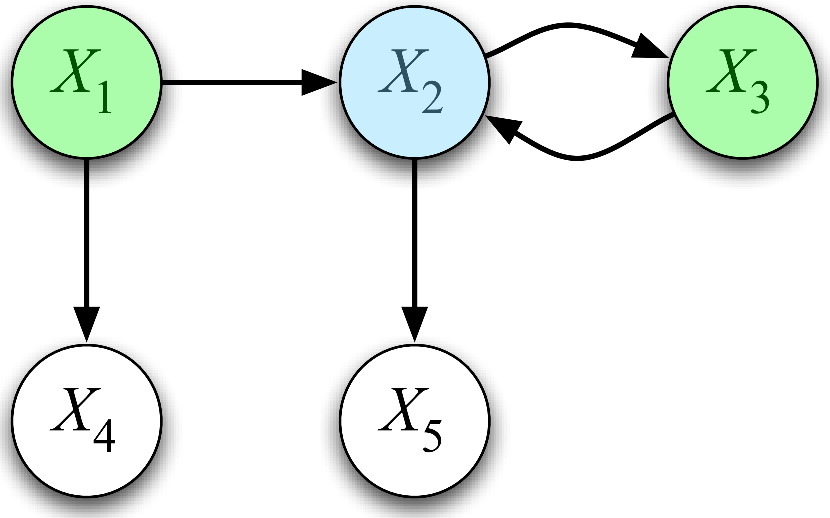
Multi component Markov process $X=(X_1, \ldots, X_N)$

Naive representation ➜ size of rate matrix is exponential in number of components

Compact representation

- Only one component changes its state in a single transition
- *Conditional rate matrices* - depend on parents $\mathbb{Q}^{i|\mathrm{Par}(i)}$

$$q_{(a_1,a_2,a_3,a_4,a_5)\rightarrow(a_1,b_2,a_3,a_4,a_5)} = q_{a_2\rightarrow b_2|(a_1,a_3)}$$



[Nodelman et al. 2002]

Continuous Time Bayesian Networks

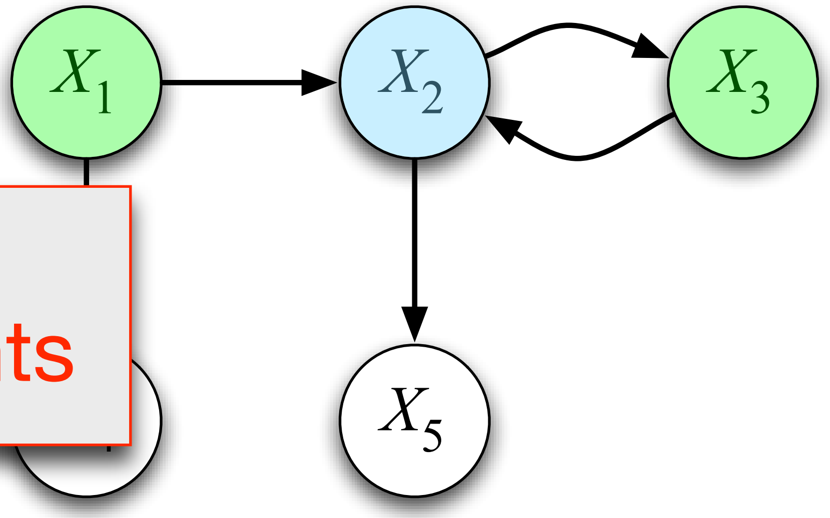
Multi component Markov process $X=(X_1, \dots, X_N)$

Naive representation ➜ size of rate matrix is exponential in number of components

Compact representation

- Only one component changes its state in a single transition
- *Conditional rate matrices* - depend on parents $\mathbb{Q}^{i|\mathrm{Par}(i)}$

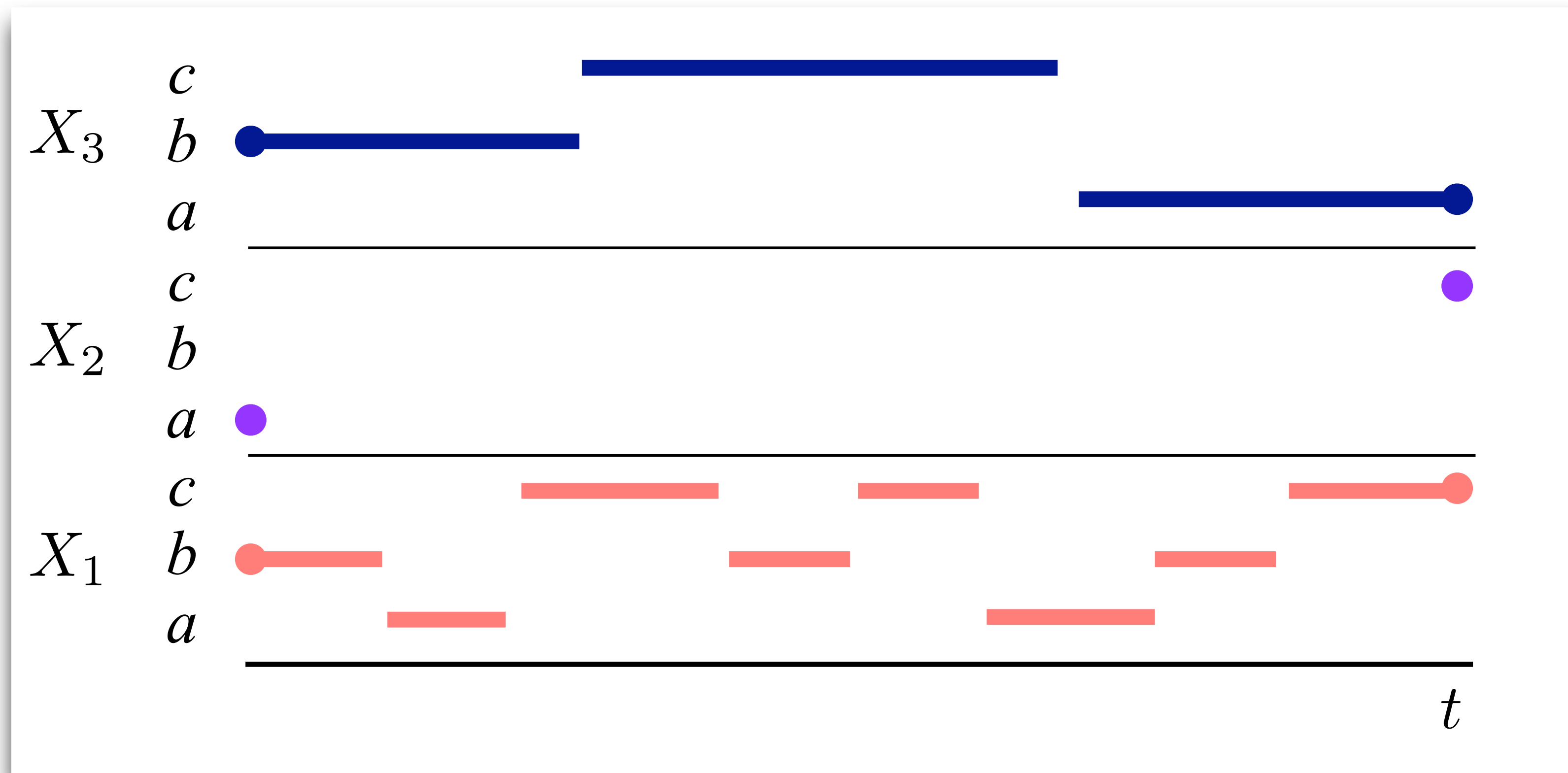
$$q_{(a_1,a_2,a_3,a_4,a_5)\to(a_1,b_2,a_3,a_4,a_5)} = q_{a_2\to b_2|(a_1,a_3)}$$

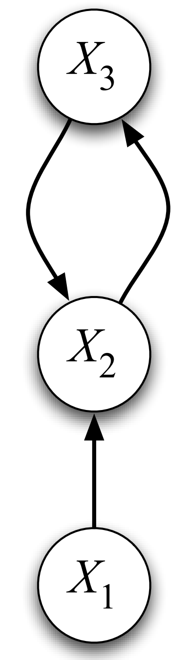


However, exact inference is still exponential in the number of components

[Nodelman et al. 2002]

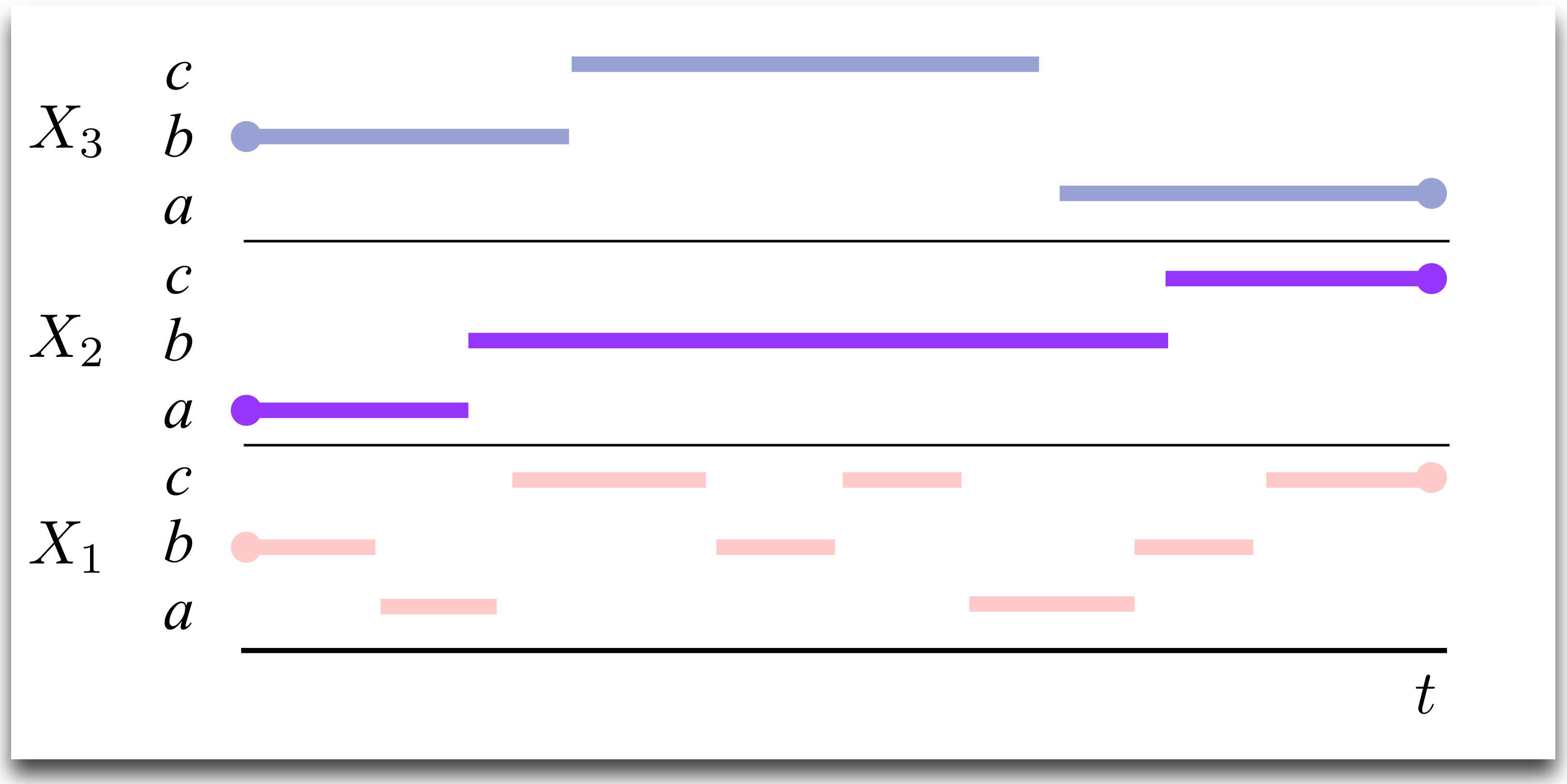
Gibbs Sampling in Continuous Time

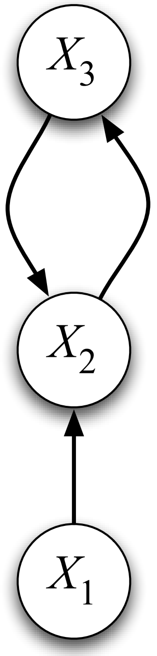




Iterate sampling each **trajectory** given the others

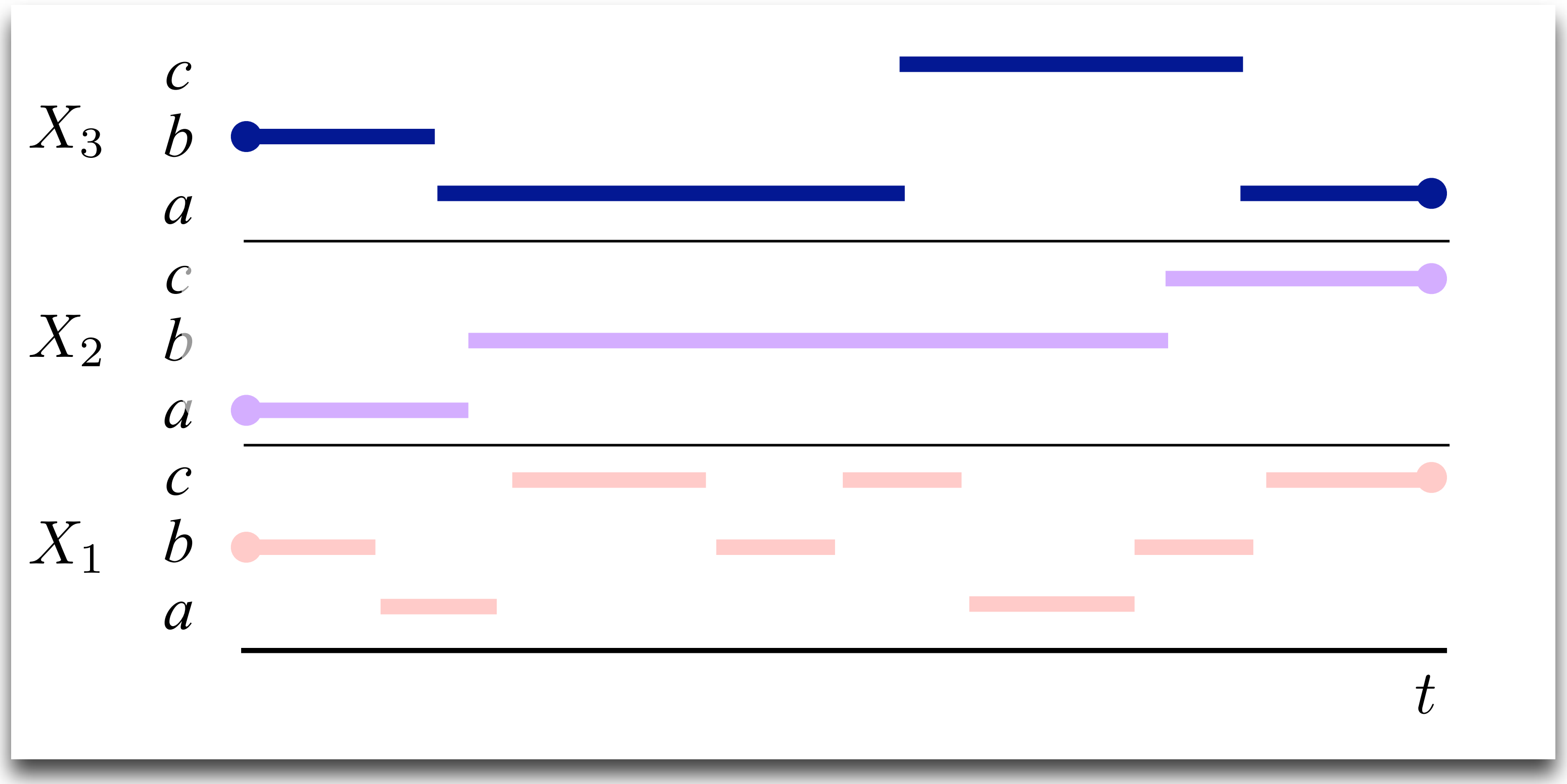
Gibbs Sampling in Continuous Time

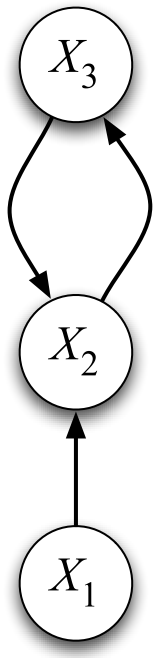




Iterate sampling each **trajectory** given the others

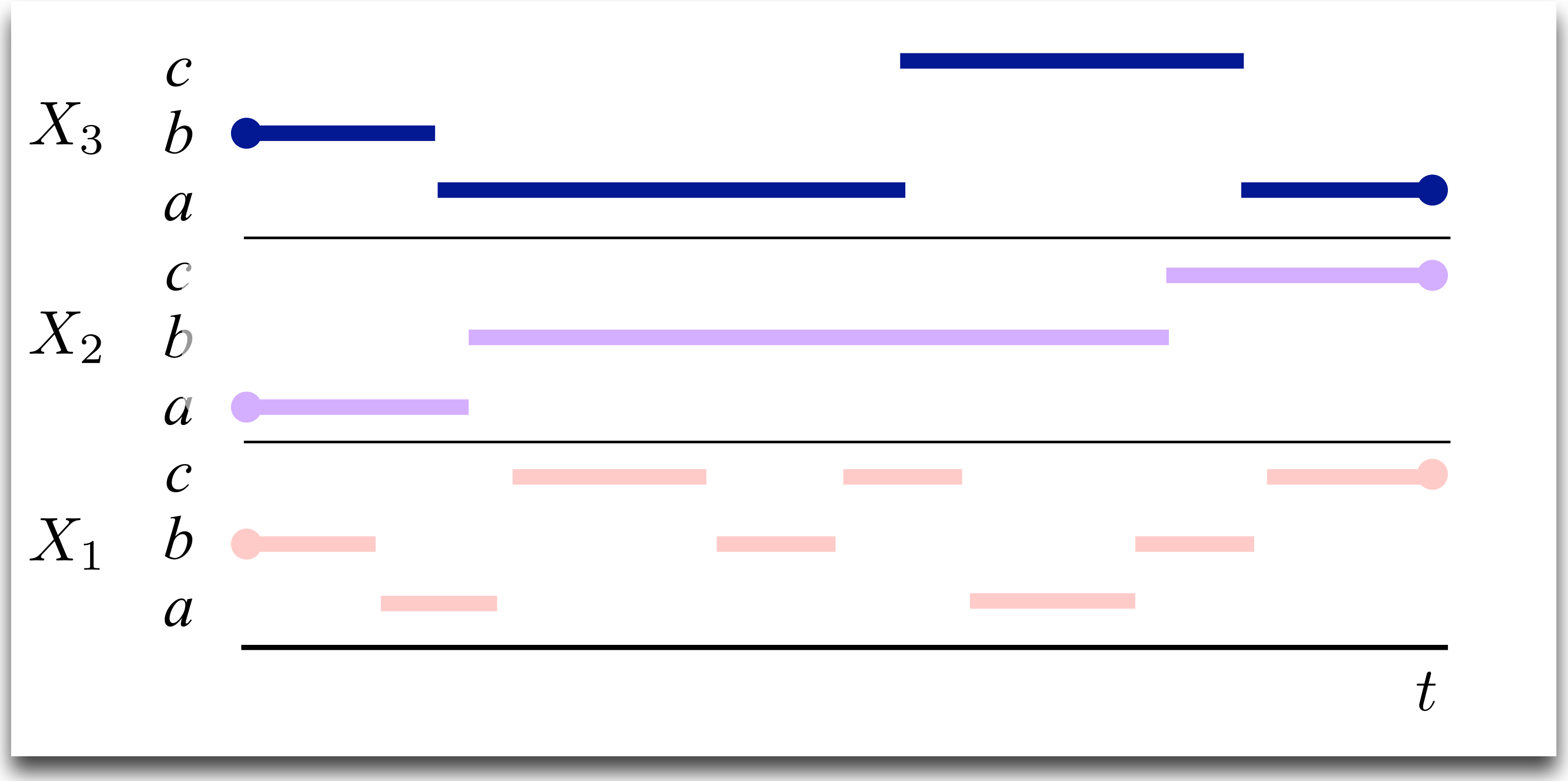
Gibbs Sampling in Continuous Time

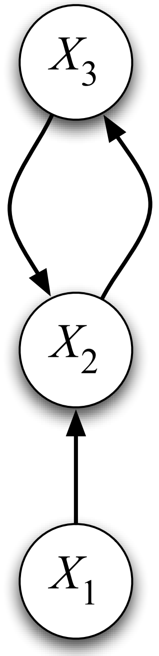




Iterate sampling each **trajectory** given the others

Gibbs Sampling in Continuous Time

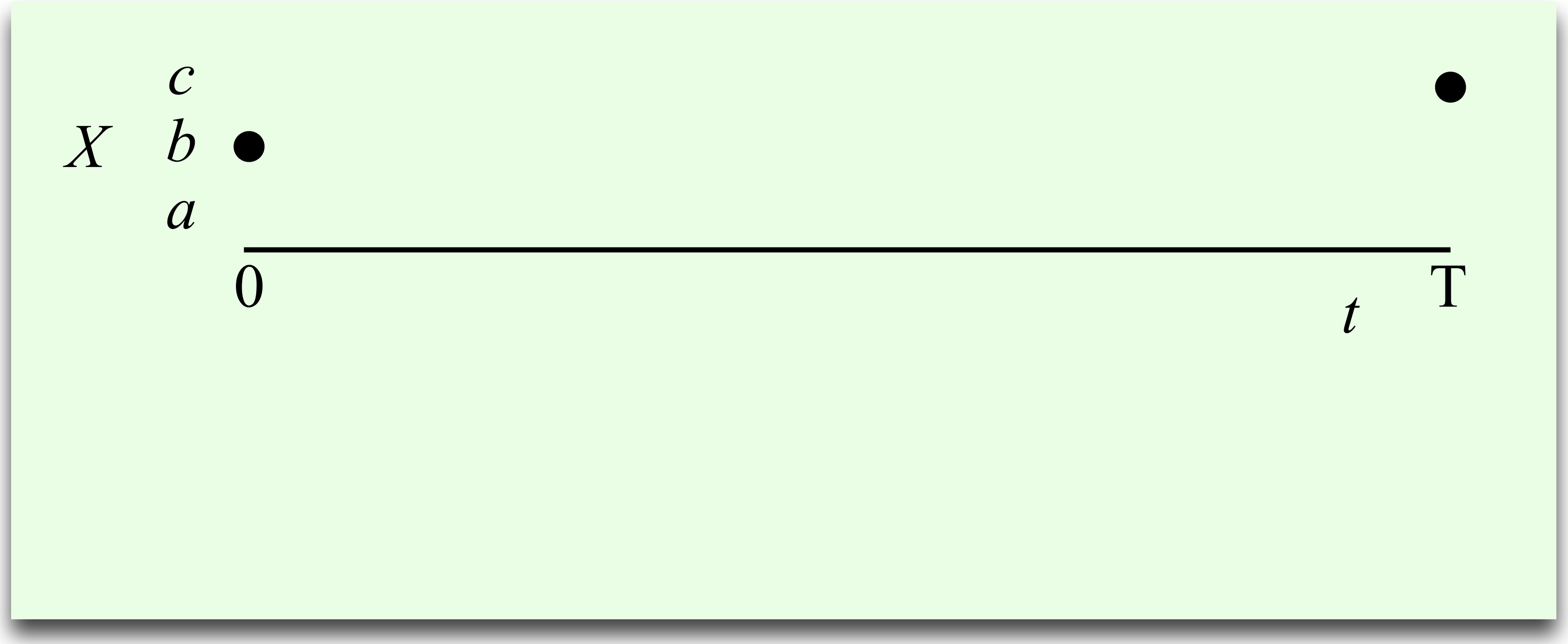




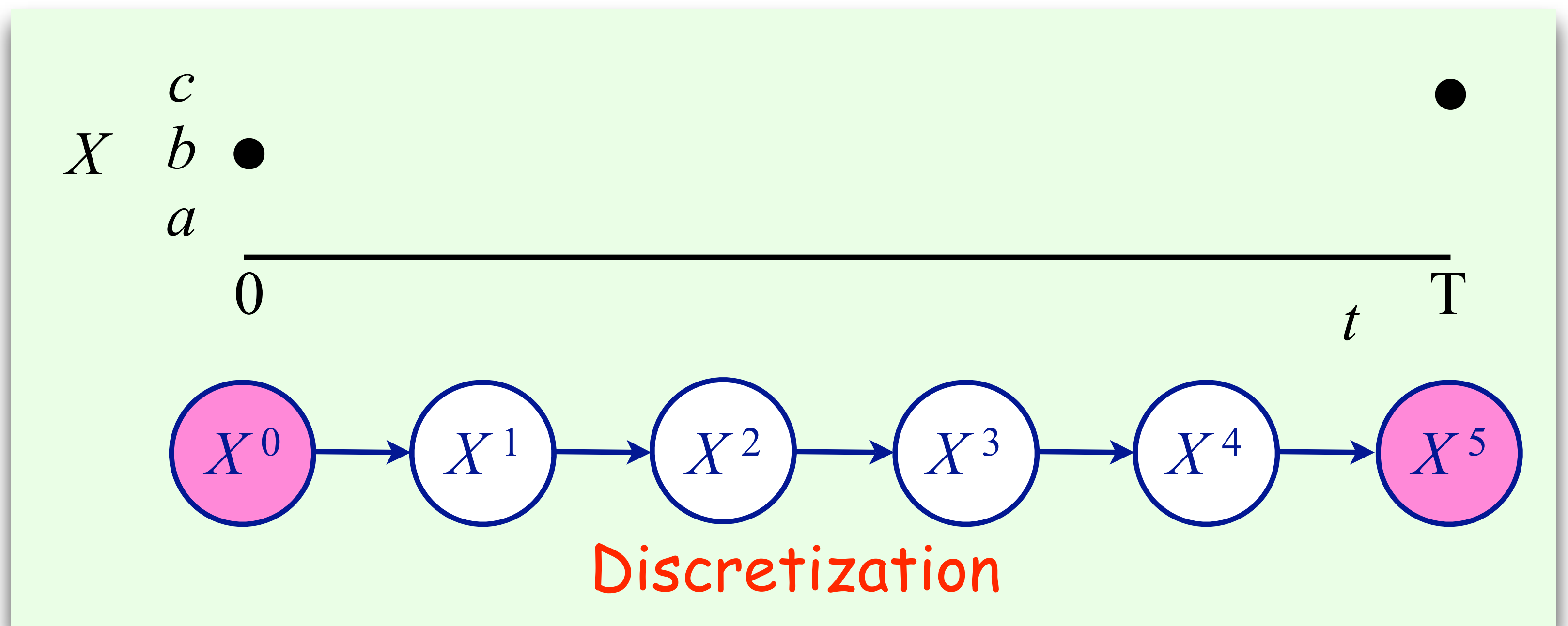
Iterate sampling each **trajectory** given the others

How do we sample a single component given the others?

Sampling From a Single Component - Naive Approach



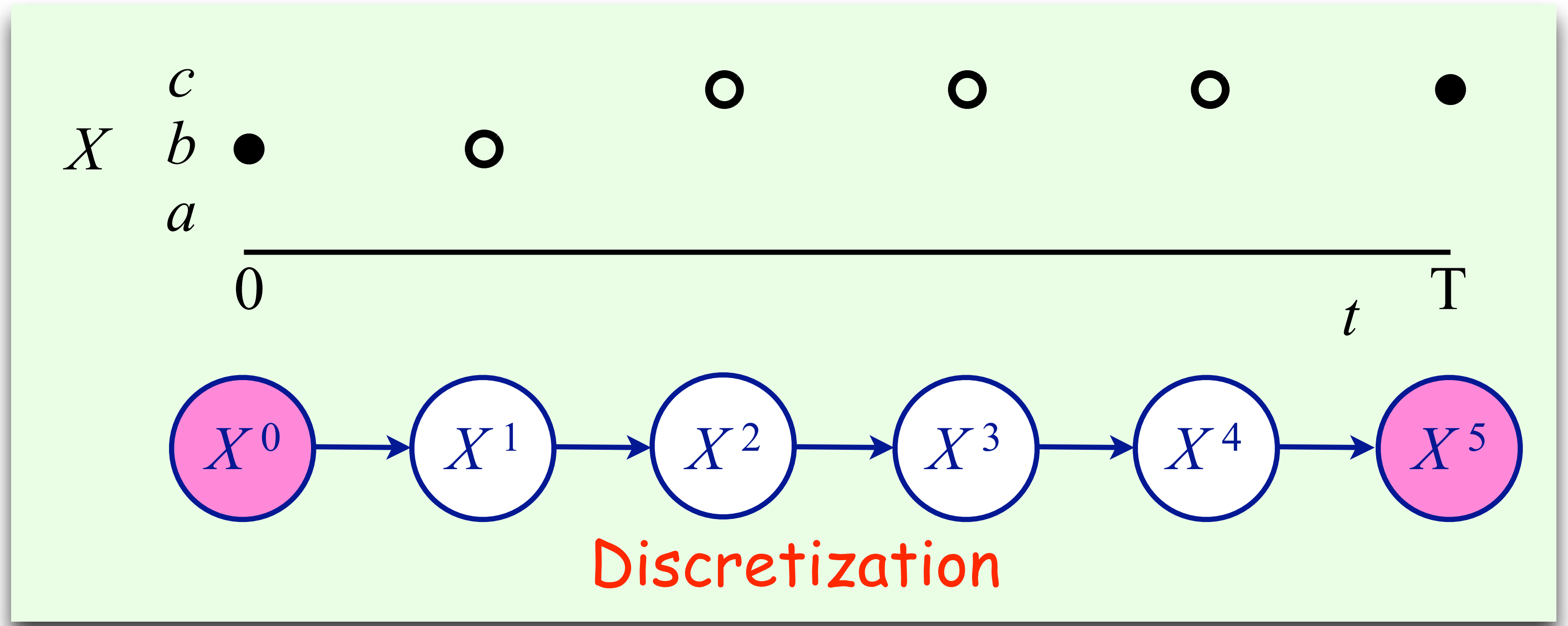
Sampling From a Single Component - Naive Approach



Transition Matrix between two discrete time points:

$$P(X^i | X^{i-1}) = \left[e^{\mathbb{Q} \frac{T}{N}} \right]$$

Sampling From a Single Component - Naive Approach

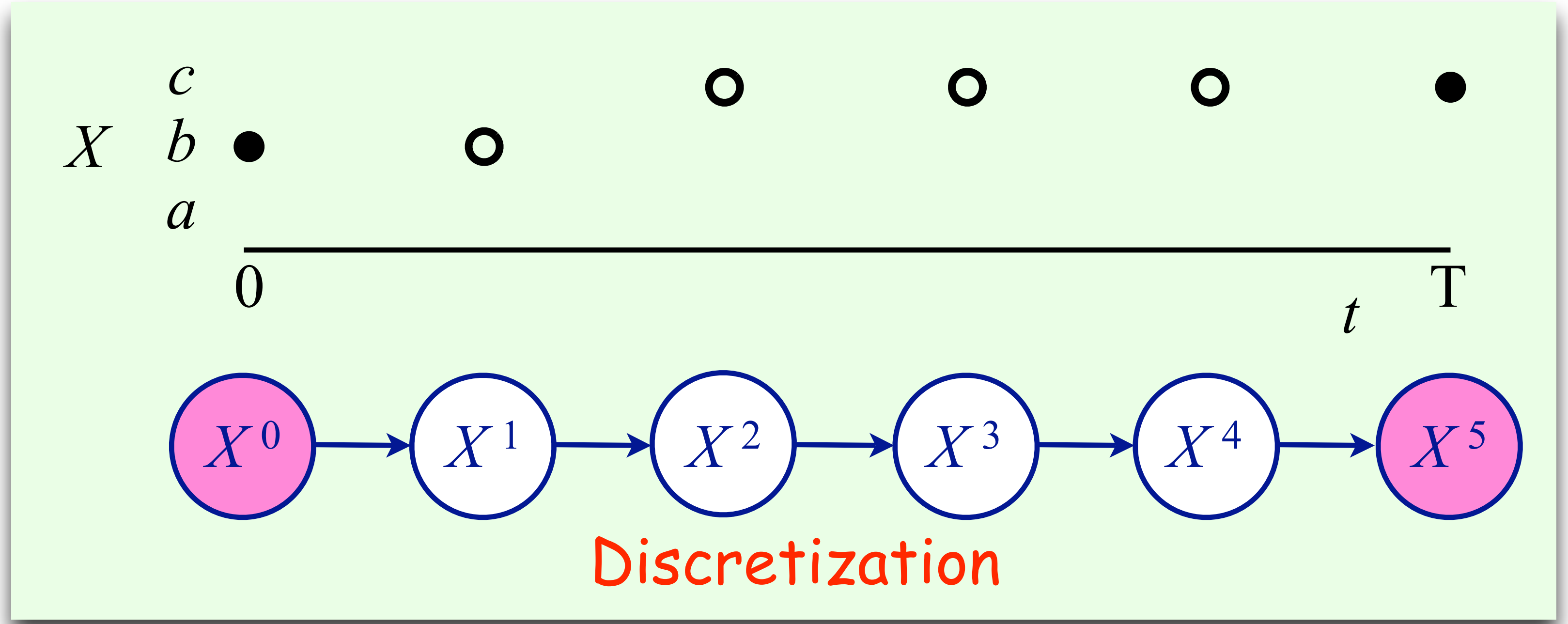


Transition Matrix between two discrete time points:

$$P(X^i|X^{i-1}) = \left[e^{\mathbb{Q}\frac{T}{N}}\right]$$

Backward propagation - forward sampling

Sampling From a Single Component - Naive Approach



Transition Matrix between two discrete time points:

$$P(X^i | X^{i-1}) = \left[e^{\mathbb{Q}\frac{T}{N}} \right]$$

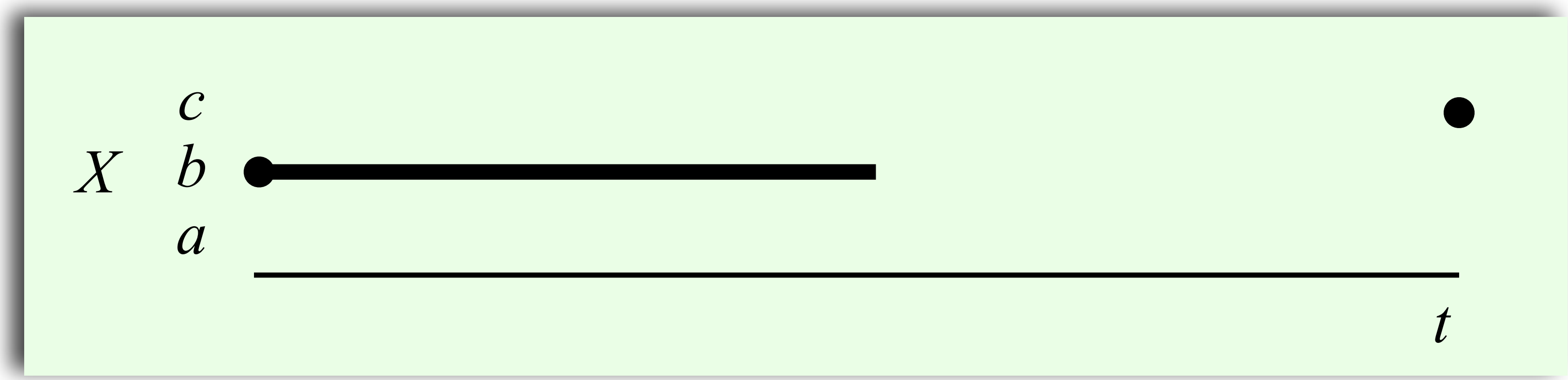
Backward propagation - forward sampling

1. Assumes a specific time scale
2. Complexity - linear in the number of sampling points

Cumulative Distribution Function of the Next Transition time

Conditional probability to leave initial state before time t

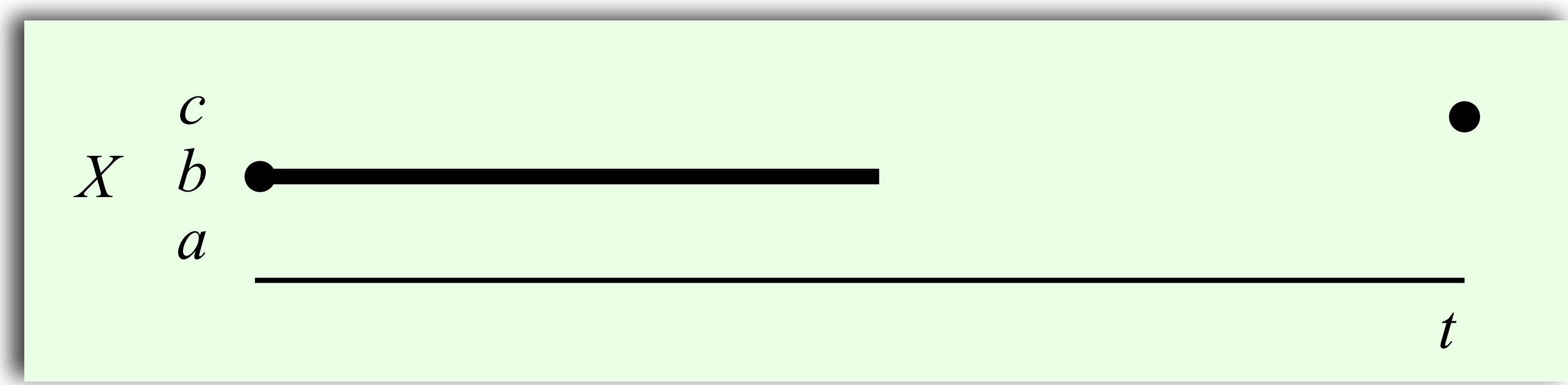
$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$



Cumulative Distribution Function of the Next Transition time

Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$



Cumulative Distribution Function of the Next Transition time

Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$

Probability to stay in x^0 at least until time t

Cumulative Distribution Function of the Next Transition time

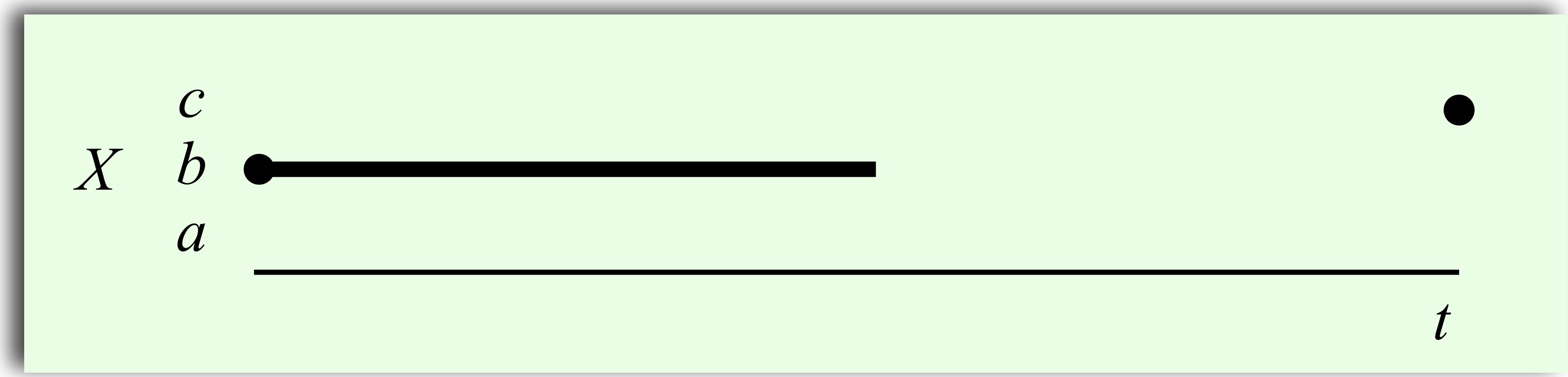
Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$

Probability to stay in x^0 at least until time t

From *Markov property*:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T) = \Pr(X^{(0,t]} = x^0 | x^0) \Pr(x^T | X^{(t)} = x^0)$$



Cumulative Distribution Function of the Next Transition time

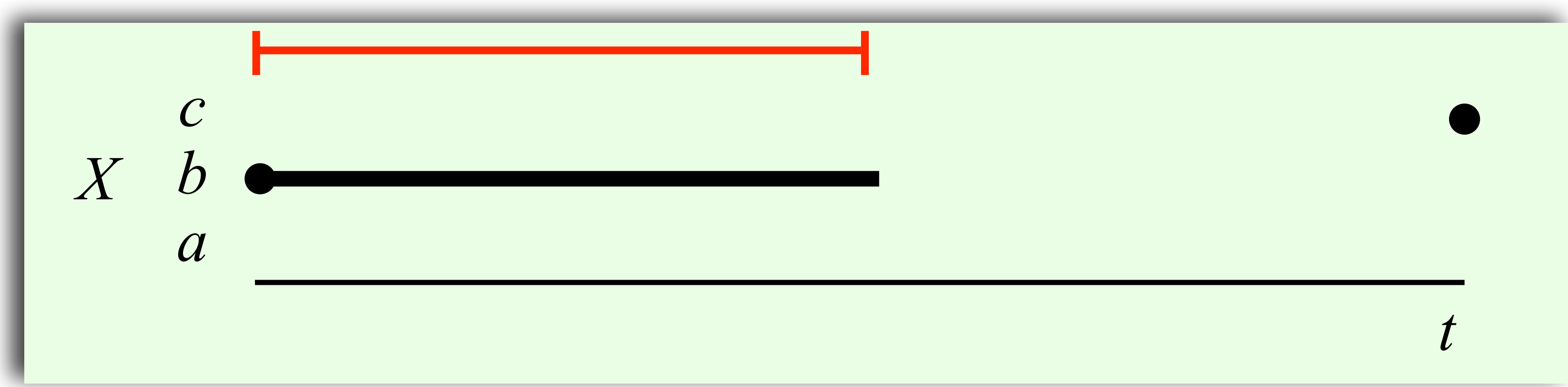
Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$

Probability to stay in x^0 at least until time t

From *Markov property*:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T) = \Pr(X^{(0,t]} = x^0 | x^0) \Pr(x^T | X^{(t)} = x^0)$$



Cumulative Distribution Function of the Next Transition time

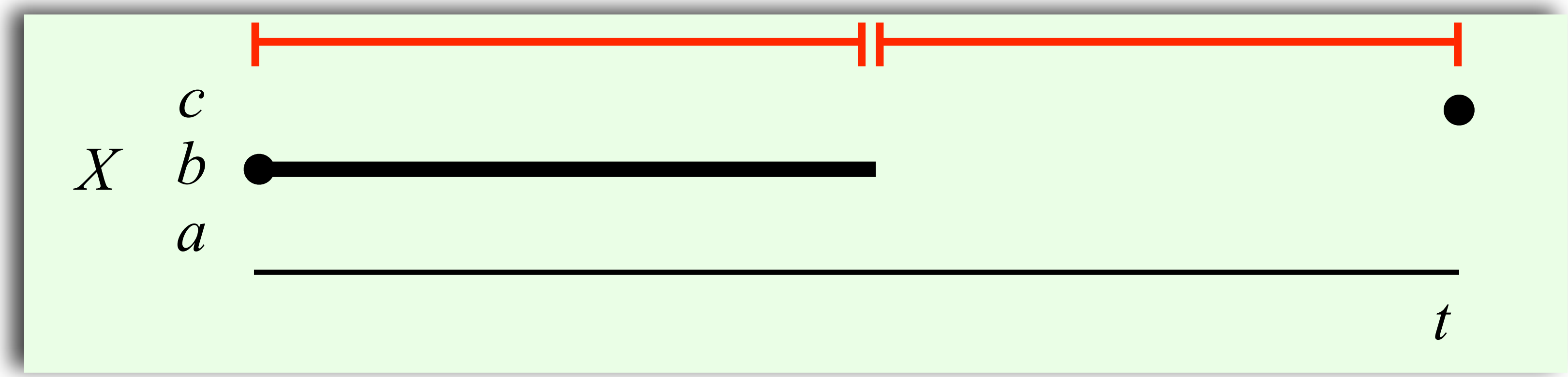
Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$

Probability to stay in x^0 at least until time t

From *Markov property*:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T) = \Pr(X^{(0,t]} = x^0 | x^0) \Pr(x^T | X^{(t)} = x^0)$$



Cumulative Distribution Function of the Next Transition time

Conditional probability to leave initial state before time t

$$F(t) = 1 - \Pr(X^{(0,t]} = x^0 | x^0, x^T)$$

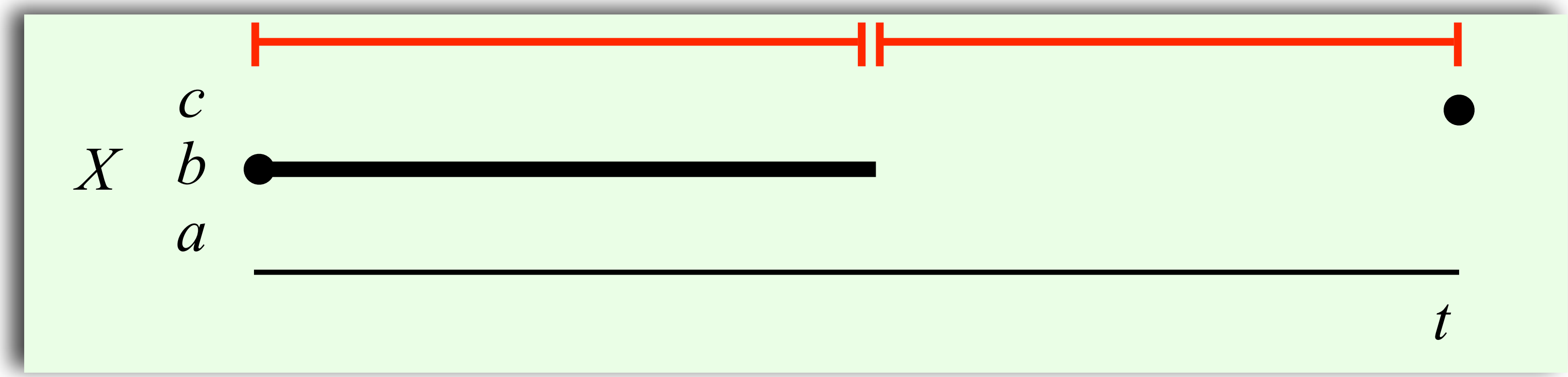
Probability to stay in x^0 at least until time t

From *Markov property*:

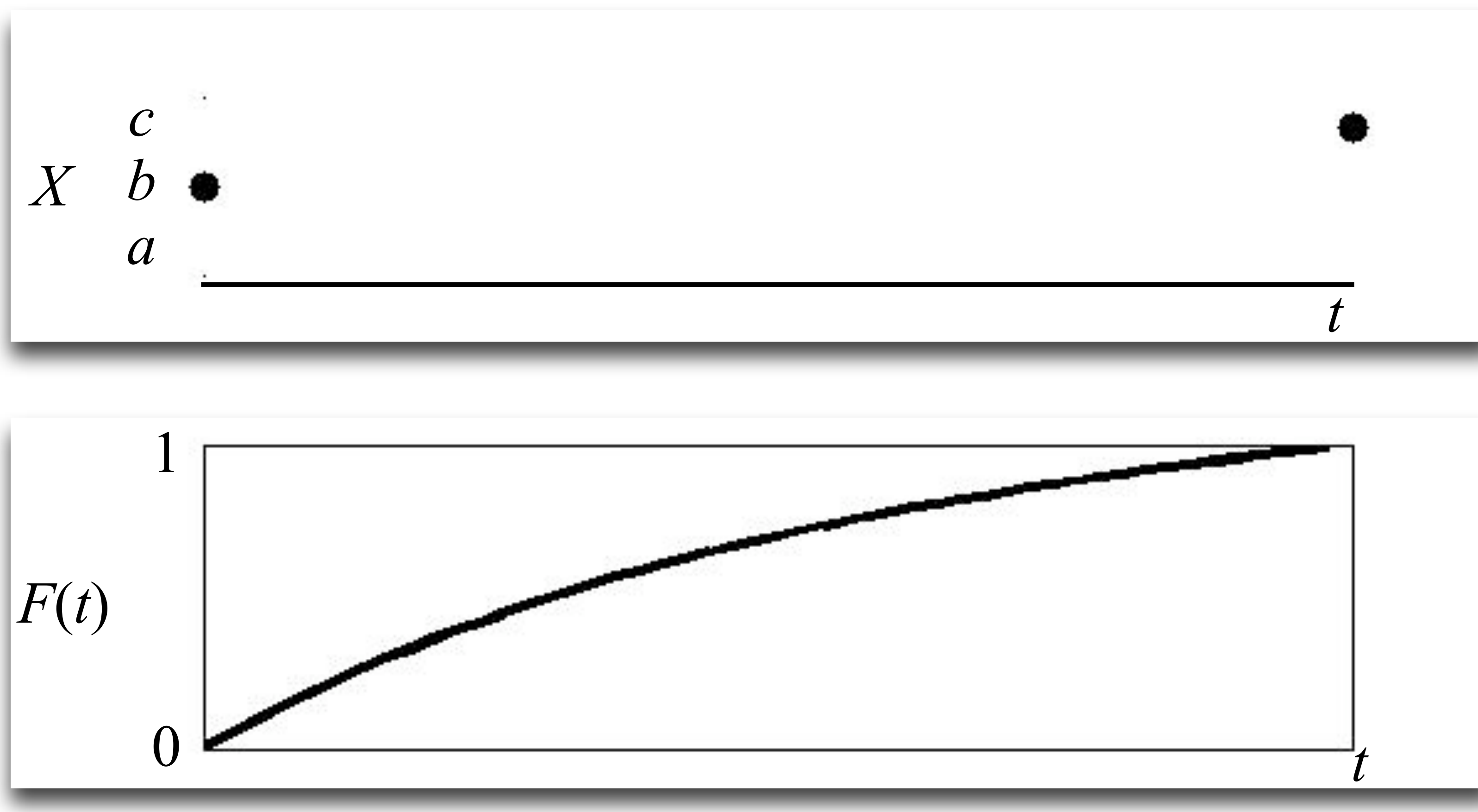
$$\Pr(X^{(0,t]} = x^0 | x^0, x^T) = \Pr(X^{(0,t]} = x^0 | x^0) \Pr(x^T | X^{(t)} = x^0)$$

Exploiting *homogeneity*:

$$e^{q_{x_0 \to x_0} t} \qquad [e^{\mathbb{Q}(T-t)}]_{x_t, x_T}$$

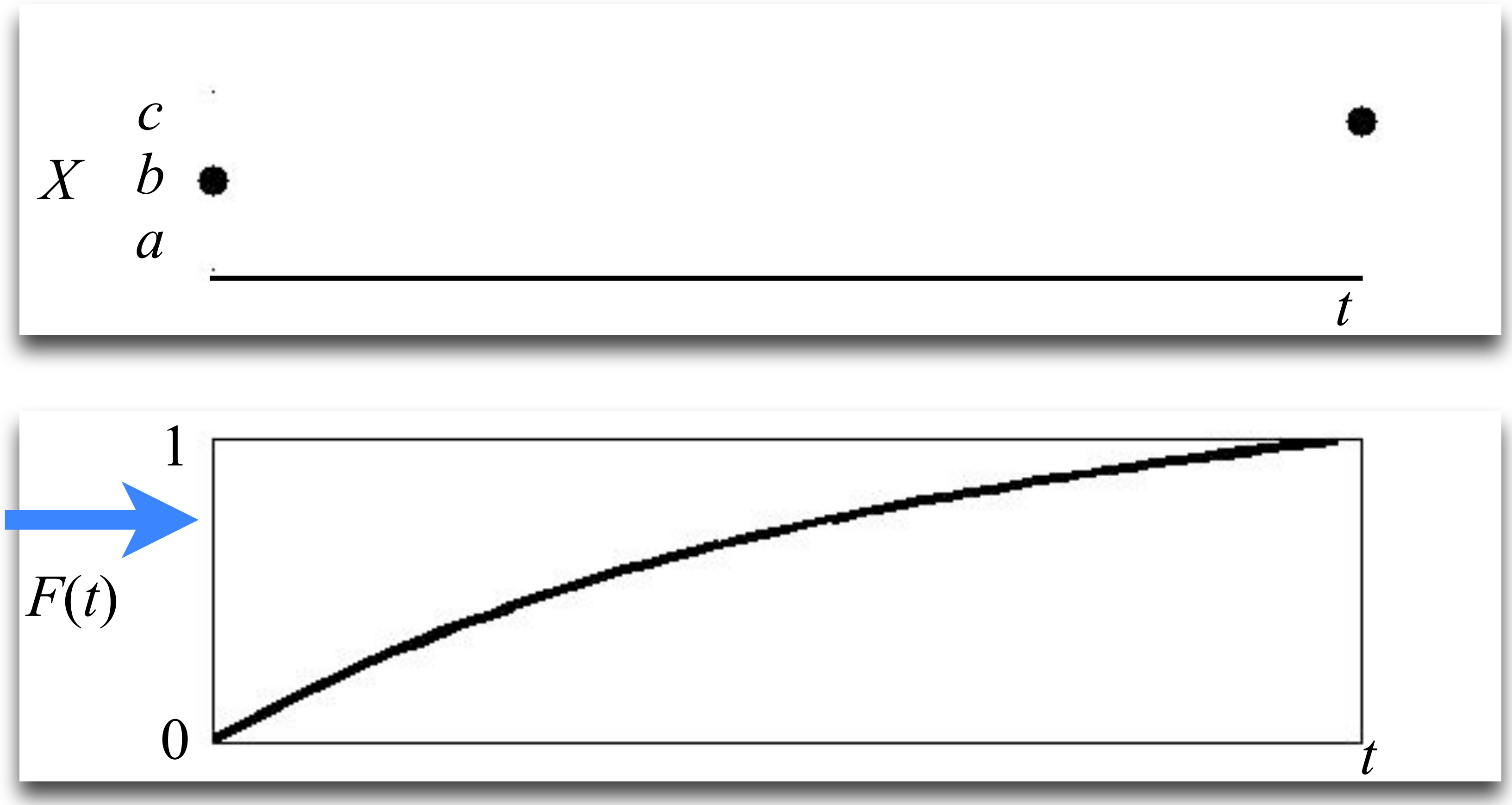


Single Component -
Sampling Next transition time

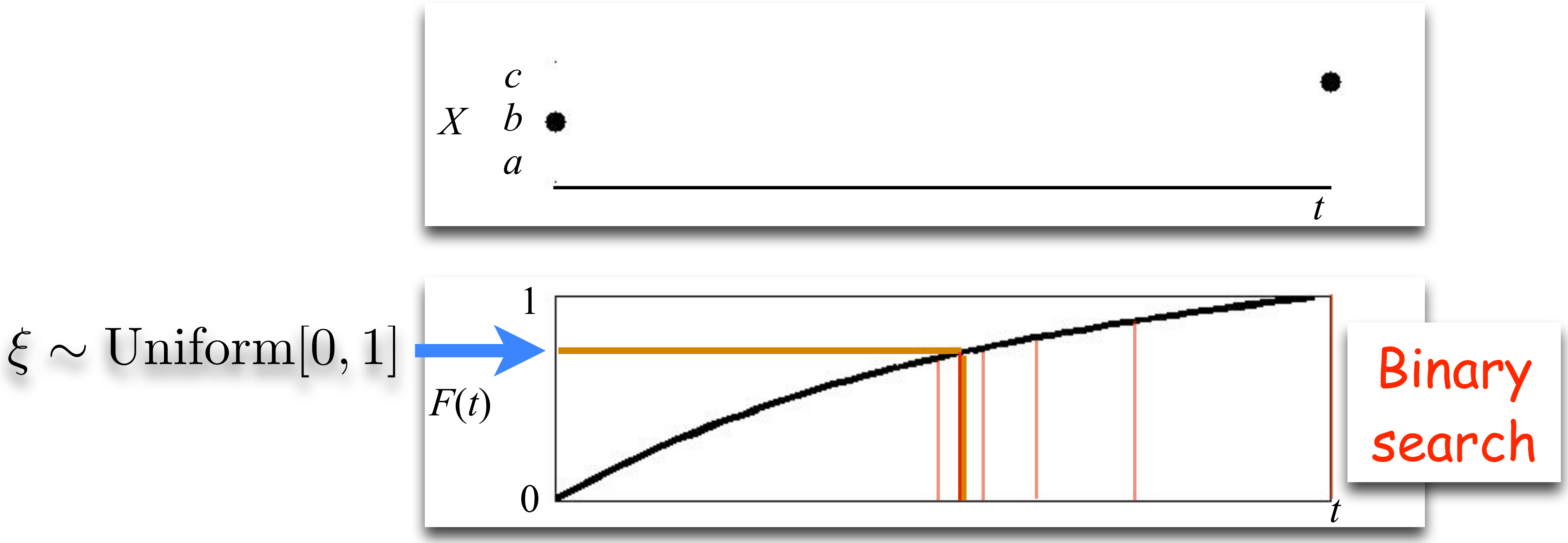


Single Component -
Sampling Next transition time

$\xi \sim \text{Uniform}[0, 1]$

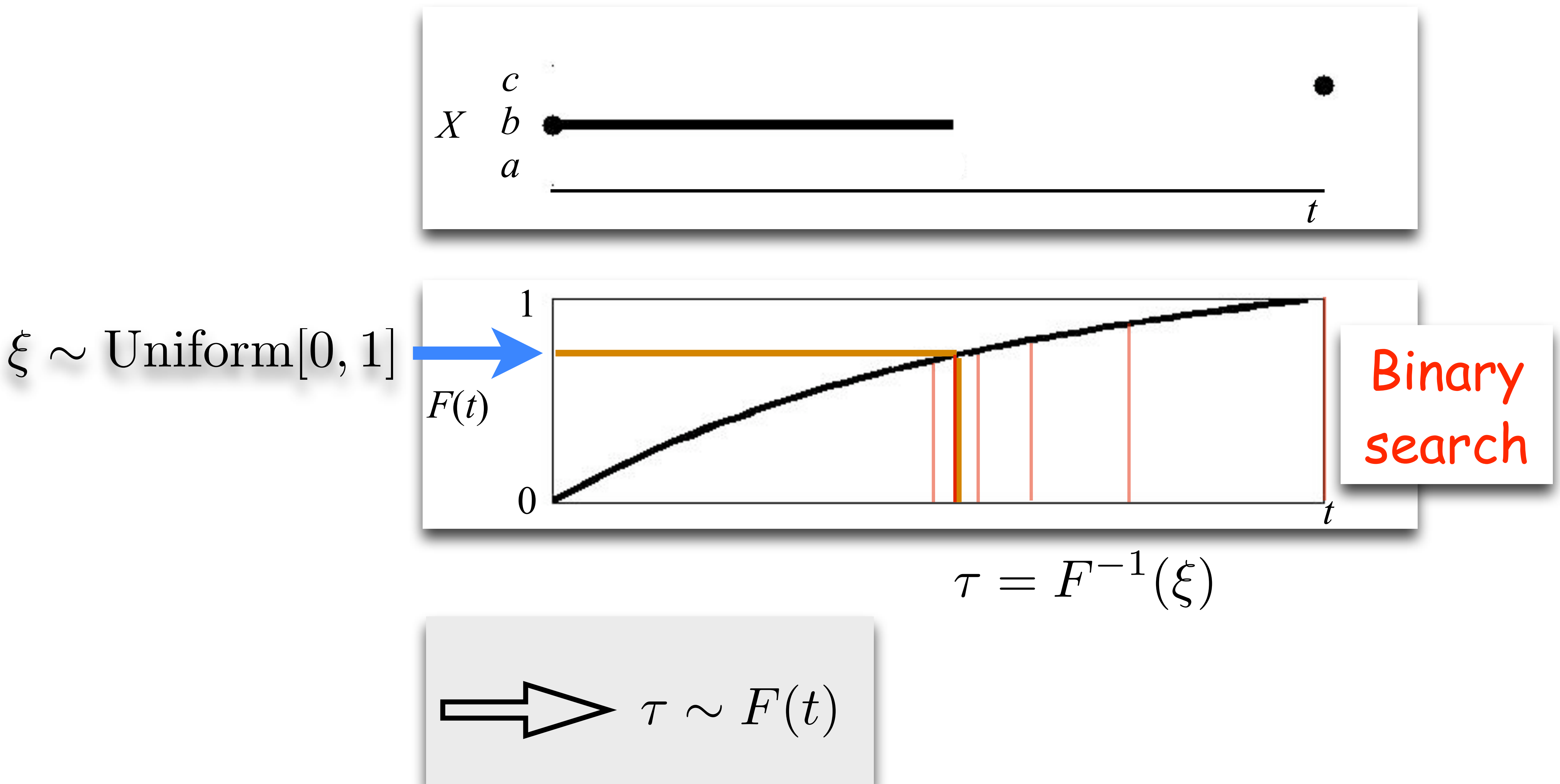


Single Component - Sampling Next transition time



$$\tau = F^{-1}(\xi)$$

Single Component - Sampling Next transition time

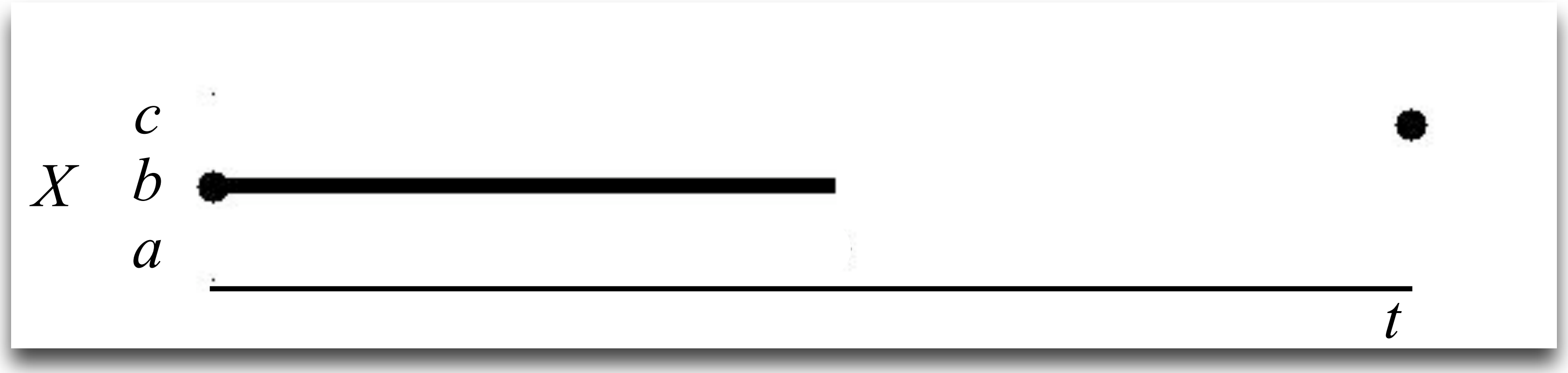


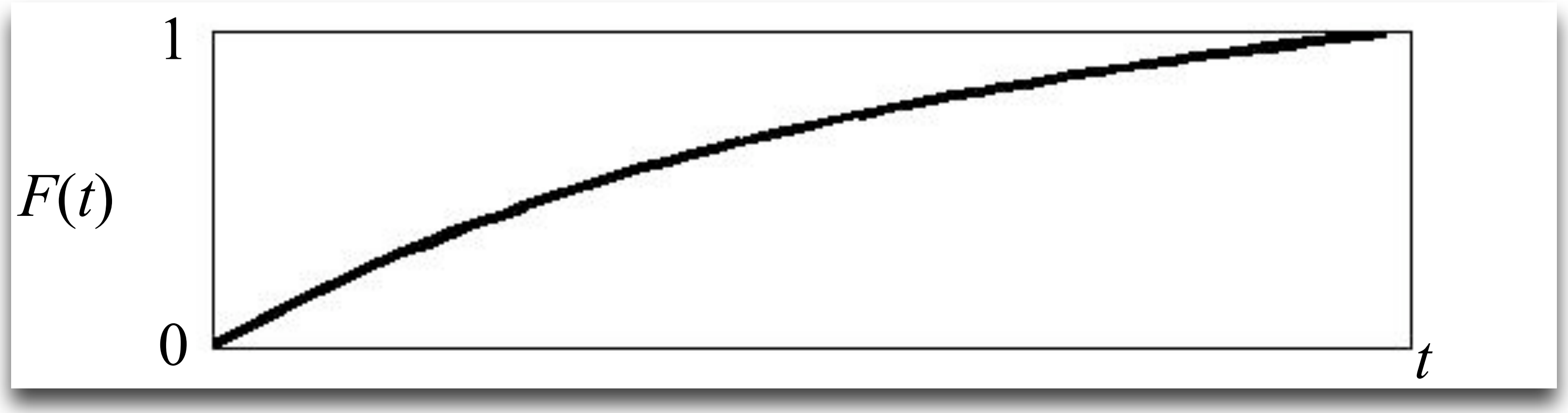
Sampling From a Single Component - Sequential procedure

Iterate between sampling transition times and sampling the next state

Sampling From a Single Component - Sequential procedure

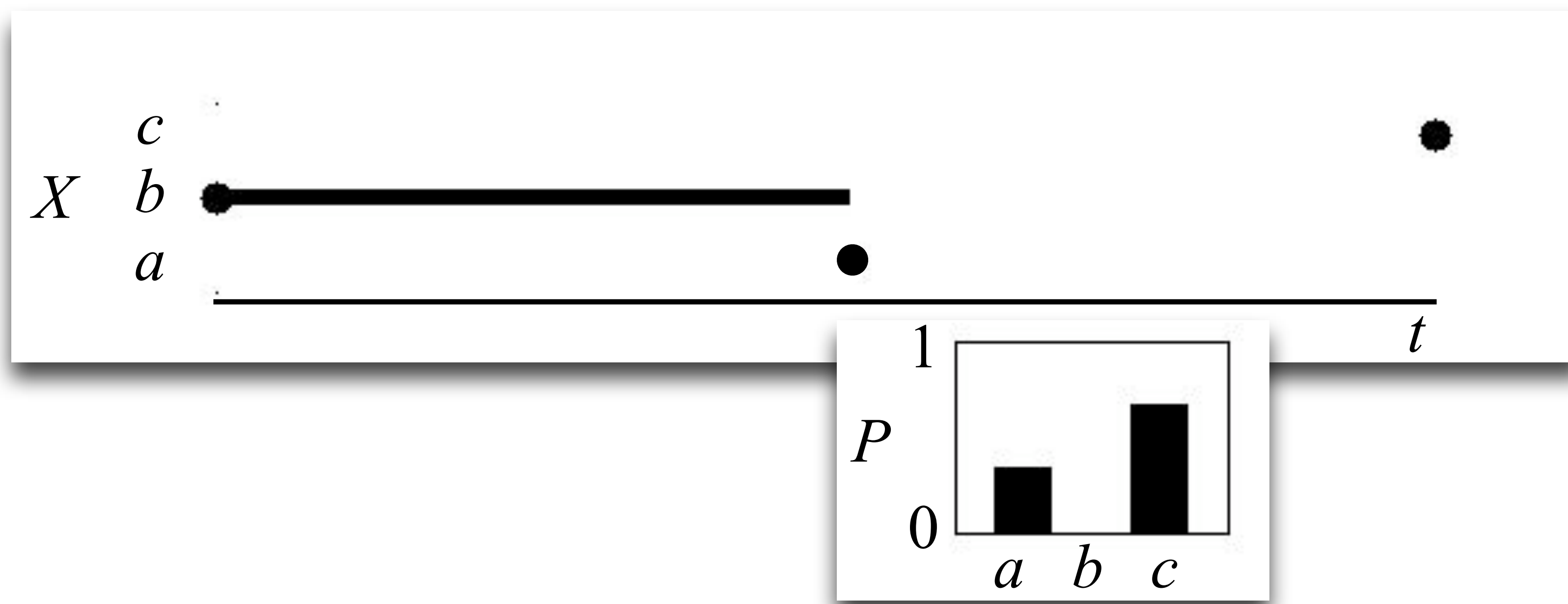
Iterate between sampling transition times and sampling the next state





Sampling From a Single Component - Sequential procedure

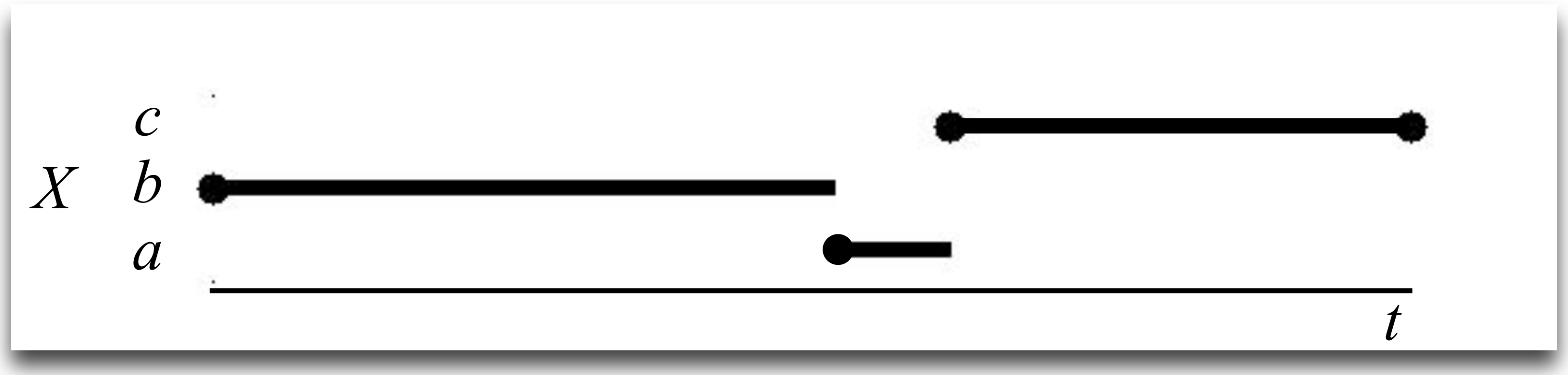
Iterate between sampling transition times and sampling the next state

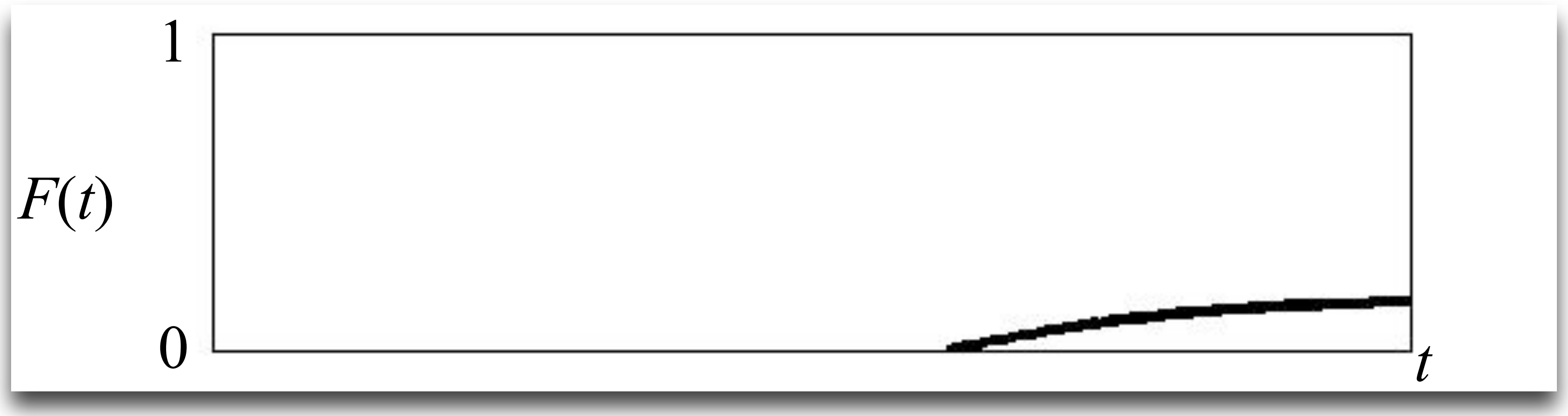


t

Sampling From a Single Component - Sequential procedure

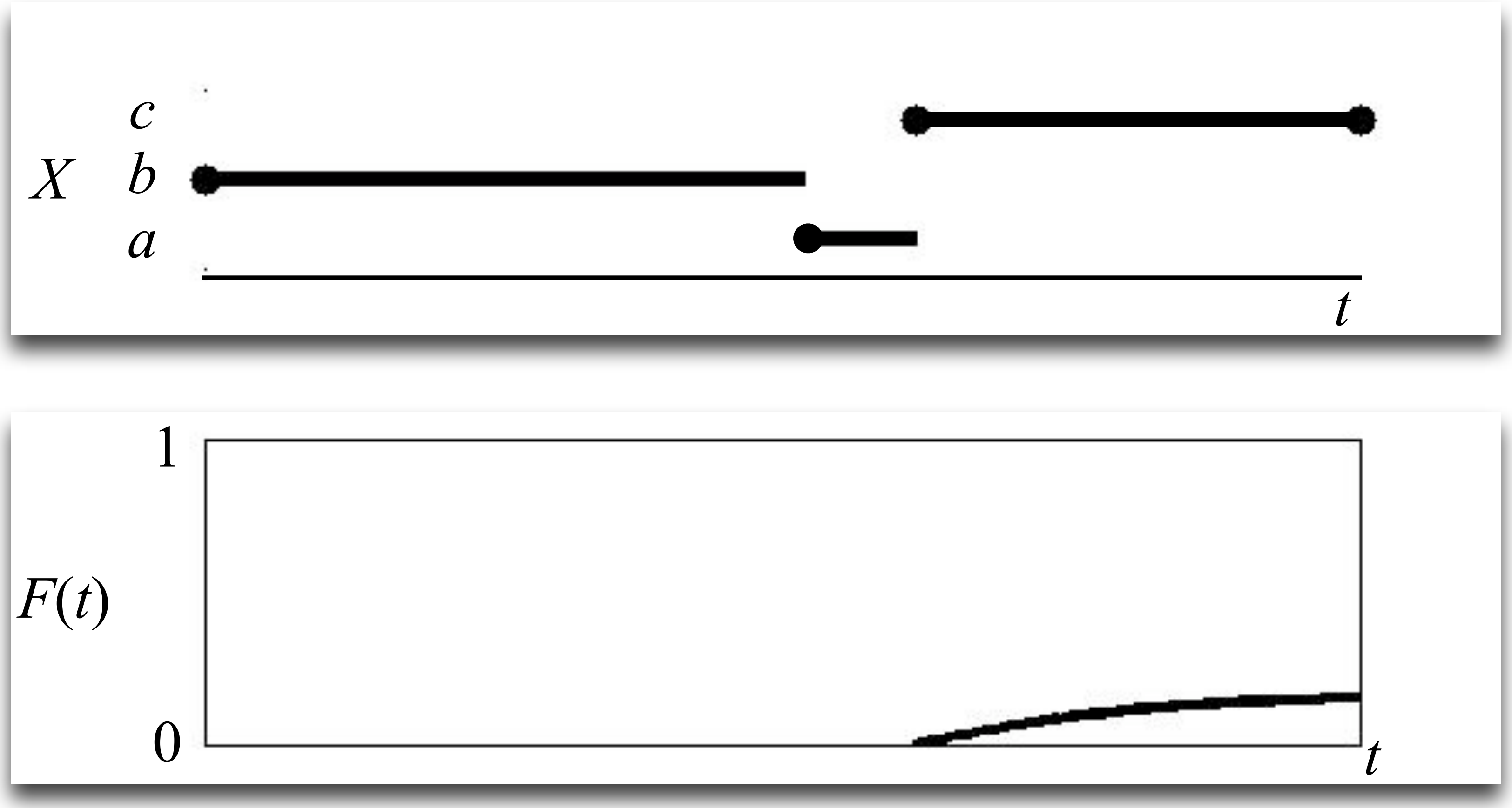
Iterate between sampling transition times and sampling the next state





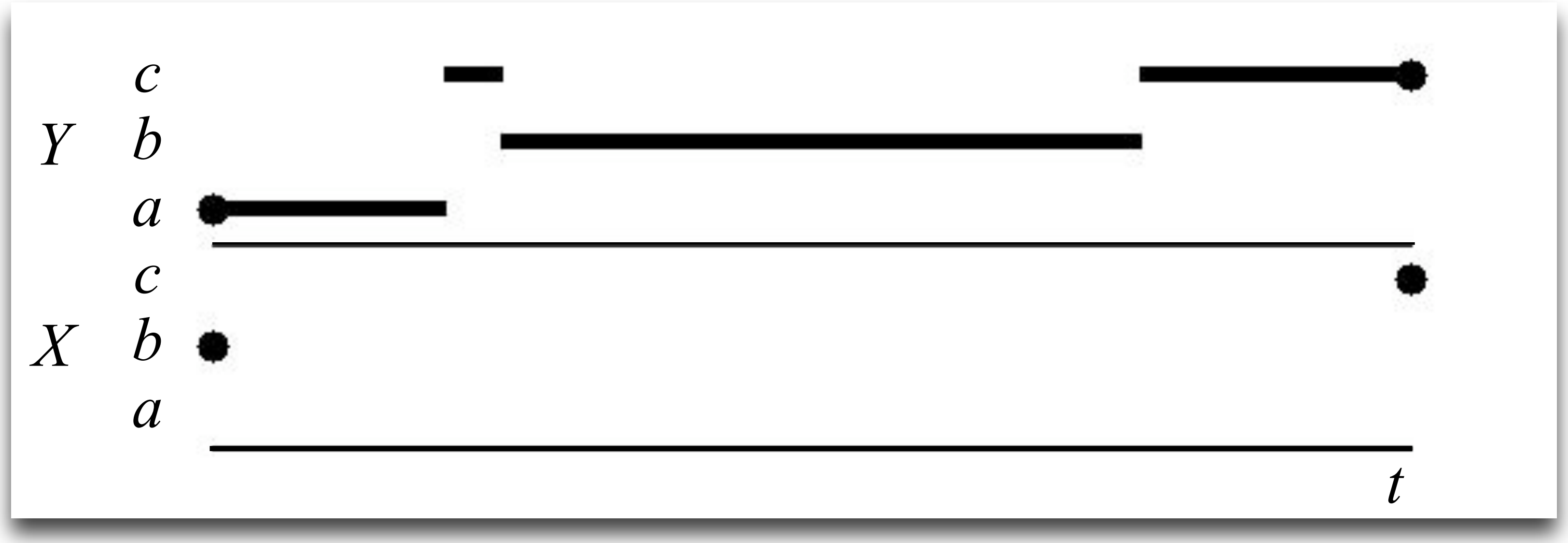
Sampling From a Single Component - Sequential procedure

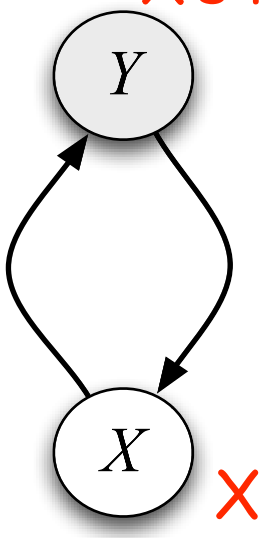
Iterate between sampling transition times and sampling the next state



➡ Linear in the number of **transitions** in X_i

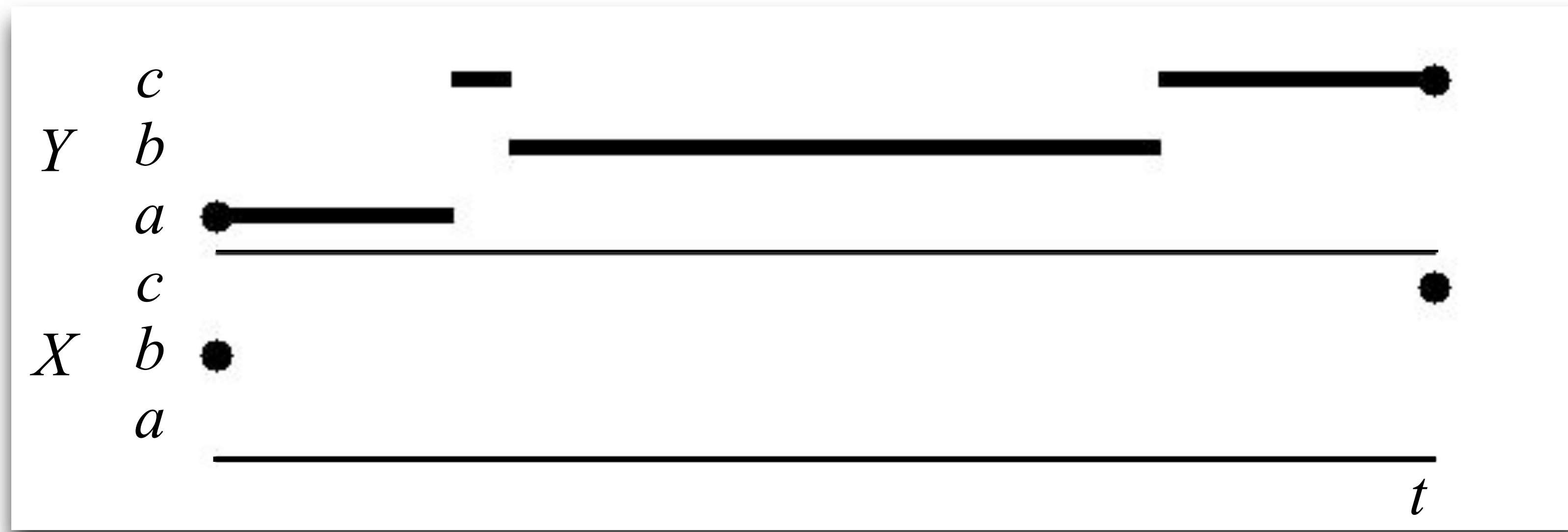
Conditional Sampling





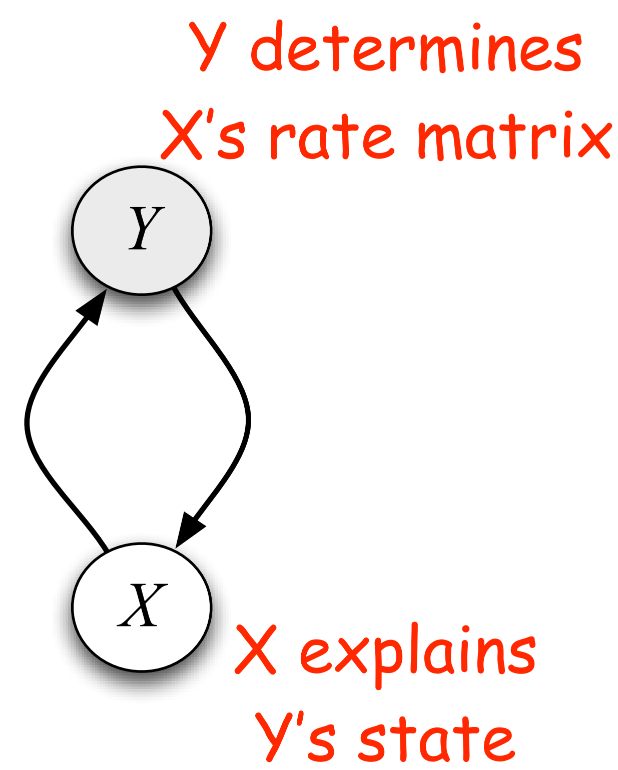
t

Conditional Sampling



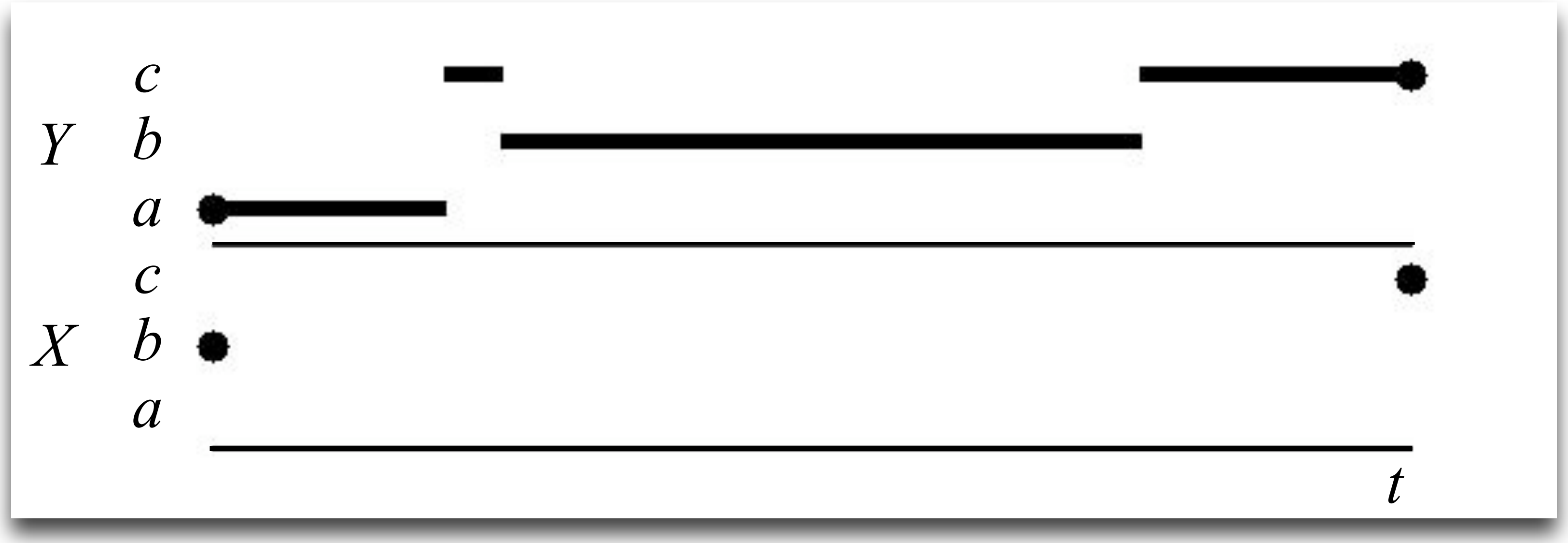
Theorem: X satisfies Markov property

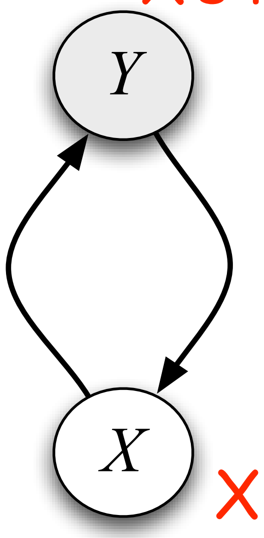
➡ Basic procedure remains the same



t

Conditional Sampling

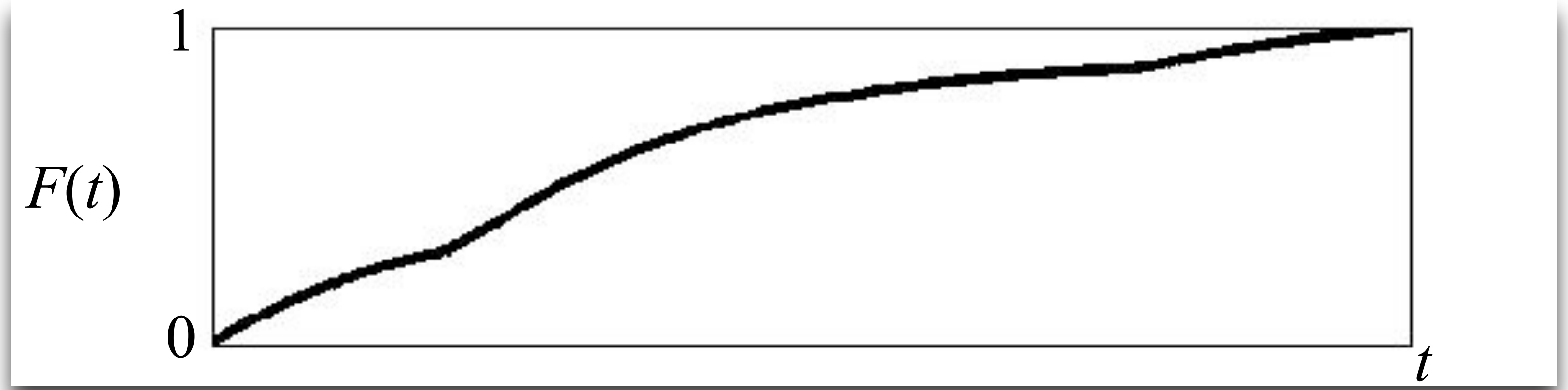




Theorem: X satisfies Markov property

➡ Basic procedure remains the same

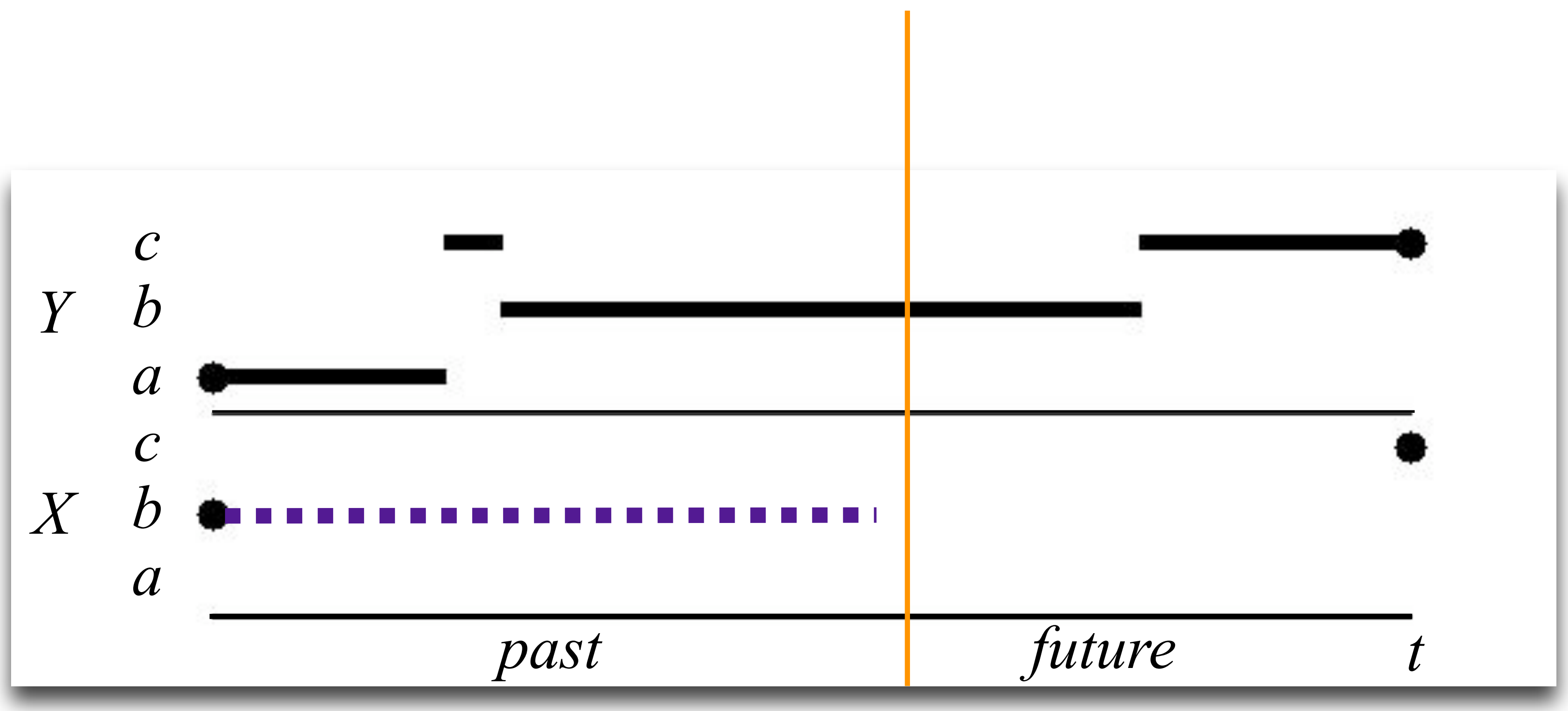
However: Posterior is not time-homogeneous



Distribution of Next Transition

Probability to remain:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T, y^{[0,T]}) = \frac{p^{\text{past}}(t) \cdot p^{\text{future}}_{x^0}(t)}{p^{\text{future}}_{x^0}(0)}$$

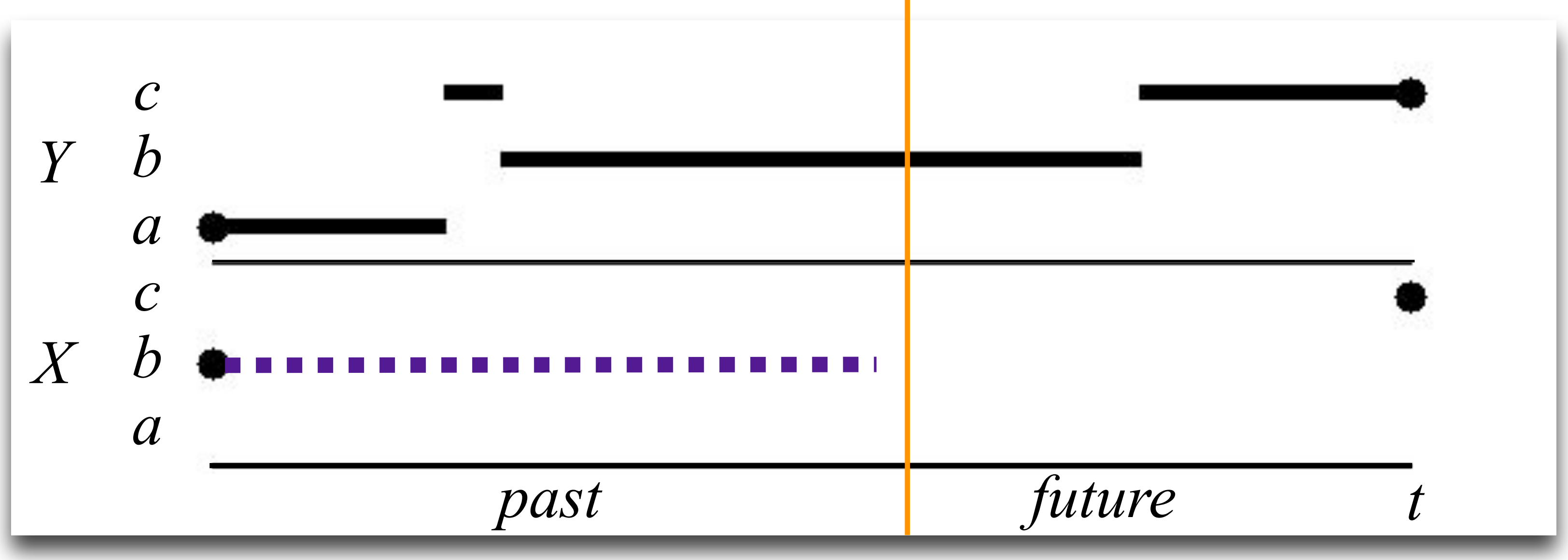


Distribution of Next Transition

Probability to remain:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T, y^{[0,T]}) = \frac{p^{\text{past}}(t) \cdot p^{\text{future}}_{x^0}(t)}{p^{\text{future}}_{x^0}(0)}$$

$$\Pr(X^{(0,t]} = x^0, y^{(0,t]} | x^0, y^0)$$

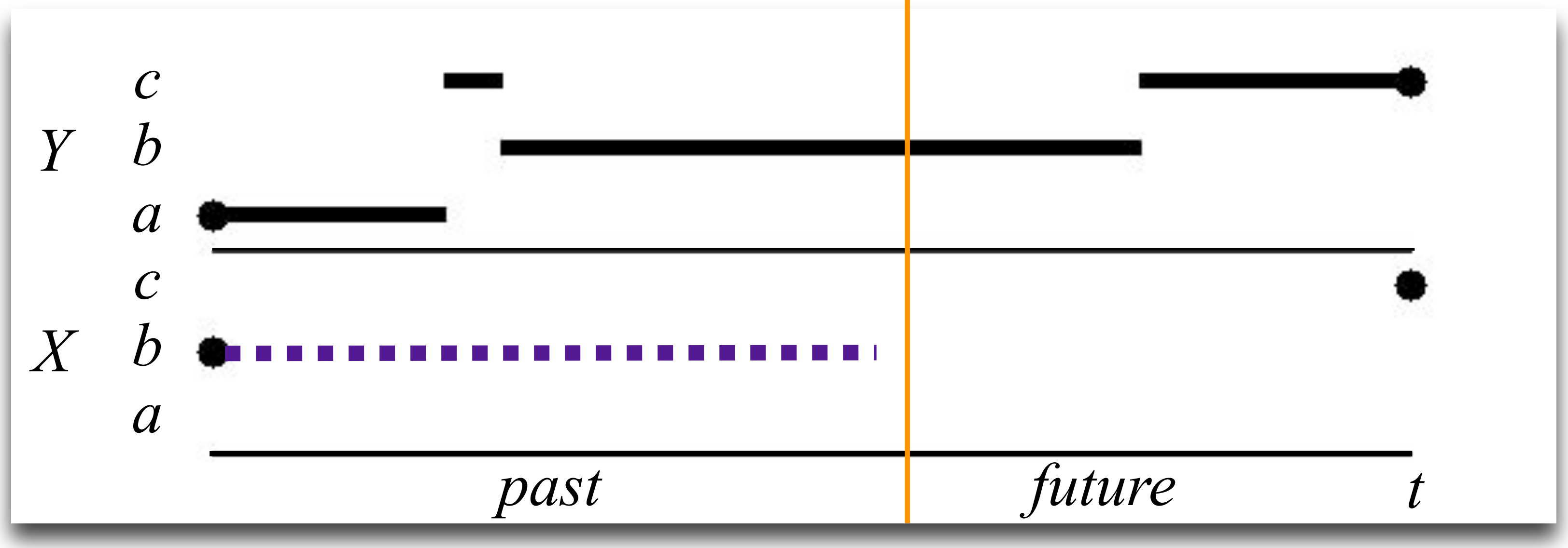


Distribution of Next Transition

Probability to remain:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T, y^{[0,T]}) = \frac{p^{\text{past}}(t)\, p^{\text{future}}_{x^0}(t)}{p^{\text{future}}_{x^0}(0)}$$

$$\Pr(X^{(0,t]} = x^0, y^{(0,t]} | x^0, y^0) \qquad \Pr(x^T, y^{(t,T]} | X^t = x^0, y^t)$$



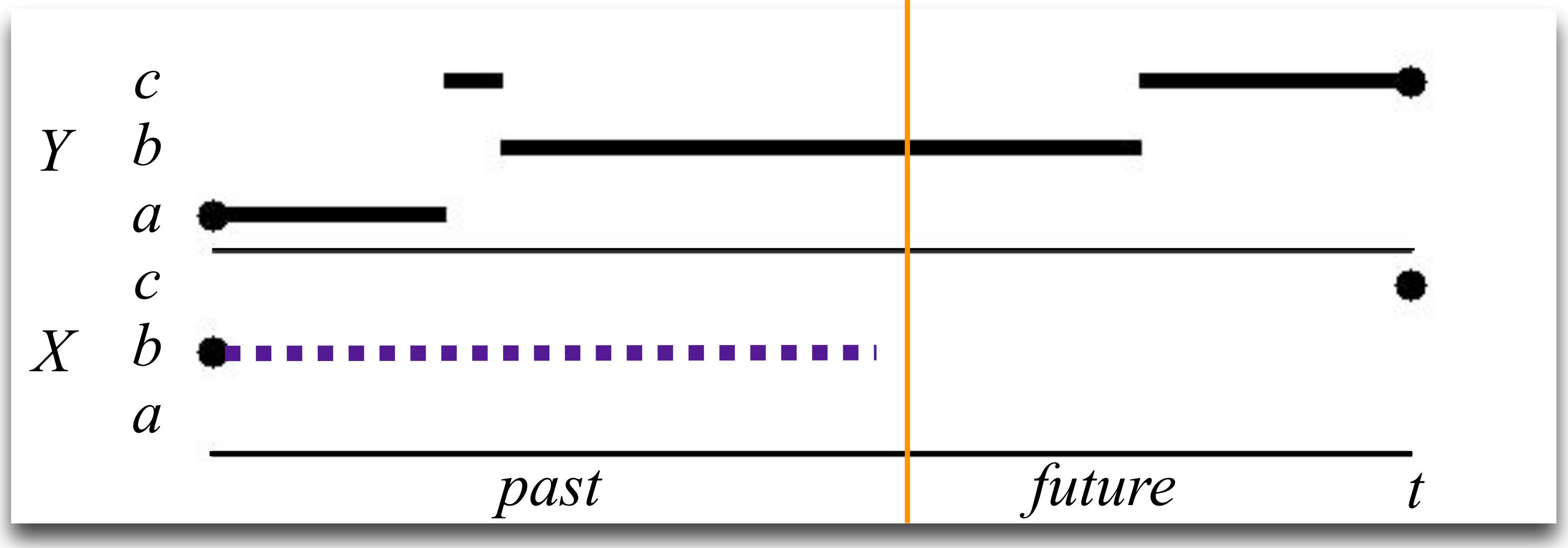
Distribution of Next Transition

Probability to remain:

$$\Pr(X^{(0,t]} = x^0 | x^0, x^T, y^{[0,T]}) = \frac{p^{\text{past}}(t)\, p^{\text{future}}_{x^0}(t)}{p^{\text{future}}_{x^0}(0)}$$

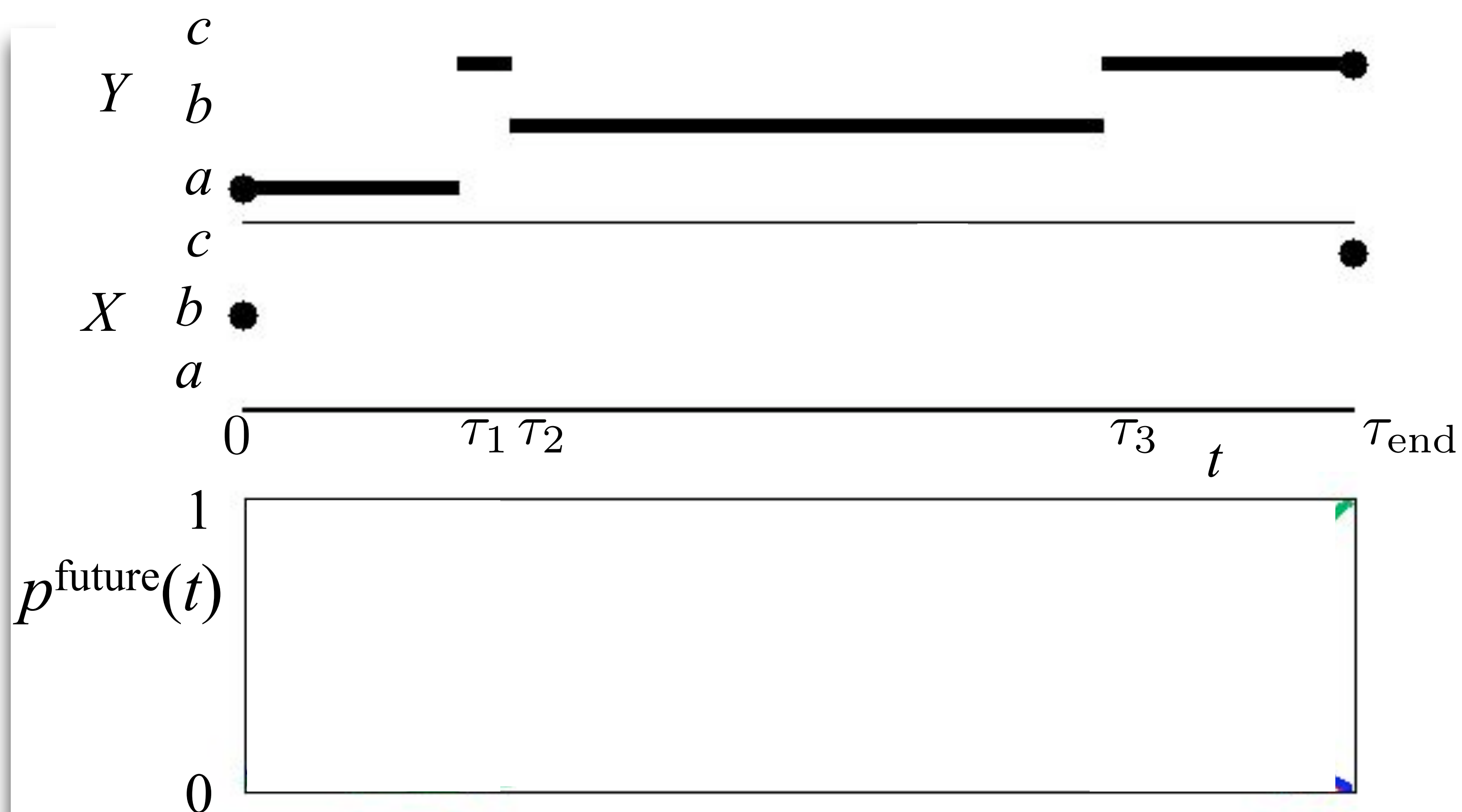
Normalization

$$\Pr(X^{(0,t]} = x^0, y^{(0,t]} | x^0, y^0) \qquad \Pr(x^T, y^{(t,T]} | X^t = x^0, y^t)$$



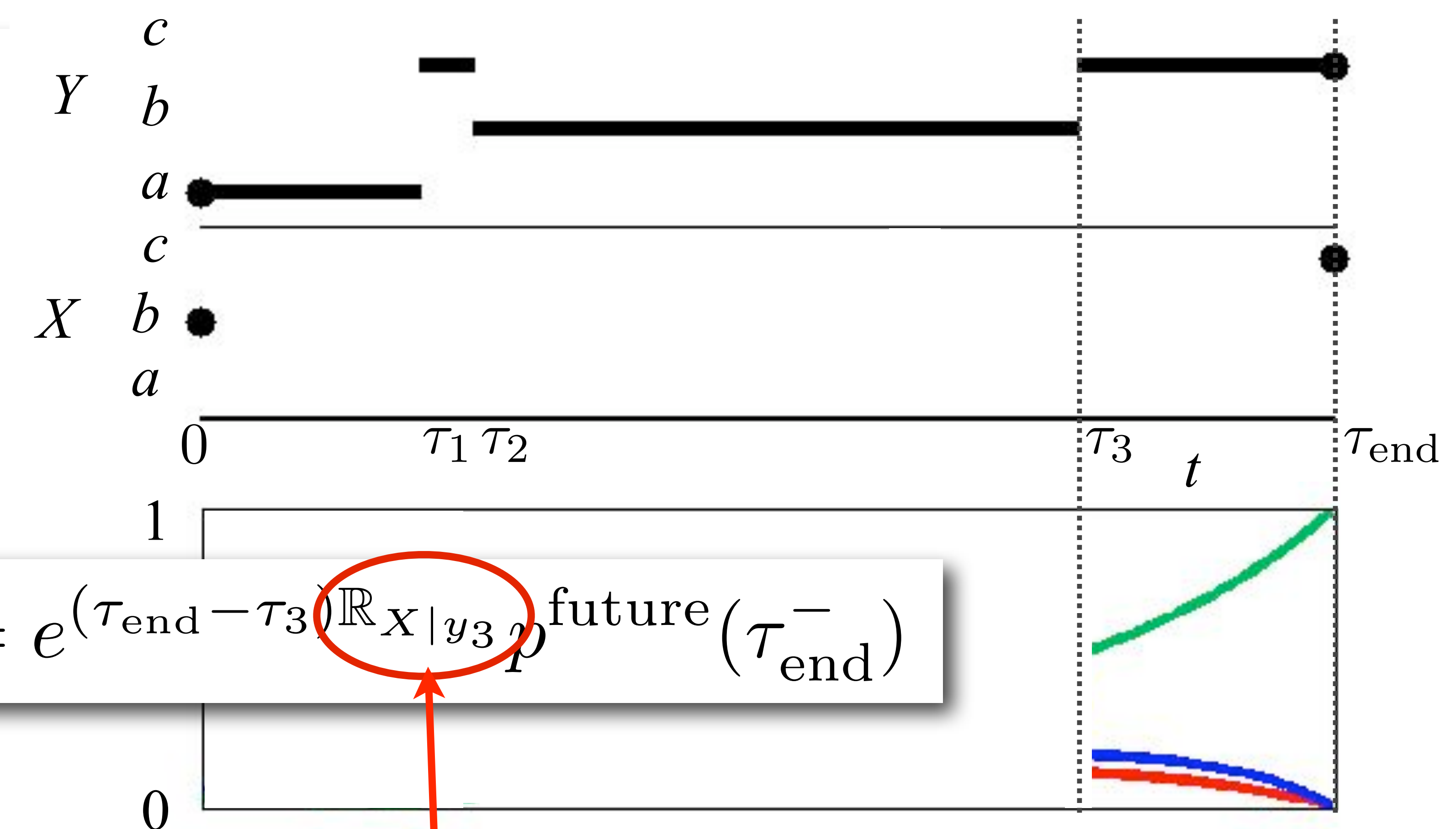
Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y



Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y

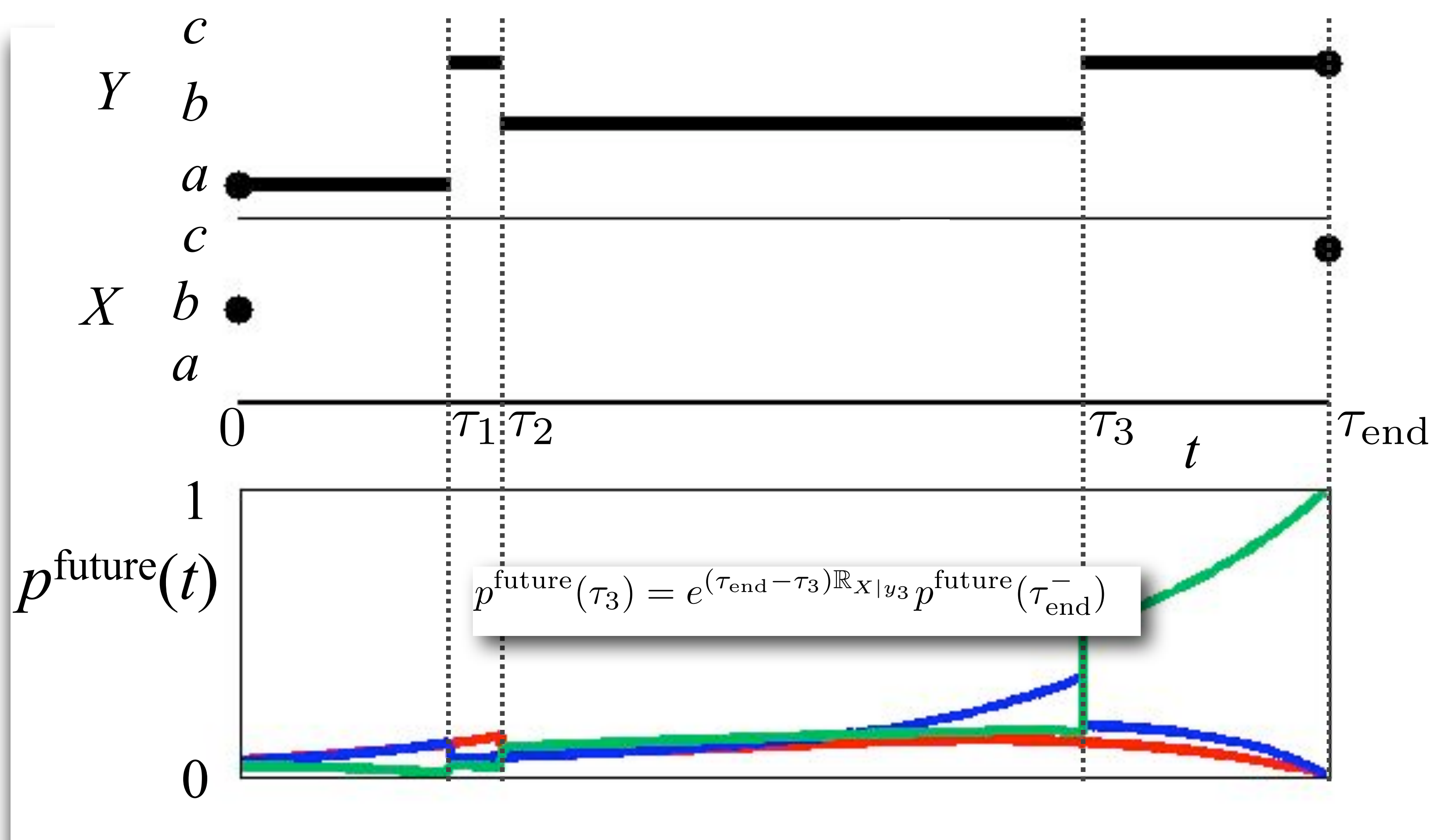


$$p^{\text{future}}(\tau_3) = e^{(\tau_{\text{end}} - \tau_3)\mathbb{R}_{X|y_3}} p^{\text{future}}(\tau_{\text{end}}^-)$$

Local unnormalized rate matrix

Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y

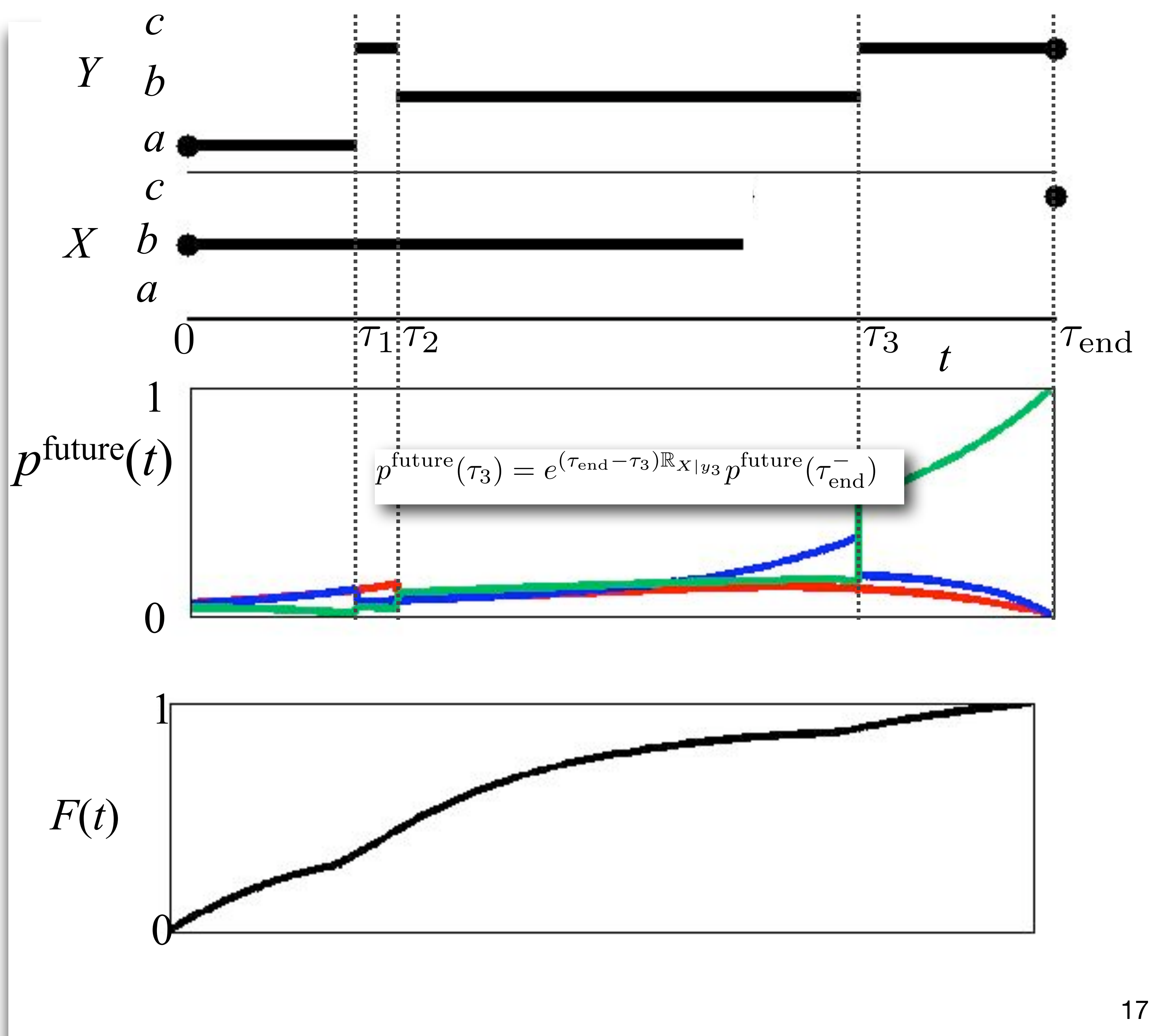


Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y

Sample next time

1. Draw γ from $U[0,T]$
2. Search k such that $F(t_k) < \gamma < F(t_k+1)$
3. Binary search within $[t_k, t_{k+1})$



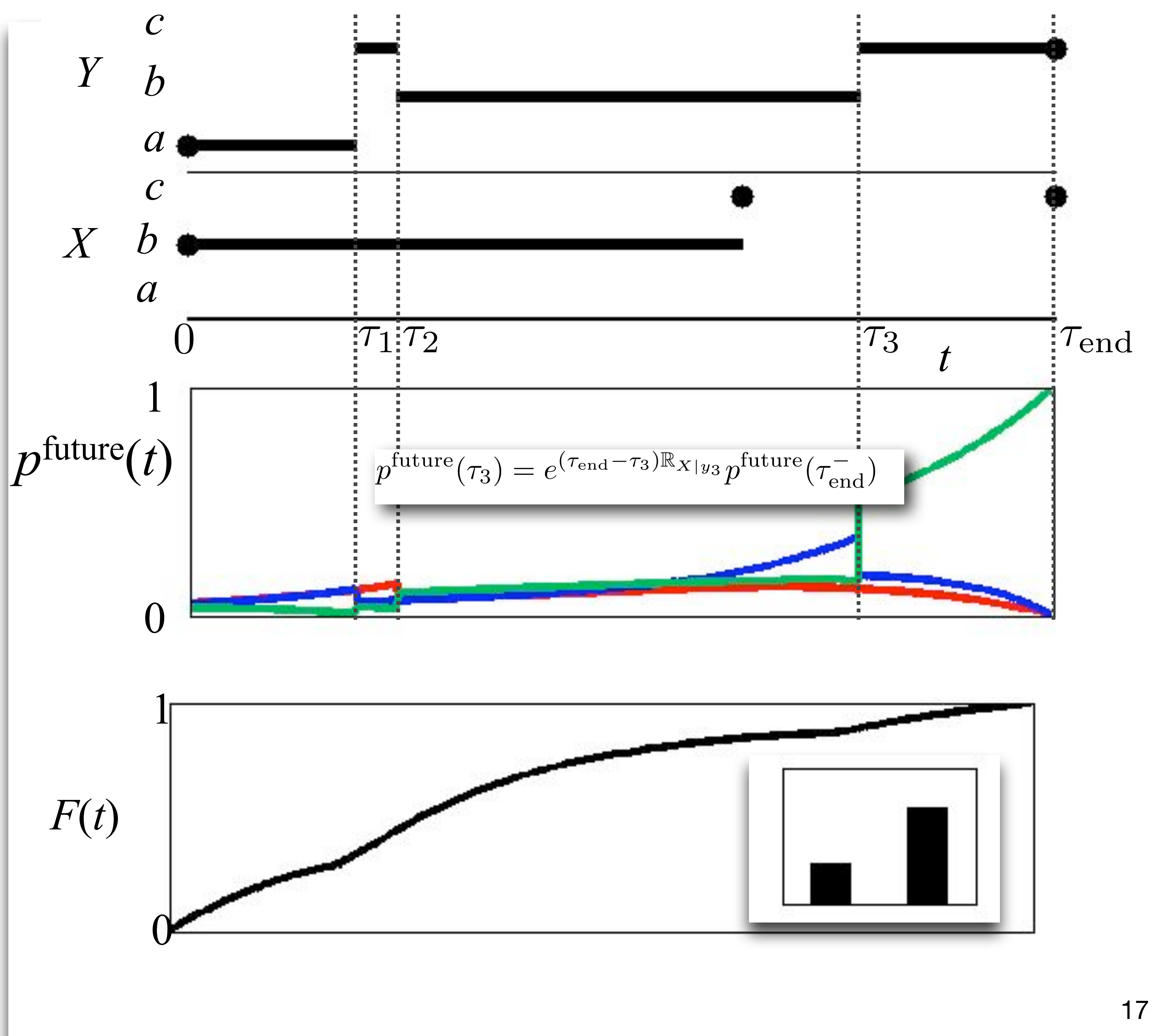
Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y

Sample next time

1. Draw γ from $U[0,T]$
2. Search k such that $F(t_k) < \gamma < F(t_k+1)$
3. Binary search within $[t_k, t_{k+1})$

Sample next state



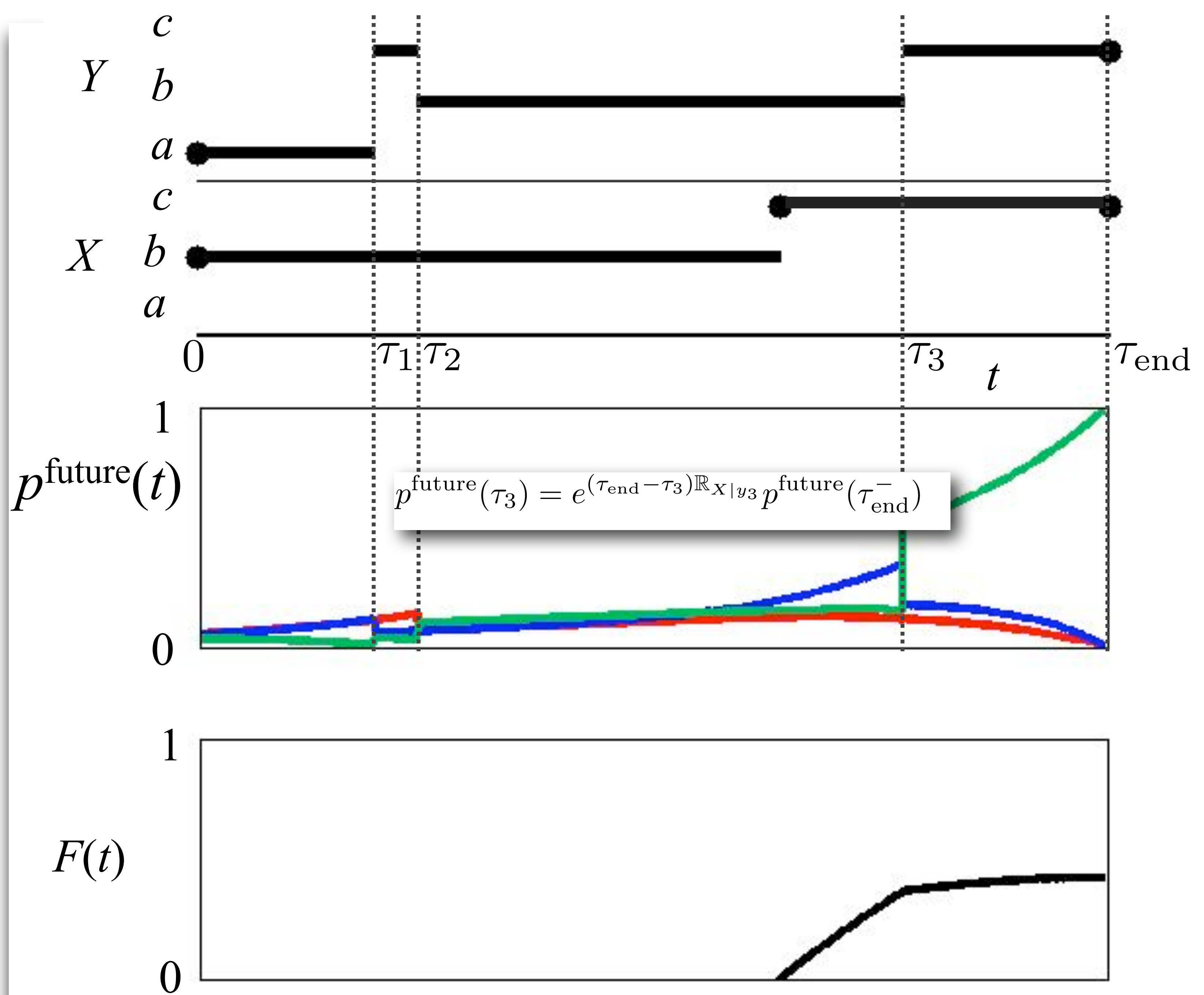
Sampling Procedure in Two Component Processes

Calculate $p^{\text{future}}(t)$ at transition points of Y

Sample next time

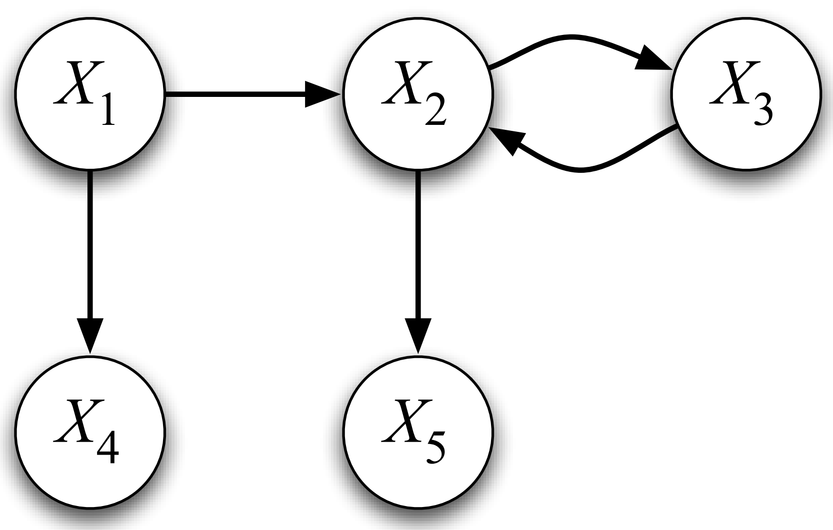
1. Draw γ from $U[0,T]$
2. Search k such that $F(t_k) < \gamma < F(t_k+1)$
3. Binary search within $[t_k, t_{k+1})$

Sample next state



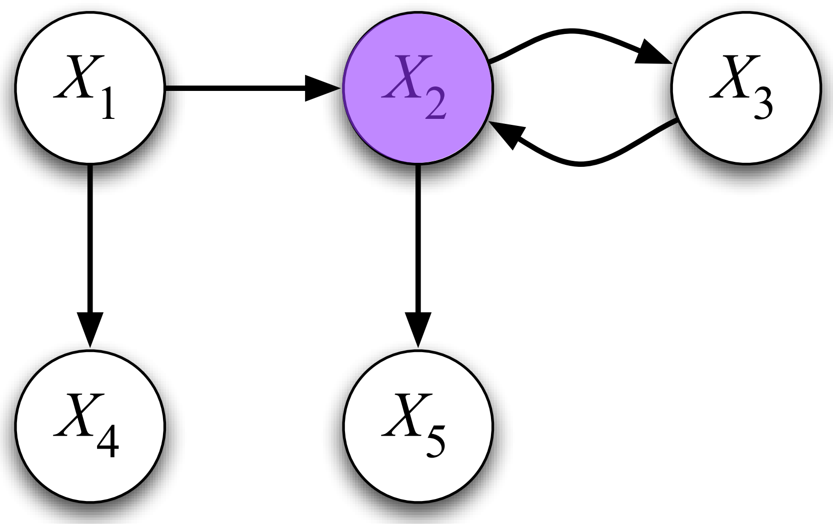
Locality

Given complete trajectories of the Markov blanket of Xi, X_i is independent of all other components [Nodelman et al 2002].



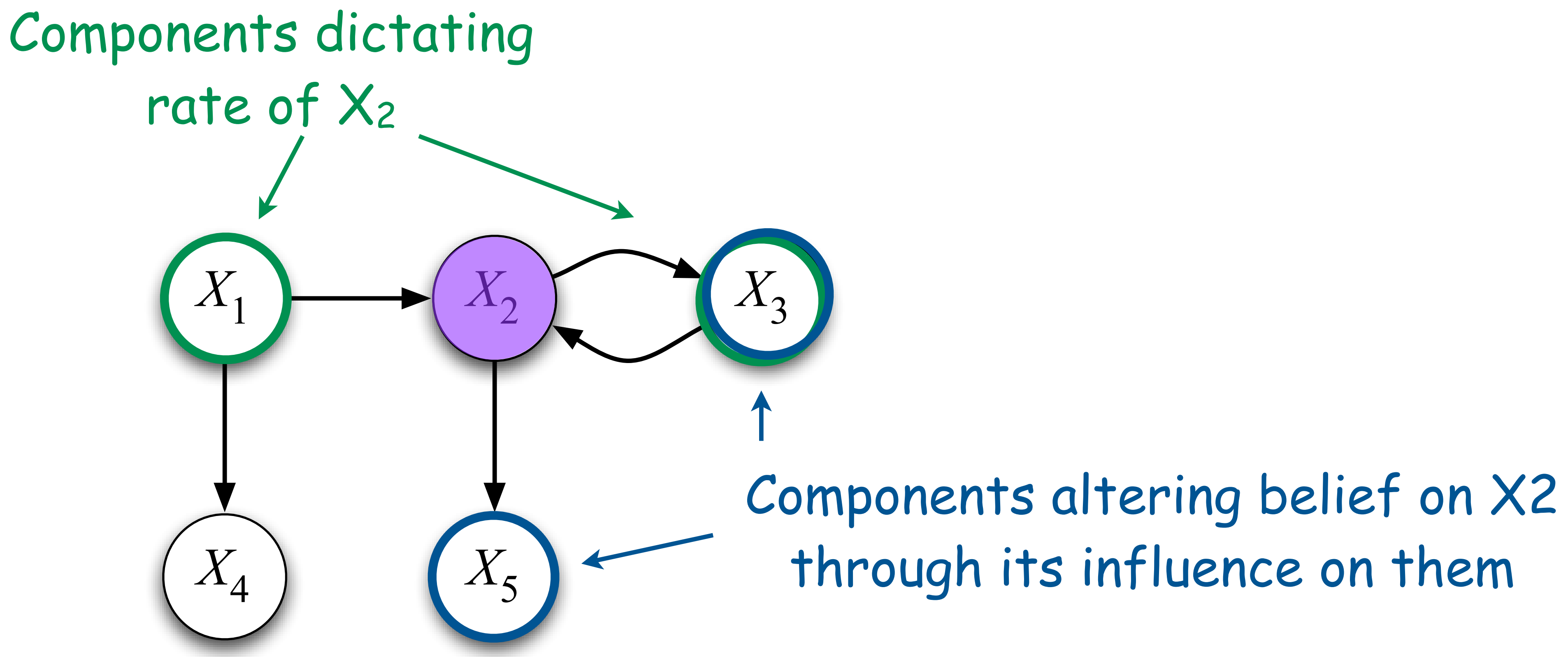
Locality

Given complete trajectories of the Markov blanket of Xi, X_i is independent of all other components [Nodelman et al 2002].



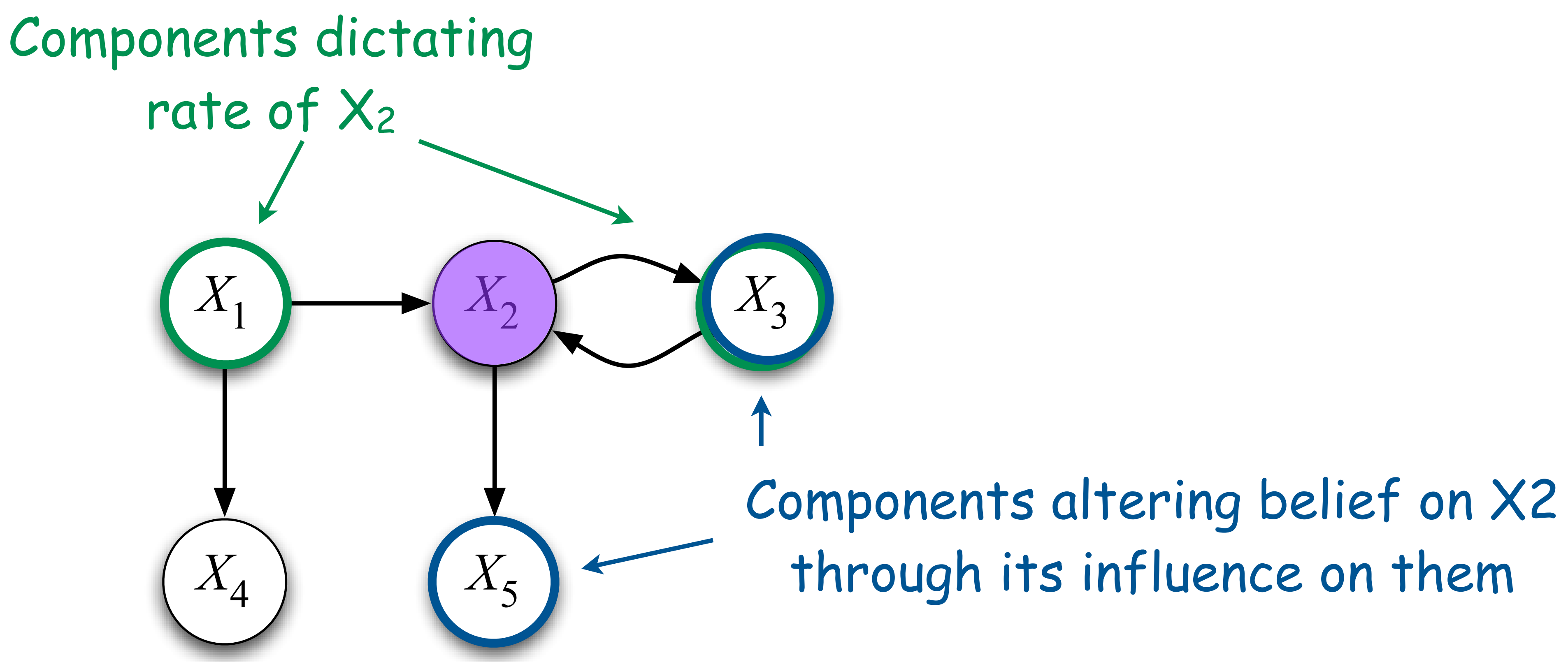
Locality

Given complete trajectories of the Markov blanket of Xi, X_i is independent of all other components [Nodelman et al 2002].



Locality

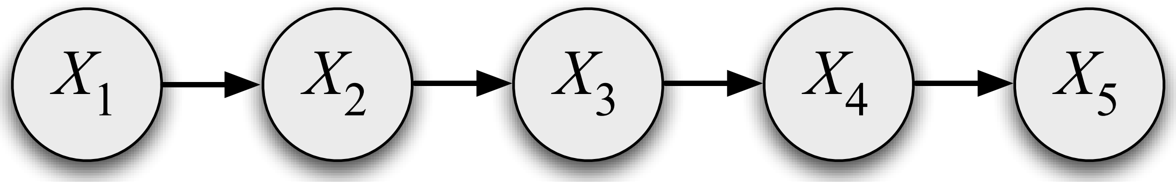
Given complete trajectories of the Markov blanket of Xi, X_i is independent of all other components [Nodelman et al 2002].



➡ **Complexity scales with the rate of X_i and its Markov Blanket**

Experimental Setup

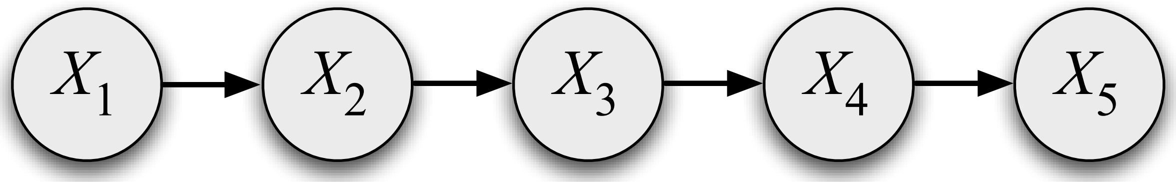
Model



- Each component with 5 states
- X_1 tends to cycle in two loops
  a→b→c→a and **a→d→e→a**
- X_i tries to follow X_{i-1}

Experimental Setup

Model



- Each component with 5 states
- X_1 tends to cycle in two loops **a→b→c→a** and **a→d→e→a**
- X_i tries to follow X_{i-1}

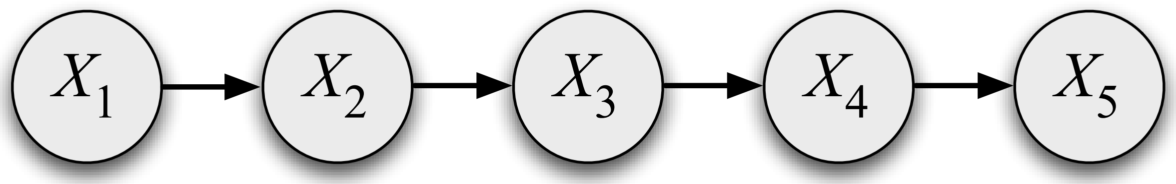
Evidence

$X^{(0)} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})$

$X^{(T)} = (\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{a}, \mathbf{b})$

Experimental Setup

Model



- Each component with 5 states
- X_1 tends to cycle in two loops
  $\mathbf{a \to b \to c \to a}$ and $\mathbf{a \to d \to e \to a}$
- X_i tries to follow X_{i-1}

Evidence

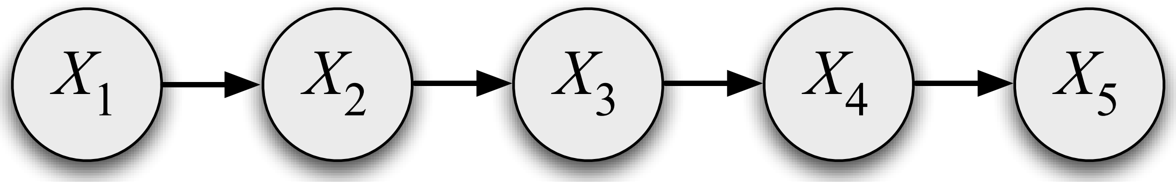
$X^{(0)} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})$

$X^{(T)} = (\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{a}, \mathbf{b})$

➡ **- Low probability evidence**
- Multi modal posterior

Experimental Setup

Model



- Each component with 5 states
- X_1 tends to cycle in two loops

  a→b→c→a and **a→d→e→a**
- X_i tries to follow X_{i-1}

Evidence

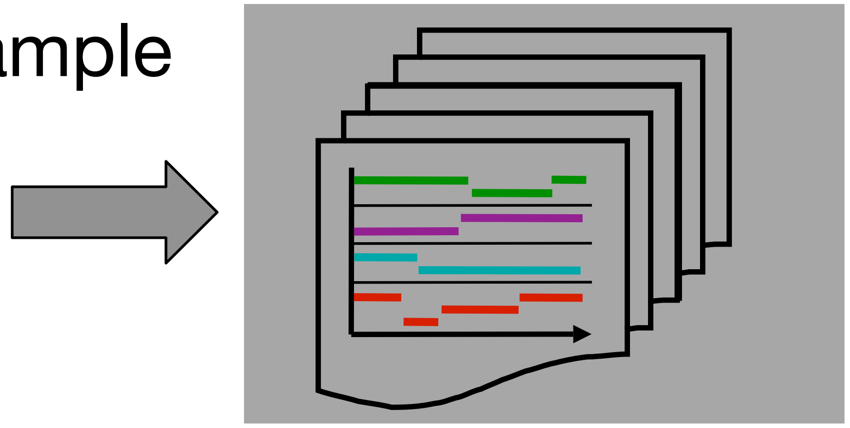
$$X^{(0)} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})$$

$$X^{(T)} = (\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{a}, \mathbf{b})$$

➡ **- Low probability evidence**

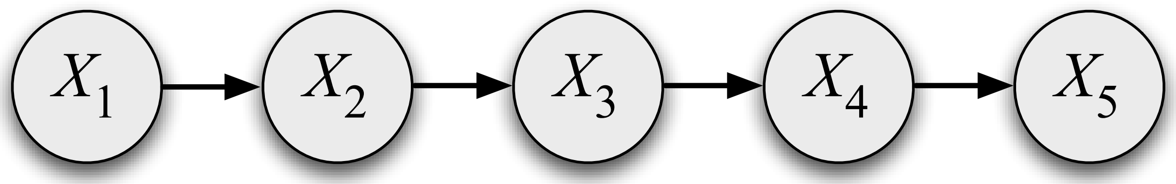
- Multi modal posterior

Sample



Experimental Setup

Model



- Each component with 5 states
- X_1 tends to cycle in two loops
  a→b→c→a and **a→d→e→a**
- X_i tries to follow X_{i-1}

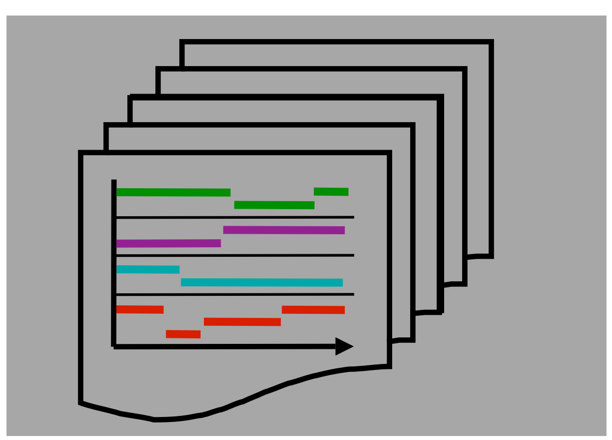
Evidence

$$X^{(0)} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})$$

$$X^{(T)} = (\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{a}, \mathbf{b})$$

➡ **- Low probability evidence**
- Multi modal posterior

Sample

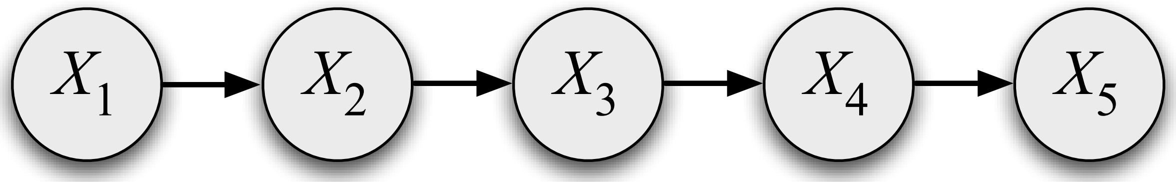


Compute Statistics

$\hat{N}[X_1 : a \to b] = 3.1$
$\hat{N}[X_2 : b \to c | X_1 = e] = .7$
$\ldots$
$\hat{T}[X_2 = a | X_1 = a] = .2$
$\hat{T}[X_2 = b | X_1 = a] = .8$
$\ldots$

Experimental Setup

Model



- Each component with 5 states
- X_1 tends to cycle in two loops
  a→b→c→a and **a→d→e→a**
- X_i tries to follow X_{i-1}

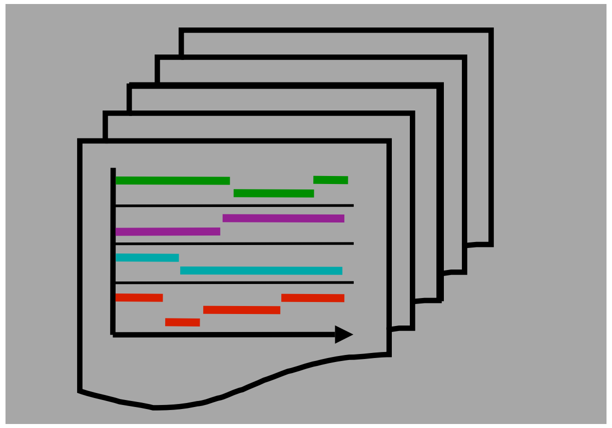
Evidence

$$X^{(0)} = (\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})$$

$$X^{(T)} = (\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{a}, \mathbf{b})$$

➡ **- Low probability evidence**
- Multi modal posterior

Sample →

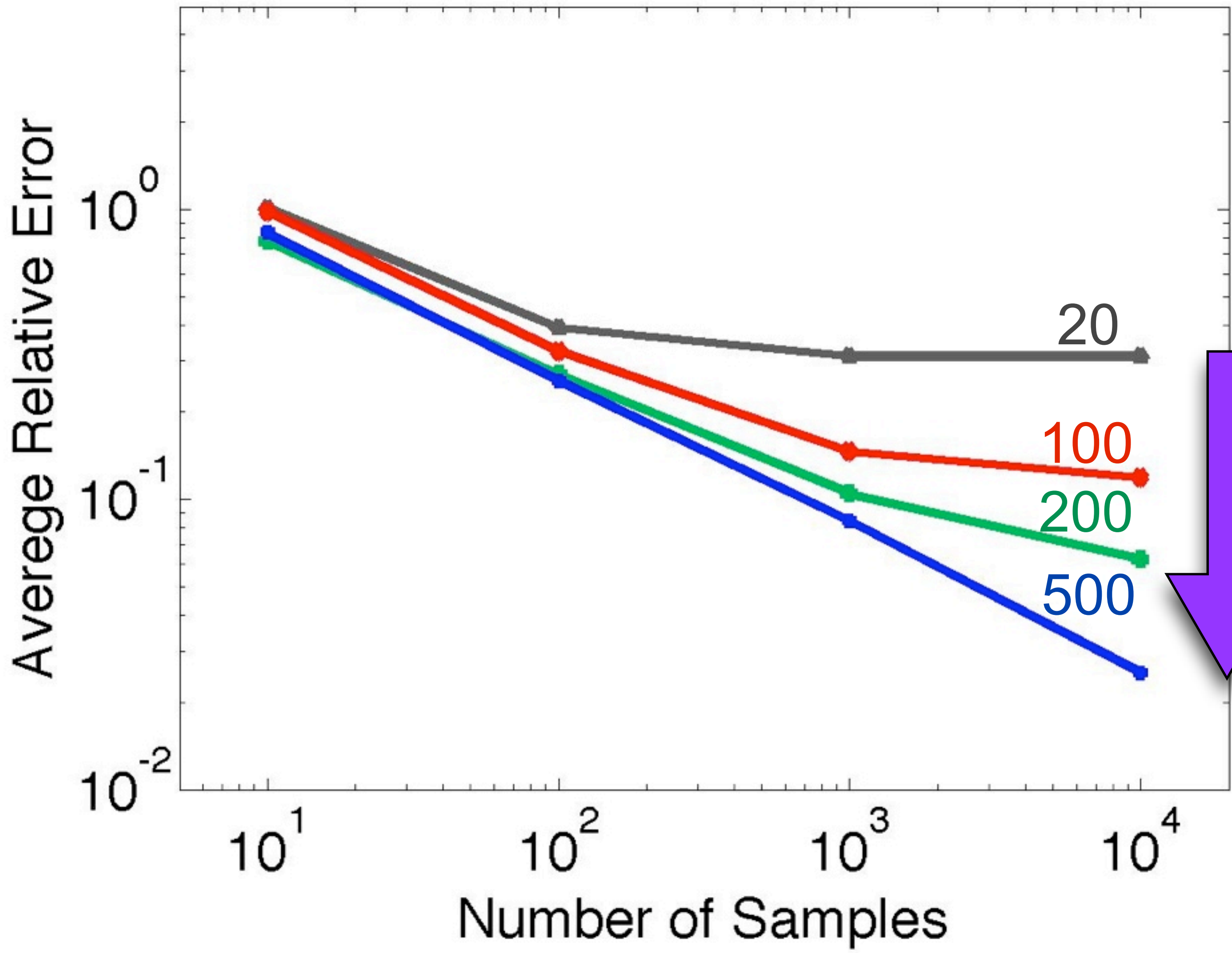


Compute Statistics →

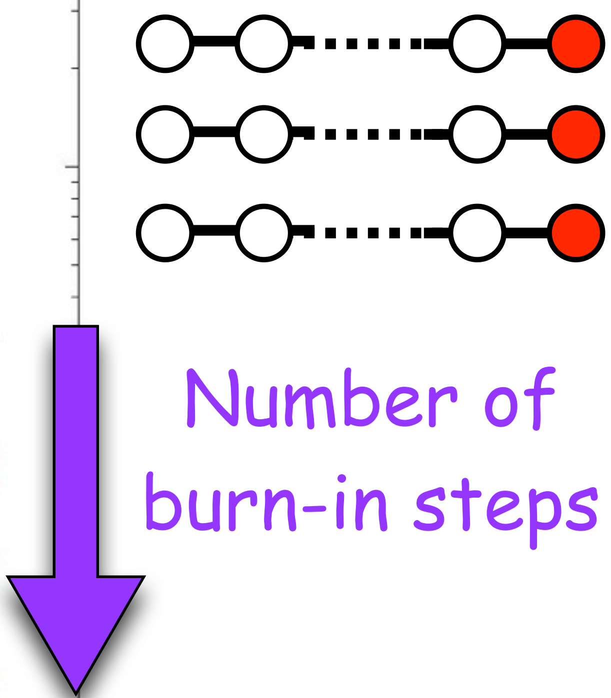
$$\hat{N}[X_1 : a \rightarrow b] = 3.1$$
$$\hat{N}[X_2 : b \rightarrow c | X_1 = e] = .7$$
$$\ldots$$
$$\hat{T}[X_2 = a | X_1 = a] = .2$$
$$\hat{T}[X_2 = b | X_1 = a] = .8$$
$$\ldots$$

→ Compare to Exact

Results

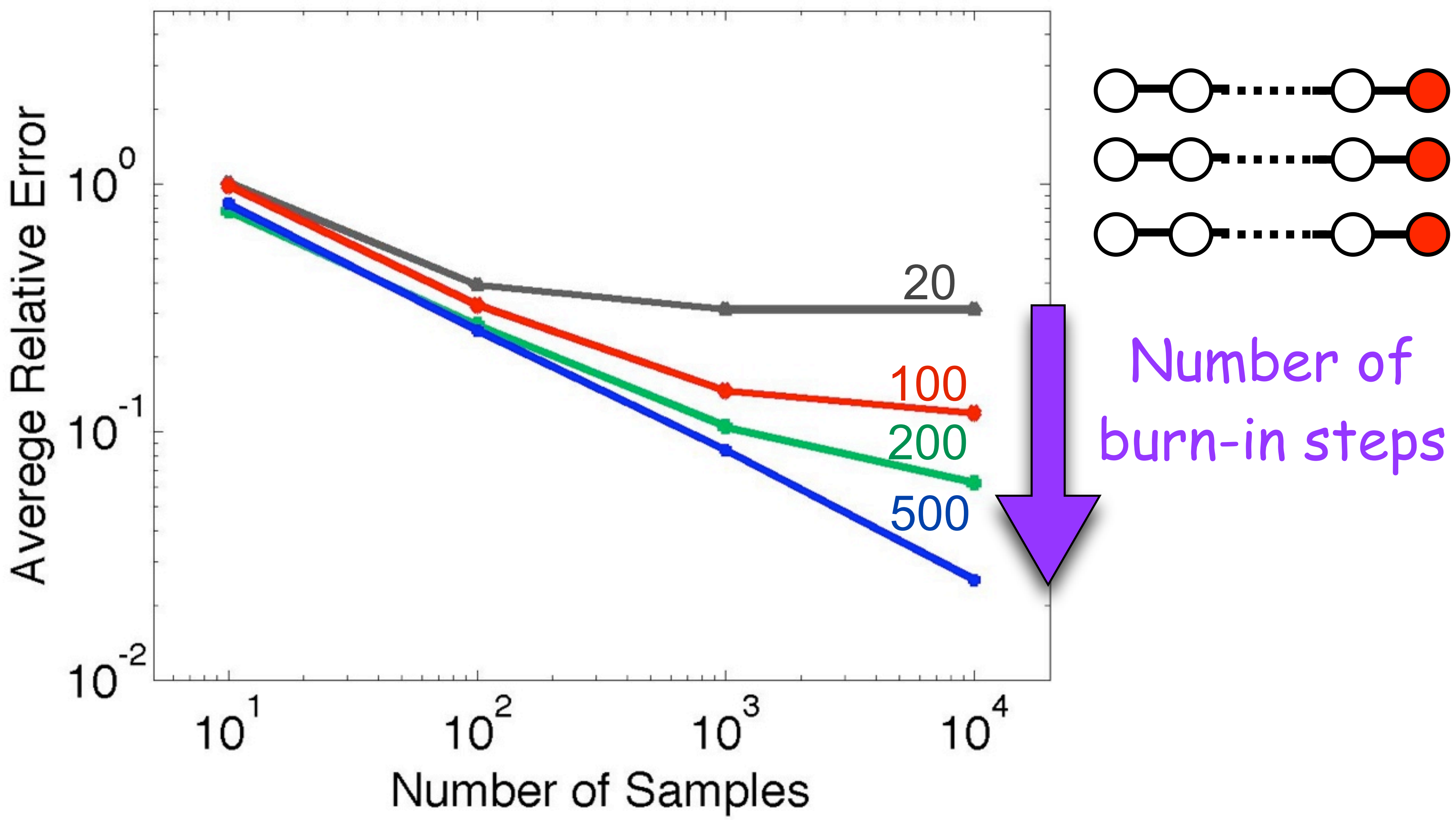


Averege Relative Error
10^0
10^{-1}
10^{-2}
10^1
10^2
10^3
10^4
Number of Samples
20
100
200
500



Number of
burn-in steps

Results



Unbiased samples after 500 burn-in steps

Summary

Continuous Time Gibbs Sampling

- Exact posterior for distinct components
- Asymptotically Unbiased
- Suitable for judging other inference methods
- Adapts to the natural time scale of the sampled process
- Exploits network structure to reduce computational cost

Further directions

- Convergence diagnostics
- Acceleration of convergence

Thank you