A reversible infinite HMM using normalised random measures

Konstantina Palla, David A. Knowles, Zoubin Ghahramani

23rd of June 2014
**Motivation**

Assume a Markov chain $X_1, \ldots, X_t, \ldots, X_T$, which is *reversible*:

$$P(X_1, \ldots, X_t, \ldots X_T) = P(X_T, \ldots, X_t, \ldots, X_1)$$

**Applications**

- Modelling physical systems e.g transitions of a macromolecule conformation at fixed temperature.
- Chemical dynamics of protein folding.

**Tasks**

- Find the transition operation (transition matrix) of the reversible Markov chain
- Put a prior on the reversible Markov chain

This work: proposes a Bayesian non-parametric prior for reversible Markov chains.
**Problem:** Put prior on reversible Markov chains. What does that mean?

Reversible chains and random walk on weighted graph

\[ \mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W}) \] weighted undirected graph

- vertex-set \( \mathcal{V} = \{i, r, q, \ldots \} \)
- edge-set \( \mathcal{E} = \{e_{ir}, e_{iq}, e_{rq}, \ldots \} \)
- weight-set \( \mathcal{W} = \{J_{ir}, J_{rq}, J_{iq}, \ldots \} \)

Discrete-time *random walk* on \( \mathcal{G} \rightarrow \) Markov chain with \( X_t \in \mathcal{V} \) and transition matrix

\[ P(i, j) := \frac{J_{ij}}{\sum_k J_{ik}}, \]

Put a prior on the transition matrix \( P \) (or on the weights \( J \)s).
Seminal work by Diaconis, Freedman and Coppersmith.

**Markov Exchangeability**

A process on a *countable* space $S$ is *Markov exchangeable* if the probability of observing a path $X_1, \ldots, X_t, \ldots, X_T$ is only a function of $X_1$ and the transition counts $C(i,j) := |\{X_t = i, X_{t+1} = j; 1 \leq t < T\}|$ for all $i, j \in S$.

**Representation Theorem (Diaconis and Freedman, 1980)**

A process is Markov exchangeable and returns to every state visited infinitely often (recurrent), if and only if it is a mixture of recurrent Markov chains

$$P(X_2, \ldots, X_t, \ldots, X_T | X_1) = \int_{\mathcal{P}} \prod_{t=1}^{T-1} P(X_t, X_{t+1}) \mu(dP|X_1)$$

where $\mathcal{P}$ is the set of stochastic matrices on $S \times S$ and the mixing measure $\mu(\cdot | X_1)$ on $\mathcal{P}$ is uniquely determined.

**Problem:** Determine the prior $\mu$. Not always easy.
Random walk with reinforcement

- **Idea:** Simulate from the prior $\mu$.
- Increase the edge weight by $+1$ each time an edge is crossed.

$$\frac{1}{T}[J_{ir}, J_{rq}, J_{iq}] \xrightarrow{T \to \infty} [L_{ir}, L_{rq}, L_{iq}] \sim \mu$$

$T$ - total number of steps, $\mu$ - measure over edge weights, the underlying prior

- Process Markov exchangeable, recurrent $\rightarrow$ mixture of recurrent MCs

**Examples**

- Edge Reinforcement Random Walk (ERRW) Diaconis and Freedman [1980], Diaconis and Rolles [2006]; conjugate prior for the transition matrix for reversible MCs.
- Edge reinforced schema by Bacallado et al. [2013] extends ERRW to countably infinite space, reversible process, prior is difficult to characterise.
Define a prior over reversible Markov chains:

1. Explicitly characterize the measure $\mu$ over transition matrix
2. Define an Edge Reinforcement schema

**Proposed work:** Explicitly construct the prior $\mu$ over the weights (or equivalently the transition matrix)
A model for reversible Markov chains

**General idea:** Define the prior over the weights using the Gamma process *hierarchically*.

**Gamma process** $\Gamma P(\alpha_0 H)$

Completely random measure on $\mathcal{X}$ with Lévy measure

$$\nu(dw, dx) = \rho(dw)H(dx) = a_0 w^{-1} e^{-a_0 w} dw H(dx).$$

on the space $\mathcal{X} \times [0, \infty)$. $H$ is the base measure and $\alpha_0$ the concentration parameter.

$$G_0 := \sum_{i=1}^{\infty} w_i \delta_{X_i} \sim \Gamma P(\alpha_0 H)$$

*Countably infinite* collection of pairs $\{X_i, w_i\}_{i=1}^{\infty}$ sampled from a Poisson process with intensity $\nu$. 
A model for reversible Markov chains

Define the prior over the weights using the Gamma process hierarchically.

Model

1. First level: $\Gamma P$ over space $\mathcal{X}$

$$G_0 = \sum_{i=1}^{\infty} w_i \delta_{x_i} \sim \Gamma P(\alpha_0, \mu_0)$$

Set of states $S := \{x_i; x_i \in \mathcal{X}, i \in \mathbb{N}\}$, countably infinite.

2. Second level: $\Gamma P$ over space $S \times S$.

$$G = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} J_{ij} \delta_{x_i} \delta_{x_j} \sim \Gamma P(\alpha, \mu),$$

$$J_{ij} | \alpha, w_i, w_j \sim \text{Gamma}(\alpha w_i w_j, \alpha)$$

Base measure atomic on $S \times S$:

$$\mu(x_i, x_j) = G_0(x_i) G_0(x_j)$$

Non-reversible: Directed edges, $J_{ij} \neq J_{ji}$
A model for reversible Markov chains

Reversibility

Impose symmetry

\[ J_{ij} = J_{ji} \sim \text{Gamma}(\alpha w_i w_j, \alpha) \]

Proof: Sufficient to prove detailed balance

\[ \pi_i P(i, j) = \pi_j P(j, i) \]

where \( \pi_i = \frac{\sum_k J_{ik}}{\sum_j \sum_k J_{jk}} \), \( 0 < \sum_k J_{jk} < \infty \)

Corollary: \( \pi \) is the invariant measure of the chain.

We call the model the Symmetric Hierarchical Gamma Process (SHGP)
A MODEL FOR REVERSIBLE MARKOV CHAINS

Properties

- **Irreducibility**
  A MC is irreducible if $\exists t \in \mathbb{N}$ s.t $P^t_{ij} > 0$, $\forall i, j \in S$

  SHGP is irreducible: $J_{ij}, \sum_k J_{ik} \in (0, \infty) \rightarrow P_{ij} = \frac{J_{ij}}{\sum_k J_{ik}} > 0$ a.s $\forall i, j \in S$

- **Recurrence** A state $i$ is positive recurrent if

  $E(\tau_{ii}) < \infty, \tau_{ij} := \min\{t > 1: X_t = j | X_1 = i\}$

  The SHGP is positive recurrent since the following applies:

Theorem (Levin et al. [2006])

An irreducible Markov chain is positive recurrent iff there exists a probability distribution $\pi$ such that $\pi = \pi P$. 

Konstantina Palla 10 / 24
Representation Theorem

A process is Markov exchangeable and returns to every state visited infinitely often (recurrent), if and only if it is a mixture of recurrent Markov chains

\[ P(X_2, \ldots, X_t, \ldots, X_T | X_1) = \int_{\mathcal{P}} \prod_{t=1}^{T-1} P(X_t, X_{t+1}) \mu(dP|X_1) \]

where \( \mathcal{P} \) is the set of stochastic matrices on \( S \times S \) and \( \mu(\cdot|X_1) \) on \( \mathcal{P} \) is the mixing measure.

SHGP

- Explicitly defined prior \( \mu \); hierarchical construction of weights
- SHGP is a mixture of recurrent, reversible Markov chains
- SHGP is recurrent, Markov exchangeable and reversible.
The SHGP Hidden Markov Model

Finite number of states $K$. Countably infinite model as $K \to \infty$.

$$G_0 = \sum_{i=1}^{K} w_i \delta_{x_i}$$

$w_i \sim \text{Gamma}(\alpha_0 \mu_0(x_i), \alpha_0)$

$$G = \sum_{i=1}^{K} \sum_{j=1}^{K} J_{ij} \delta_{x_i, x_j}$$

$J_{ij} = J_{ji} \sim \text{Gamma}(\alpha w_i w_j, \alpha)$

$X_t \in \{1, \ldots, K\} - \text{hidden state sequence.}$

$E$ - emission matrix

$Y_{t}, t = 1, \ldots, T$ - observed sequence with observation model $F(\cdot|E)$

$Y_t|X_t, E \sim \text{iid} F(\cdot|E_{X_t})$

$\{E_k, k = 1, \ldots, K\}$ state emission parameters. $F$; multinomial, Poisson and Gaussian observation models
We ran the SHGP Hidden Markov model on 2 real world datasets with reversible underlying systems. Comparison against

- SHGP HMM non-reversible
- infinite HMM (HDP)
ChIP-seq allows us to measure what proteins, with what chemical modifications, are bound to DNA along the genome.

- $Y$ matrix $T \times L$, $T = 2 \cdot 10^4$ and $L = 6$: counts, how many reads for the protein of interest $l$ map to bin $t$.
- Poisson (multivariate) likelihood model $F$.

**Figure:** ChipSeq data for a small section of length 300 of the whole chromosome region, along with the $L = 6$ identifiers (proteins of interest)
Task: Predict held out values in $Y$.

**Table:** ChipSeq results for 10 runs using different hold out patterns (20%), a truncation level of $K = 30$, 1000 iterations and a burnin of 700.

<table>
<thead>
<tr>
<th>Model</th>
<th>Alogirthm</th>
<th>Train error</th>
<th>Test error</th>
<th>Train log likelihood</th>
<th>Test log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reversible</td>
<td>HMC</td>
<td>0.9122 ± 0.0032</td>
<td>1.1158 ± 0.0097</td>
<td>−1.0488 ± 0.0009</td>
<td>−3.2422 ± 0.0023</td>
</tr>
<tr>
<td>Non-rev</td>
<td></td>
<td>0.9127 ± 0.0033</td>
<td>1.1167 ± 0.0095</td>
<td>−1.0494 ± 0.0009</td>
<td>−3.2478 ± 0.0022</td>
</tr>
<tr>
<td>iHMM</td>
<td>Beam Sampler</td>
<td>0.9383 ± 0.0061</td>
<td>1.1365 ± 0.0107</td>
<td>−1.0727 ± 0.0041</td>
<td>−3.3047 ± 0.0027</td>
</tr>
</tbody>
</table>
SHGP recovers known types of regulatory regions

- promoters.
- enhancers.

**Figure**: Learnt emission matrix $L \times K$ for ChIP-seq dataset. Element $E_{lk}$ is the Poisson rate parameter for protein $l$ in state $k$. Brighter indicates higher values.
Single Ion Channel Recordings Dataset

- Patch clamp recordings is a method for measuring conformational changes in ion channels. These changes are accompanied by changes in electrical potential (measurements).
- $Y$ matrix $1 \times T, T = 10^4$: 10KHz recording of electrical potential measurements of a single alamethicin channel.
- Gaussian likelihood model $F$.

\[ Y_t|X_t, E \sim N(Y_t; \mu, \sigma), \]

where $\mu = E(X_t, 1)$ and $\sigma = E(X_t, 2)$ with $K \times 2$ emission matrix $E$.

Table: Ion channel results across 10 different random hold out patterns, a truncation of $K = 15$, 1000 iterations and a burnin of 700.

<table>
<thead>
<tr>
<th>Model</th>
<th>Algorithm</th>
<th>Train error</th>
<th>Test error</th>
<th>Train log likelihood</th>
<th>Test log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reversible</td>
<td>HMC</td>
<td>0.023 ± 0.001</td>
<td>0.030 ± 0.002</td>
<td>2.204 ± 0.055</td>
<td>2.034 ± 0.058</td>
</tr>
<tr>
<td>Non-reversible</td>
<td>HMC</td>
<td>0.027 ± 0.007</td>
<td>0.033 ± 0.007</td>
<td>2.108 ± 0.084</td>
<td>1.970 ± 0.078</td>
</tr>
<tr>
<td>iHMM</td>
<td>Beam sampler</td>
<td>0.038 ± 0.005</td>
<td>0.045 ± 0.004</td>
<td>2.134 ± 0.070</td>
<td>2.008 ± 0.058</td>
</tr>
</tbody>
</table>
Figure: Clusters found by the SHGP-HMM for the ion channel dataset, shown relative to a histogram of levels across the recording. The smaller clusters at higher currents are often merged in the model.
CONCLUSION AND FUTURE WORK

- Constructed non-parametric prior for reversible Markov chains
- Presented a finite approximation
- Experimental results using SHGP as part of HMM
- Experimental results underline the importance of accounting for reversibility

Future Work

- Construct sampler for the infinite case. Use of sampling process proposed by Favaro and Teh [2013].
- Look at the corresponding edge reinforcement schema (?)
Thank you!


Gamma process $\Gamma P(\alpha_0 H)$

Completely random measure on $\mathcal{X}$ with Lévy measure

$$
\nu(dw, dx) = \rho(dw)H(dx) = a_0w^{-1}e^{-a_0w} dw H(dx).
$$
on the space $\mathcal{X} \times [0, \infty)$. $H$ is the base measure and $\alpha_0$ the concentration parameter.

$$
G_0 := \sum_{i=1}^{\infty} w_i \delta_{X_i} \sim \Gamma P(\alpha_0 H)
$$

*Countably infinite* collection of pairs $\{X_i, w_i\}_{i=1}^{\infty}$ sampled from a Poisson process with intensity $\nu$. 
**Appendix B - Relation to Hierarchical Dirichlet and Hierarchical Gamma Process**

<table>
<thead>
<tr>
<th>HDP</th>
<th>HGP</th>
<th>SHGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G'_0 \sim DP(\alpha_0, \mu_0)$</td>
<td>$G_0 \sim \Gamma P(\alpha_0, \mu_0)$</td>
<td>$G_0 \sim \Gamma P(\alpha_0, \mu_0)$</td>
</tr>
<tr>
<td>$P_j \sim DP(\alpha' G'_0)$</td>
<td>$\tilde{J}_j \sim \Gamma P(\tilde{\alpha}, G_0)$</td>
<td>$J_j \sim \Gamma P(\alpha \omega_j, G_0)$</td>
</tr>
</tbody>
</table>

**Table:** HDP, HGP and SHGP. $P_j$ & $J_j$ refer to the $j$th row of the transition and weight matrix respectively.

- The HDP puts a prior over the *transition* matrix. SHGP puts a prior over the *weight* matrix, imposes symmetry, allows reversibility.
- The SHGP modulo the symmetrization is equivalent to the HDP with specific gamma distributions over the concentration parameters between the levels;

  $$\alpha'_j \sim \text{Gamma}(\alpha_0 \mu_0(\mathcal{X}), \frac{\alpha_0}{\alpha \omega_j})$$
Inference

- Hybrid Monte Carlo (HMC) to sample the weights $J_{ij}$
- Forward filtering, backward sampling, to sample state sequence $X_1, \ldots, X_T$
- iHMM : Beam sampler