Priors over Recurrent Continuous Time Processes

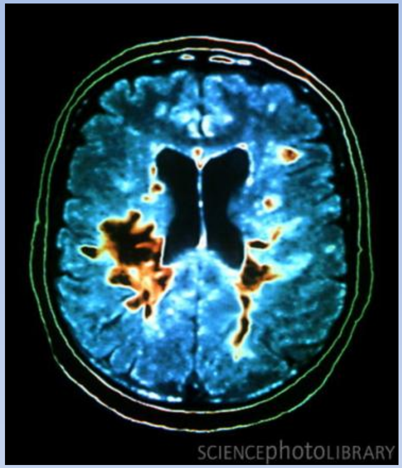
Speaker: Ardavan Saeedi

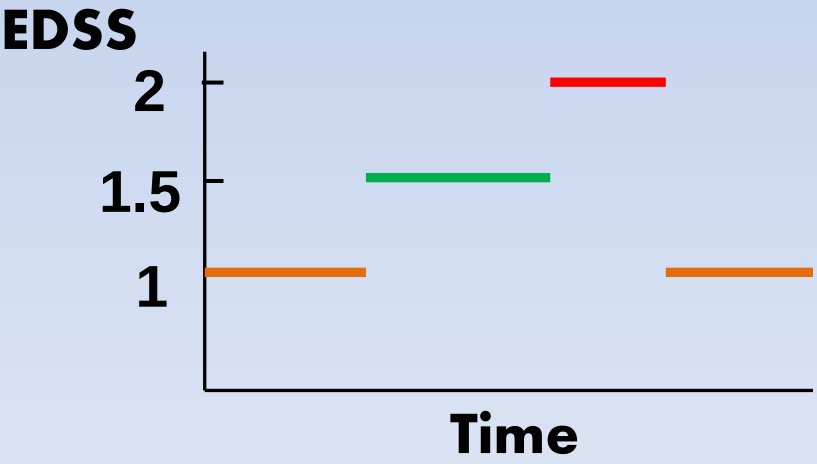
Joint work with: Alexandre Bouchard-Côté

Department of Statistics
University of British Columbia (UBC)

Example of data we are interested in

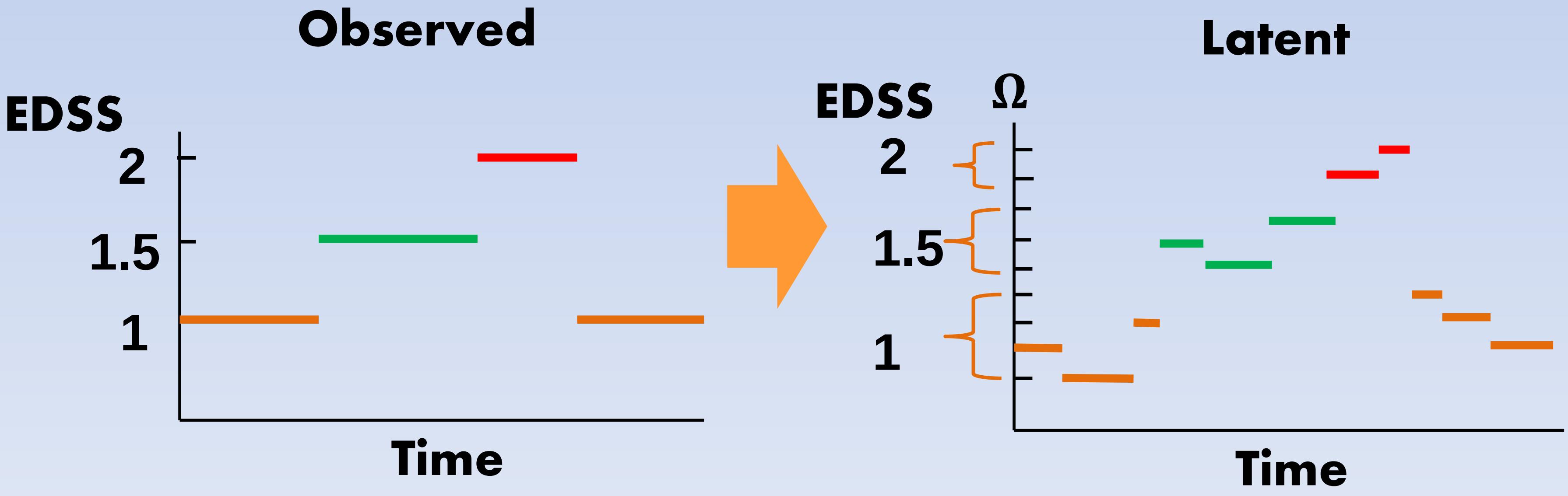
- Multiple Sclerosis (MS): a complex, poorly understood, neurological disease
- EDSS: a measure of disability in MS
- EDSS changes abruptly.
- Patients undergo cycles of relapses and remissions.





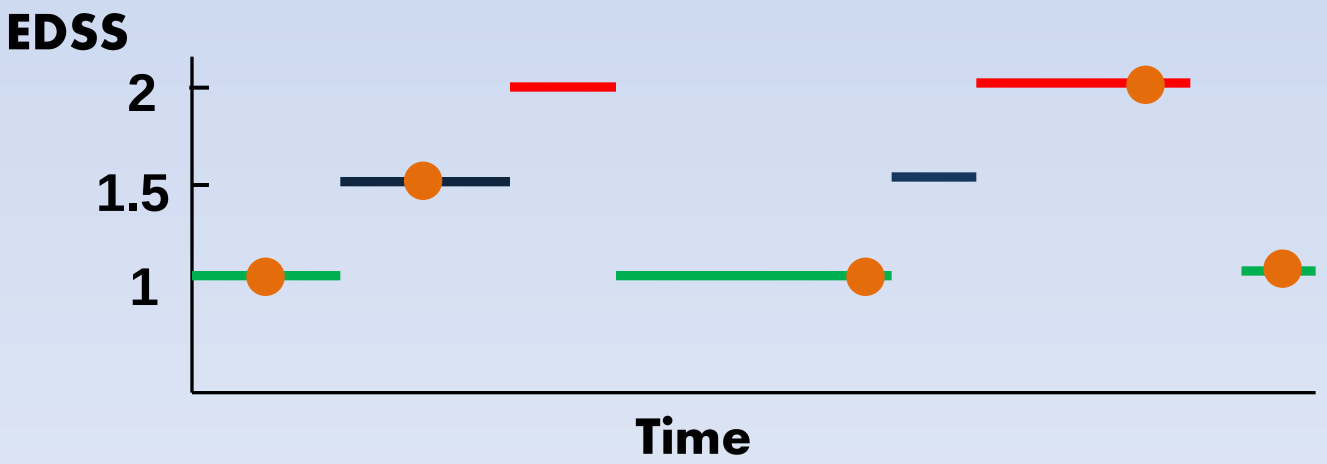
Example of data we are interested in

- EDSS sequence is not Markovian, but augmenting it with latent structure can make Markovian.
- Latent structure complexity is unknown.



Example of data we are interested in

- Continuous-time framework
- EDSS values only available at potentially unequal time intervals



General Properties of the Data

- Latent structure with unknown complexity
- Latent process is a jump process
- Recurrence
- Continuous-time framework
- Observations at potentially unequal time intervals

Our Model: a nonparametric Bayesian model

Outline

- Related Models & Background
- Gamma-Exponential Processes (GEP)
- Hierarchical GEP (HGEP)
- Inference
- Experiments
- Conclusion

Related Models

	Transient	**Recurrent**
Continuous-Time	• **Dependent Dirichlet Processes (DDP)** (MacEachern, 1999) • **Ornstein-Uhlenbeck Dirichlet Processes** (Griffin, 2008) • **Stick-breaking autoregressive processes** (Griffin & Steel, 2011)	?
Discrete-Time		• **Infinite HMM** (Beal et al. 2002) • **Sticky-HDP-HMM** (Fox et al. 2008)

Continuous-Time Markov Processes (CTMP)

- CTMP with M states ($\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$) is characterized by the off-diagonal entries of a rate matrix $Q_{M\times M}$:

$$
\begin{array}{c}
 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_M
\end{array}
\begin{array}{c}
\begin{array}{cccc} \omega_1 & \omega_2 & \dots & \omega_M \end{array} \\
\begin{pmatrix}
* & q_{\omega_1\omega_2} & \dots & q_{\omega_1\omega_M} \\
q_{\omega_2\omega_1} & * & \dots & \vdots \\
\vdots & & \ddots & \vdots \\
q_{\omega_M\omega_1} & \dots & \dots & *
\end{pmatrix}
\end{array}
$$

Continuous-Time Markov Processes (CTMP)

- Samples from CTMPs:

  a list of pairs of

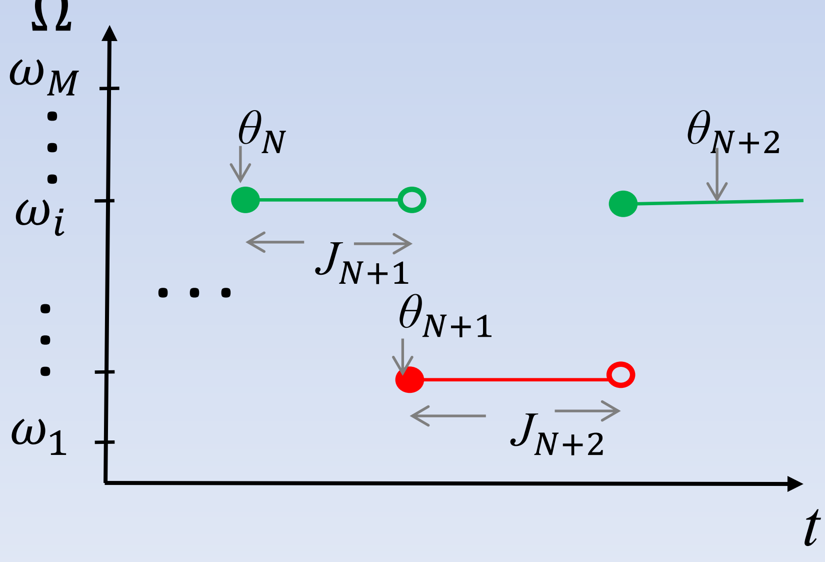
  - **latent states**: θ_n (from state space Ω)
  - **waiting times:** J_n

  All the latent information for N transitions: $X = (\theta_n, J_n)_{n=1}^{N}$
- Sampling from CTMPs:

  Doob-Gillespie algorithm given the rate matrix Q and $\theta_N = \omega_i$:

$$J_{N+1}|\theta_N = \omega_i \sim Exp\left(\sum_j q_{\omega_i \omega_j}\right)$$

$$p(\theta_{N+1} = \omega_j | \theta_N = \omega_i) \propto q_{\omega_i \omega_j}$$



Outline

- Related Models & Background ✓
- Gamma-Exponential Processes (GEP)
- Hierarchical GEP (HGEP)
- Inference
- Experiments
- Conclusion

Infinite Rate Matrix

- We want models with an adaptive complexity: we allow for any number of states, using **infinite dimensional rate matrix**.

$$\begin{array}{c} \\ \omega_1 \\ \omega_2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccc} \omega_1 & \omega_2 & \cdots \end{array} \\ \begin{pmatrix} q_{\omega_1\omega_1} & q_{\omega_1\omega_2} & \cdots \\ q_{\omega_2\omega_1} & q_{\omega_2\omega_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

How to develop priors on such infinite rate matrices ?

Priors on Infinite Rate Matrices

Big picture:

- View off-diagonal elements for the ith row of Q as a positive measure μ_i.

- Use (Moran) Gamma Process to generate each row of Q.

(Moran) Gamma Process (MGP)

- To get a conjugate prior over Q we base our priors on MGPs, a measure valued process with gamma marginals.
- Parameters of an MGP:
  1. Concentration parameter: $\alpha_0 > 0$
  2. Base probability distribution: $P_0 : \mathcal{F}_\Omega \to [0,1]$
  3. Rate parameter: $\beta_0 > 0$

(1–2) $H_0 = \alpha_0 P_0$ Base measure

- Some notation:

Normalization constant of μ: $||\boldsymbol{\mu}||$ Normalized measure: $\bar{\mu} = \frac{\mu}{||\mu||}$

Similar to Dirichlet Process (DP), but with one extra degree of freedom

Gamma-Exponential Process: Definition

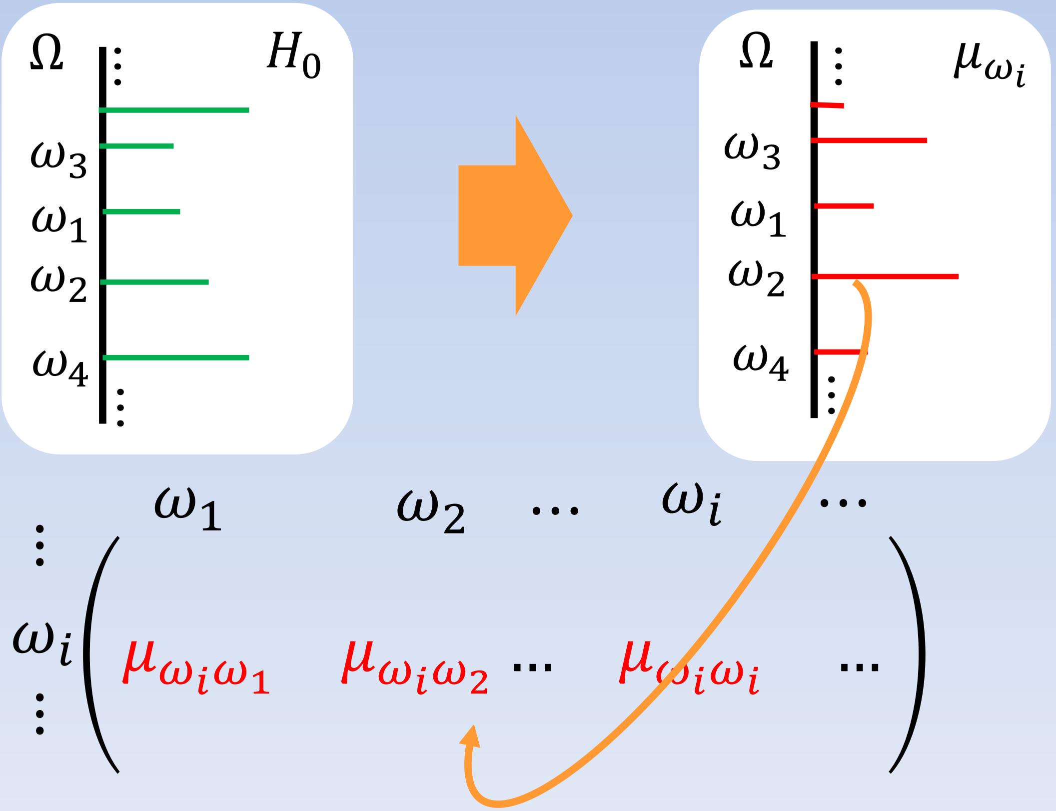
In GEPs:

- The rows of a rate matrix Q are obtained by *i.i.d.* samples from an MGP.
- Events are then generated from Q with the Doob-Gillespie algorithm.

Gamma-Exponential Process: Generative Process

Suppose H_0 is the base measure with countable support Ω:

1. For each row μ_ω of the rate matrix ($\omega \in \Omega$) sample *i.i.d.* from $\mathrm{MGP}(H_0, \beta_0)$:



Gamma-Exponential Process: Generative Process

2. Given the current state θ_N and the rates $\{\mu_\omega\}_{\omega\in\Omega}$:

$$\begin{array}{c} \vdots \\ \theta_N \\ \vdots \end{array} \overset{\begin{array}{cccc} \omega_1 & \omega_2 \;\cdots & \omega_i & \cdots \end{array}}{\begin{pmatrix} \\ \mu_{\theta_N\omega_1} & \mu_{\theta_N\omega_2} \cdots & \mu_{\theta_N\omega_i} & \cdots \\ \\ \end{pmatrix}} \xrightarrow{\;\;\Sigma\;\;} ||\mu_{\theta_N}||$$

sample waiting time until the next transition:

$$J_{N+1}|X,\{\mu_\omega\}_{\omega\in\Omega} \sim Exp(||\mu_{\theta_N}||)$$

Gamma-Exponential Process: Generative Process

3. Sample the next state θ_{N+1} from an infinite multinomial distribution:

$$\begin{array}{c} \vdots \\ \theta_N \\ \vdots \end{array} \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_i & \cdots \\ \mu_{\theta_N \omega_1} & \mu_{\theta_N \omega_2} & \cdots & \mu_{\theta_N \omega_i} & \cdots \\ & & & & \end{pmatrix} \xrightarrow{/\ ||\mu_{\theta_N}||} \bar{\mu}_{\theta_N}$$

$$\theta_{N+1} | X, \{\mu_\omega\}_{\omega \in \Omega} \sim \bar{\mu}_{\theta_N}$$

Gamma-Exponential Process: Properties

- Conjugacy: posterior of each row $\mu_\theta|X$ is also MGP with updated parameters.

We assume all the events are observed for now.

- Sufficient statistics for parameters of $\mu_\theta|X$ are:
  1. Number of transitions from state θ: F_θ
  2. Total waiting time at state θ: T_θ

GEP is a conjugate family, $\mu_\theta|X \sim MGP(\boldsymbol{\mu'_\theta}, \boldsymbol{\beta'_\theta})$ where $\boldsymbol{\mu'_\theta = H_0 + F_\theta}$ and $\boldsymbol{\beta'_\theta = \beta_0 + T_\theta}$

Gamma-Exponential Process: Properties

- Predictive distribution: GEP has a closed form expression for its predictive distribution $(\theta_{N+1}, J_{N+1})|X$.

The predictive distribution of the GEP is given by:

$$(\theta_{N+1}, J_{N+1})|X \sim \bar{\mu}'_{\theta_N} \times TP(\|\mu'_{\theta_N}\|, \boldsymbol{\beta}'_{\boldsymbol{\theta}_N})$$

Translated Pareto Distribution

$TP(\alpha, \beta)$ is the distribution with density:

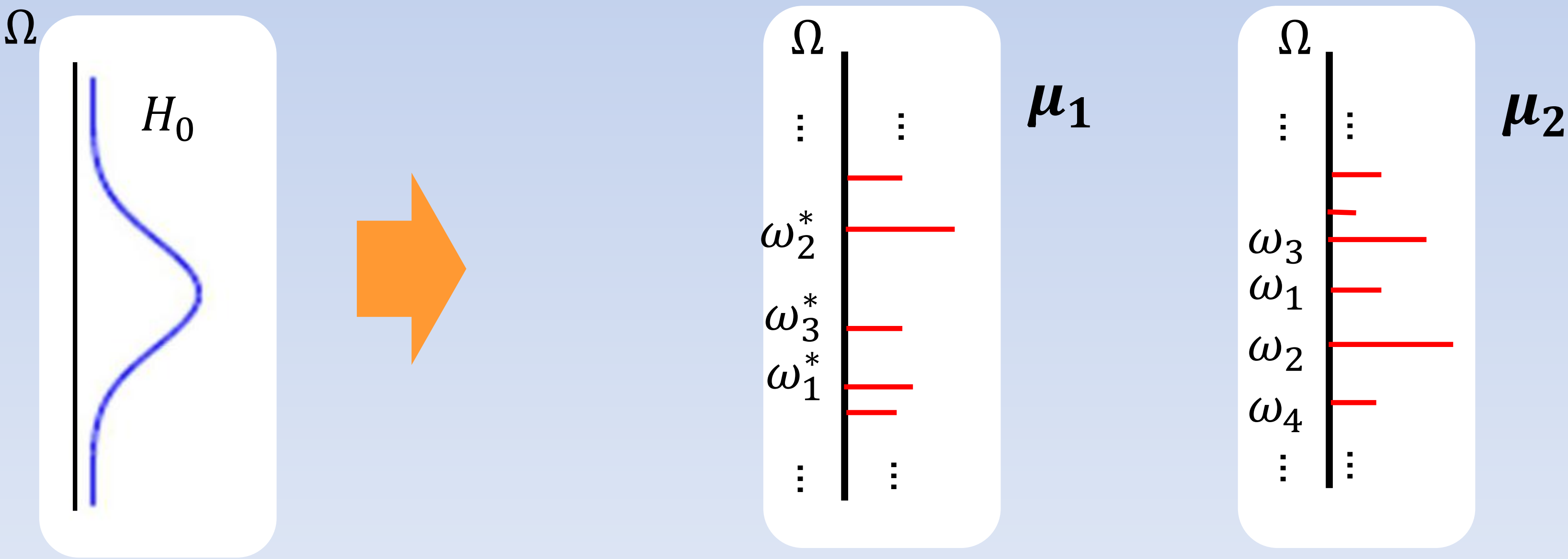
$$f(t) = \frac{\alpha\beta^{\alpha}}{(t+\beta)^{\alpha+1}}$$

Outline

- Related Models & Background ✓
- Gamma-Exponential Processes (GEP) ✓
- Hierarchical GEP
- Inference
- Experiments
- Conclusion

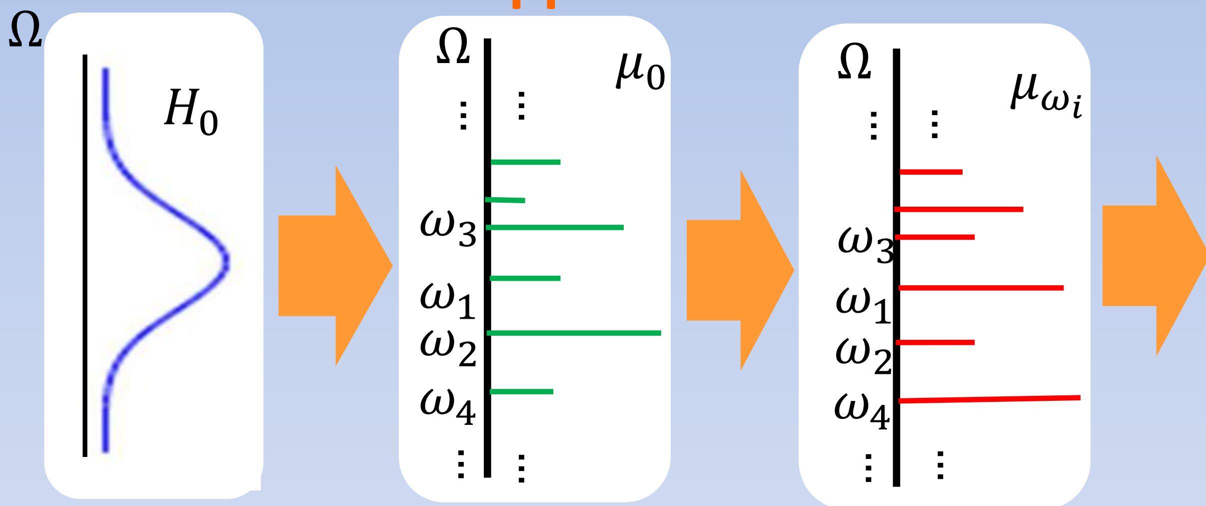
Hierarchical GEP (HGEP)

- Why do we need a hierarchical model?
  - To have the rows share information on what states are frequently visited.
  - To support Ω with uncountable support: two samples from an MGP will have disjoint supports with probability one.



Hierarchical GEP (HGEP)

- Solution: using a **random shared** base measure with **countable support** for the rows.



$$J_{N+1}|X,\{\mu_\omega\}_{\omega\in\Omega} \sim Exp(\|\mu_{\theta_N}\|)$$

$$\theta_{N+1}|X,\{\mu_\omega\}_{\omega\in\Omega} \sim \overline{\mu}_{\theta_N}$$

$\mu_0 \sim MGP(H_0, \gamma_0)$ $\mu_\omega \sim MGP(\mu_0, \beta_0)$ $i.i.d.$

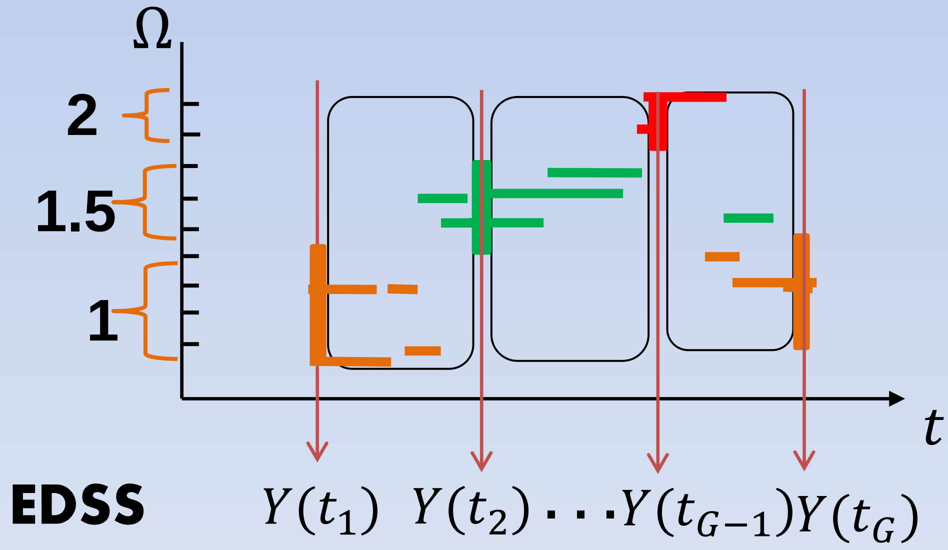
HGEP has a closed form predictive distribution similar to that of GEP.

Outline

- Related Models & Background ✓
- Gamma-Exponential Processes (GEP) ✓
- Hierarchical GEP ✓
- Inference
- Experiments
- Conclusion

Inference on Partially Observed Sequences

- In most applications the states are not directly observed, instead we only have a set of measurements for a finite set of times $\boldsymbol{y} = \left(\mathrm{y}(\mathrm{t}_1), \mathrm{y}(\mathrm{t}_2), \dots, \mathrm{y}(\mathrm{t}_G)\right)$.



Inference on Partially Observed Sequences

- At a high-level, our inference algorithm works by resampling the hidden events X by:
  - Using a Sequential Monte Carlo (SMC) algorithm to construct a proposal over sequences of hidden events.
  - Using a Particle MCMC (PMCMC) (Andrieu et al. 2010) method, to compute an acceptance ratio that makes this proposal a valid MCMC move.

Outline

- Related Models & Background ✓
- Gamma-Exponential Processes (GEP) ✓
- Hierarchical GEP ✓
- Inference ✓
- Experiments
- Conclusion

Experiments

- Three evaluation datasets obtained by holding out each observed datapoint with a 10% probability.
- Observations at these held-out times were reconstructed, and the mean error was measured.
- For HGEP the reconstruction was done using Bayes estimator.
- The results were compared to the standard maximum likelihood rate matrix estimator learned by EM (Hobolth & Jensen 2005).

Sample Dataset

Seq. 1

Time points	0	1	1.5	2.3	4.1
Observations	A	■	T	C	C

Seq. 2

Time points	1.1	2	3.1	4	5.2
Observations	T	C	T	■	C

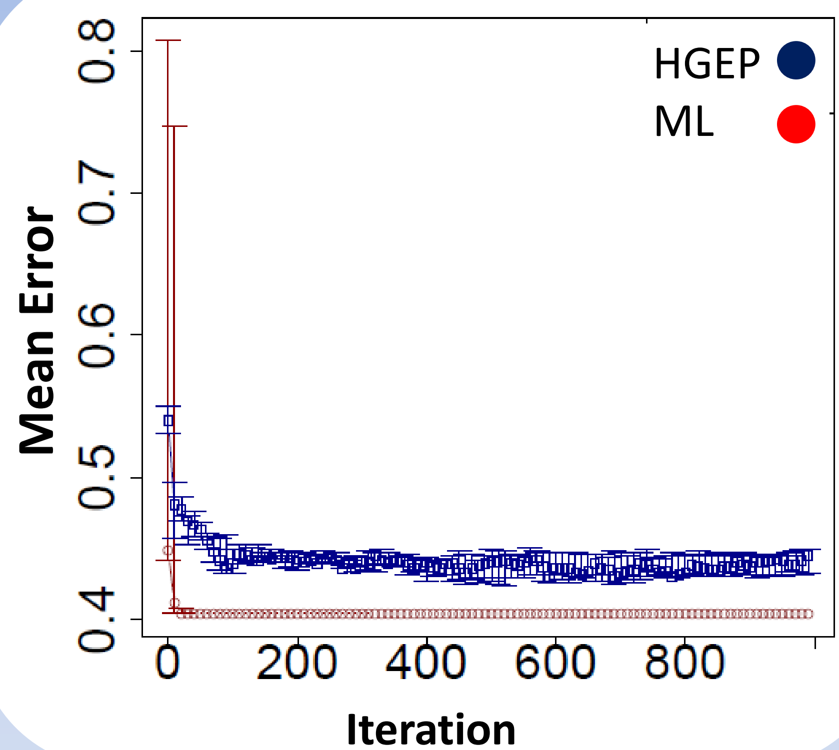
⋮

Seq. M

Time points	0	2	2.5	4.3	4.7
Observations	G	A	T	C	■

Experiments

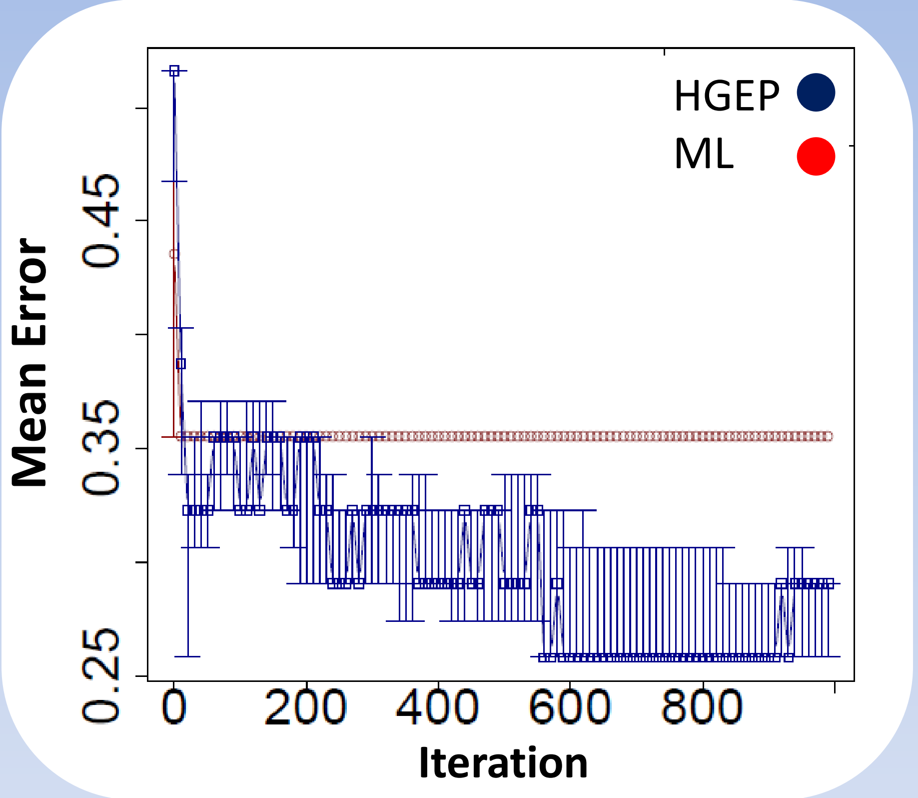
- Synthetic data set: An Erdos-Renyi model was used to generate a random sparse matrix which was perturbed with uniform noise to get a random rate matrix.



#Sequences	#Datapoints	#Characters	Mean Error		
			Baseline	ML	HGEP
1000	10000	4	0.703	**0.404**	0.446

Experiments

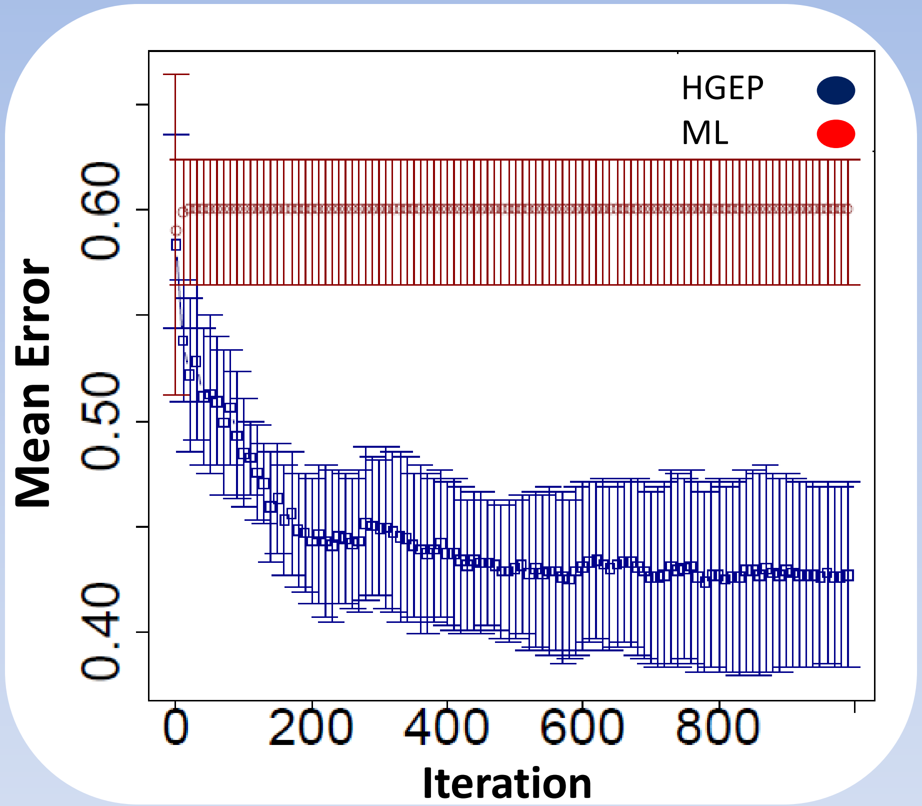
- MS data set: disease progression over 3 years in 72 MS patients



#Sequences	#Datapoints	#Characters	Mean Error			Relative Error Reduction
			Baseline	ML	HGEP	
72	384	3	0.516	0.355	**0.277**	**22%**

Experiments

- RNA data set: time series consisting of paths from one modern leaf to the root in a reconstructed phylogenetic tree (from a sample of 30 species)



#Sequences	#Datapoints	#Characters	Mean Error			Relative Error Reduction
			Baseline	ML	HGEP	
1000	6167	4	0.648	0.596	**0.426**	**29%**

Conclusion

- We introduced a nonparametric Bayesian model for continuous-time recurrent processes.
- The model has attractive properties such as conjugacy and closed form predictive distribution.
- We showed how inference can be done efficiently, using PMCMC methods.
- Our experiments showed the model is useful for analyzing complex real world datasets.
- We are currently working on extending the model by adding a layer of decision making; the resulting model is a Partially Observable Markov Decision Process (POMDP).

Thank You!

POSTER T091