Moment-Based Variational Inference for Markov Jump Processes

Christian Wildner and Heinz Koeppl

Department of Electrical Engineering and Information Technology
Technische Universität Darmstadt, Germany

Introduction

Model Class: Markov jump process / continuous time Markov chain

- Applications in many domains (finance, social networks, healthcare, systems biology, etc.)
- Data-driven modelling requires latent state estimation

Introduction

Model Class: Markov jump process / continuous time Markov chain

- Applications in many domains (finance, social networks, healthcare, systems biology, etc.)
- Data-driven modelling requires latent state estimation

Problem: Hard/intractable for large state spaces

Introduction

Model Class: Markov jump process / continuous time Markov chain

- Applications in many domains (finance, social networks, healthcare, systems biology, etc.)
- Data-driven modelling requires latent state estimation

Problem: Hard/intractable for large state spaces

Proposed solution: new variational inference approach based on

- transition space partitioning
- gradient-based optimization

Markov Jump Processes

An MJP X is fully defined by

- an initial distribution p_0
- a transition function Q^X with

Markov Jump Processes

An MJP X is fully defined by

- an initial distribution p_0
- a transition function Q^X with

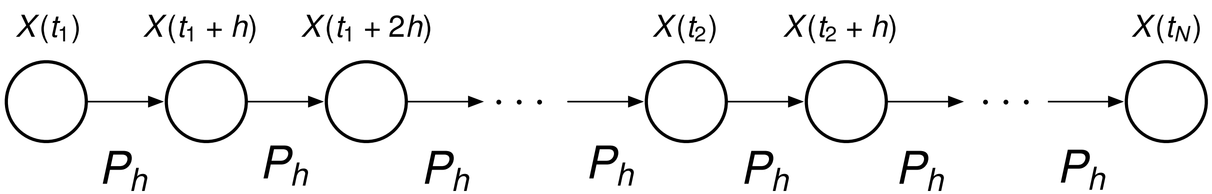
$$Pr(X(t+h) = y \mid X(t) = X) = \underbrace{\delta(x,y) + Q^X(x,y,t)h}_{=:\, P_h} + o(h)$$

Markov Jump Processes

An MJP X is fully defined by

- an initial distribution p_0
- a transition function Q^X with

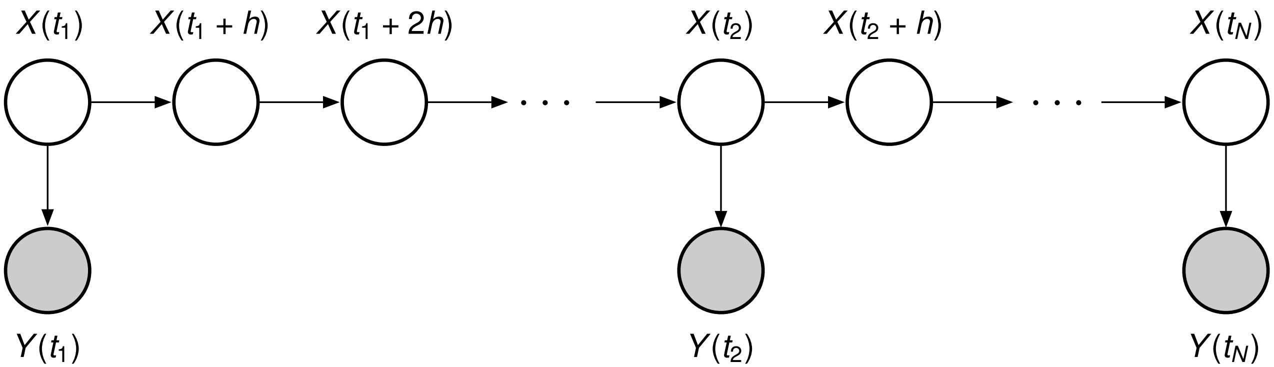
$$Pr(X(t+h) = y \mid X(t) = X) = \underbrace{\delta(x, y) + Q^X(x, y, t)h}_{=:\, P_h} + o(h)$$



discretized representation of hidden MJP

Exact Inference

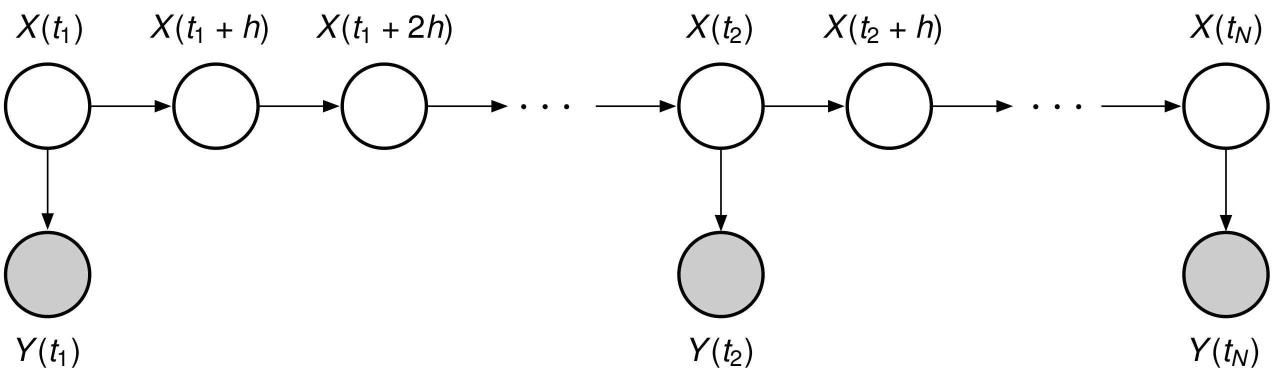
Goal: Compute posterior *path* distribution $P(X_{[0,T]} \mid Y_1, \ldots, Y_n)$



discretized representation of hidden MJP

Exact Inference

Goal: Compute posterior *path* distribution $P(X_{[0,T]} \mid Y_1, \dots, Y_n)$

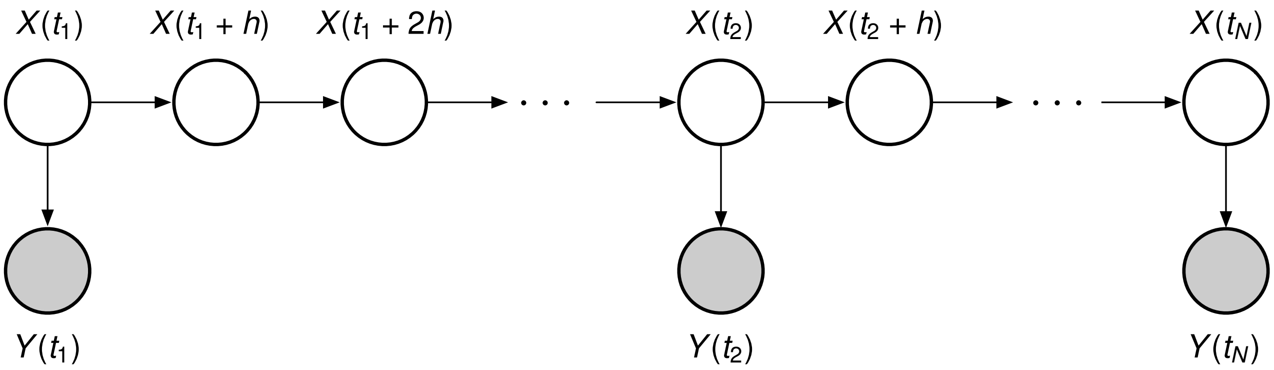


discretized representation of hidden MJP

Posterior paths are realized by smoothing process $\tilde{X}$ with modified transition function

Exact Inference

Goal: Compute posterior *path* distribution $P(X_{[0,T]} \mid Y_1, \ldots, Y_n)$



discretized representation of hidden MJP

hard to compute

Posterior paths are realized by smoothing process $\tilde{X}$ with modified transition function

$$\tilde{Q}(x, y, t) = \frac{\sigma(y, t)}{\sigma(x, t)} Q^X(x, y)$$

Variational Inference

TECHNISCHE UNIVERSITÄT DARMSTADT

Minimize path level KL divergence $D_{KL}[P^{Z} || P^{\tilde{X}}]$

exact smoothing process

Variational Inference

Minimize path level KL divergence $D_{KL}[P^Z || P^{\tilde{X}}]$

exact smoothing process

The usual decomposition applies

negative ELBO

$$D_{KL}[P^Z || P^{\tilde{X}}] = \underbrace{D_{KL}[P^Z || P^X]}_{\text{variational to prior}} - \sum_{k=1}^{N} \mathrm{E}[\log p(y_k \mid Z(t_k)] + \underbrace{\log p(y_1, \ldots, y_n)}_{\text{marginal likelihood (evidence)}}$$

variational to prior

marginal likelihood
(evidence)

Variational Inference

Minimize path level KL divergence $D_{KL}[P^Z || P^{\tilde{X}}]$ ← exact smoothing process

The usual decomposition applies

$$D_{KL}[P^Z || P^{\tilde{X}}] = \underbrace{\overbrace{D_{KL}[P^Z || P^X]}^{}}_{\text{variational to prior}} \overbrace{- \sum_{k=1}^{N} \mathrm{E}[\log p(y_k \mid Z(t_k))]}^{\text{negative ELBO}} + \underbrace{\log p(y_1, \dots, y_n)}_{\text{marginal likelihood (evidence)}}$$

Hard to construct suitable variational process class

Transition Space Partitioning

Smoothing process is in the class of controlled MJP with

$$Q^Z(x, y, t) = \underbrace{\lambda(x, y, t)}_{\text{time and state dependent control factor}} \underbrace{Q^X(x, y)}_{\text{prior transition function}}$$

Transition Space Partitioning

TECHNISCHE UNIVERSITÄT DARMSTADT

Smoothing process is in the class of controlled MJP with

$$Q^Z(x, y, t) = \underbrace{\lambda(x, y, t)}_{\text{time and state dependent control factor}} \underbrace{Q^X(x, y)}_{\text{prior transition function}}$$

Partition transitions into groups Π_i and set

$$Q^Z(x, y, t) = \lambda_i(t) Q^X(x, y)\,, \quad (x, y) \in \Pi_i$$

Transition Space Partitioning

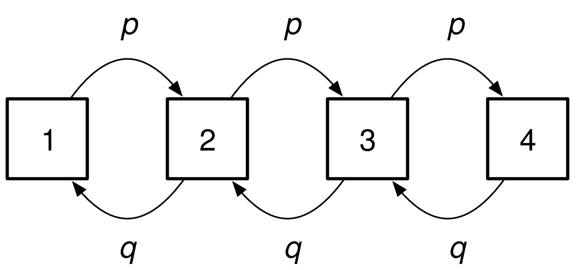
Smoothing process is in the class of controlled MJP with

$$Q^Z(x, y, t) = \underbrace{\lambda(x, y, t)}_{\substack{\text{time and state dependent} \\ \text{control factor}}} \underbrace{Q^X(x, y)}_{\substack{\text{prior transition} \\ \text{function}}}$$

Partition transitions into groups Π_i and set

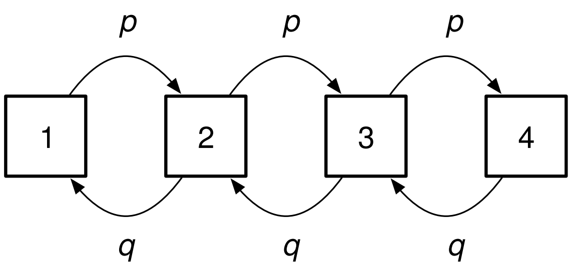
$$Q^Z(x, y, t) = \lambda_i(t) Q^X(x, y) \,, \quad (x, y) \in \Pi_i$$

Example: random walk



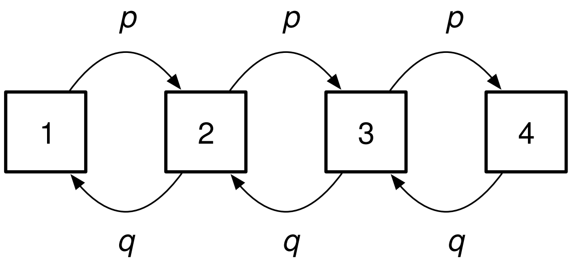
Partitioning Example

Example: random walk



Partitioning Example

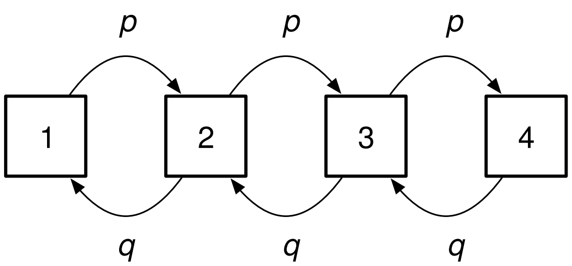
Example: random walk



$$Q^X = \begin{pmatrix} -p & p & 0 & 0 \\ q & -(p+q) & p & 0 \\ 0 & q & -(p+q) & p \\ 0 & 0 & q & -q \end{pmatrix}$$

Partitioning Example

Example: random walk



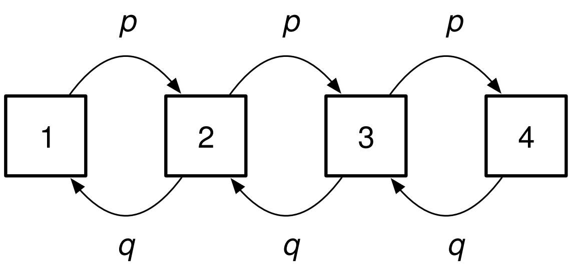
$$Q^X = \begin{pmatrix} -p & p & 0 & 0 \\ q & -(p+q) & p & 0 \\ 0 & q & -(p+q) & p \\ 0 & 0 & q & -q \end{pmatrix}$$

$\lambda_1(t)$: common scaling for rightward transitions

$\lambda_2(t)$: common scaling for leftward transitions

Partitioning Example

Example: random walk

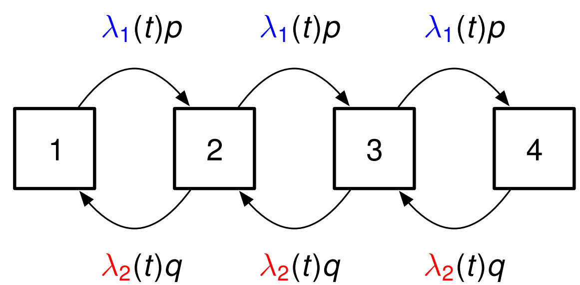


$$
Q^X = \begin{pmatrix} -p & p & 0 & 0 \\ q & -(p+q) & p & 0 \\ 0 & q & -(p+q) & p \\ 0 & 0 & q & -q \end{pmatrix}
$$

$\lambda_1(t)$: common scaling for rightward transitions

$\lambda_2(t)$: common scaling for leftward transitions

$\Rightarrow$



Complexity Reduction

$$D_{KL}[P^Z \,||\, P^X] = \int_0^T \sum_x p^Z(x,t) \sum_{y \neq x} \Big[Q^X(x,y) - Q^Z(x,y,t) - Q^Z(x,y,t) \log \left(\frac{Q^Z(x,y,t)}{Q^X(x,y)} \right) \Big] \, dt$$

Complexity Reduction

TECHNISCHE UNIVERSITÄT DARMSTADT

$$D_{KL}[P^Z \,||\, P^X] = \int_0^T \sum_x p^Z(x,t) \sum_{y \neq x} \Big[Q^X(x,y) - Q^Z(x,y,t) - Q^Z(x,y,t) \log\left(\frac{Q^Z(x,y,t)}{Q^X(x,y)}\right)\Big] \, dt$$

transition space partitioning into r classes

Complexity Reduction

$$D_{KL}[P^Z \,||\, P^X] = \int_0^T \sum_x p^Z(x,t) \sum_{y \neq x} \Big[Q^X(x,y) - Q^Z(x,y,t) - Q^Z(x,y,t) \log\left(\frac{Q^Z(x,y,t)}{Q^X(x,y)}\right)\Big]\, dt$$

transition space partitioning into r classes

$$D_{KL}[P^Z \,||\, P^X] = \sum_{i=1}^{r} \int_0^T \varphi_i(t)\,(1 - \lambda_i(t) + \lambda_i(t) \log \lambda_i(t))\, dt$$

Complexity Reduction

TECHNISCHE UNIVERSITÄT DARMSTADT

$$D_{KL}[P^Z \,||\, P^X] = \int_0^T \sum_x p^Z(x,t) \sum_{y \neq x} \Big[Q^X(x,y) - Q^Z(x,y,t) - Q^Z(x,y,t) \log\left(\frac{Q^Z(x,y,t)}{Q^X(x,y)}\right)\Big] \, dt$$

transition space partitioning into r classes

$$D_{KL}[P^Z \,||\, P^X] = \sum_{i=1}^{r} \int_0^T \varphi_i(t)\,(1 - \lambda_i(t) + \lambda_i(t) \log \lambda_i(t))\, dt$$

expected summary statistic

Control Problem

Using the Markov property, derive moment equations for φ

Control Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

Using the Markov property, derive moment equations for φ

Obtain non-linear, deterministic optimal control problem

$$\begin{aligned} \text{minimize} \quad & L[\lambda, \varphi] - F[\varphi] \\ \text{subject to} \quad & \frac{d}{dt}\varphi(t) = f(\lambda(t), \varphi(t)) \end{aligned}$$

Control Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

Using the Markov property, derive moment equations for φ

Obtain non-linear, deterministic optimal control problem

$$\begin{aligned} \text{minimize} \quad & L[\lambda, \varphi] - F[\varphi] \\ \text{subject to} \quad & \frac{d}{dt}\varphi(t) = f(\lambda(t), \varphi(t)) \end{aligned}$$

Solve via natural gradient descent in the controls $\lambda(t)$

Example Application

Scenario:

Stochastic gene expression

Studied by fluorescence microscopy

Goals:

Parameter estimation

Model comparison

Optimal experiment design