A reversible infinite HMM using normalised random measures

Konstantina Palla, David A. Knowles, Zoubin Ghahramani

23rd of June 2014

UNIVERSITY OF CAMBRIDGE

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.

REVERSIBLE MARKOV CHAINS

Problem: Put prior on reversible Markov chains. *What does that mean?*

Reversible chains and random walk on weighted graph

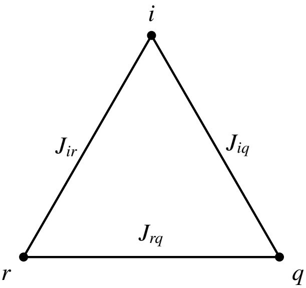
$\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$ weighted undirected graph

- 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 Js).



Basic Theory

Seminal work by Diaconis, Freedman and Coppersmith.

Markov Exchangeability

A process on a *countable* space $\mathcal{S}$ 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 $i, j \in \mathcal{S}$.

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

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

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

Problem: Determine the prior μ. Not always easy.

RELATED WORK

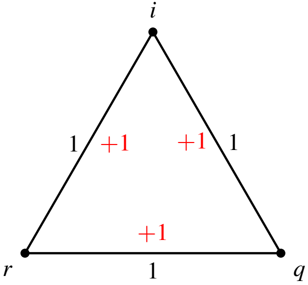
Random walk with reinforcement

- **Idea:** Simulate from the prior μ.
- 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, μ - 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.

Related work

Define a prior over reversible Markov chains:

1. Explicitly characterize the measure μ over transition matrix
2. Define an Edge Reinforcement schema

Proposed work: Explicitly construct the prior μ 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 $\Gamma\mathrm{P}(\alpha_0 H)$

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 α_0 the concentration parameter.

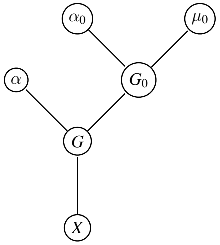
$$G_0 := \sum_{i=1}^{\infty} w_i \delta_{X_i} \sim \Gamma\mathrm{P}(\alpha_0 H)$$

Countably infinite collection of pairs $\{X_i, w_i\}_{i=1}^{\infty}$ sampled from a Poisson process with intensity ν.

A model for reversible Markov chains

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

Model



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

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

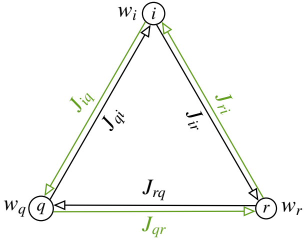
   Set of states $\mathcal{S} := \{x_i; x_i \in \mathcal{X},\ i \in \mathbb{N}\}$, *countably infinite*.

2. Second level: ΓP over space $\mathcal{S} \times \mathcal{S}$.

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

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

   Base measure atomic on $\mathcal{S} \times \mathcal{S}$: $\mu(x_i, x_j) = G_0(x_i) G_0(x_j)$



Non-reversible: Directed edges, $J_{ij} \neq J_{ji}$

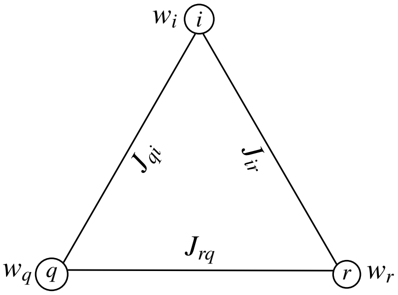
A model for reversible Markov chains

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: π is the invariant measure of the chain.



We call the model the **Symmetric Hierarchical Gamma Process (SHGP)**

Properties

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

  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 $\forall i, j \in \mathcal{S}$
- **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 π such that $\pi = \pi P$.

A MODEL FOR REVERSIBLE MARKOV CHAINS

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

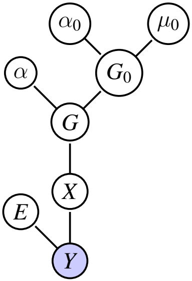
$$\mathrm{P}(X_2, \ldots, X_t, \ldots, X_T | X_1) = \int_{\mathcal{P}} \prod_{t=1}^{T-1} P(X_t, X_{t+1}) \mu(dP | X_1)$$

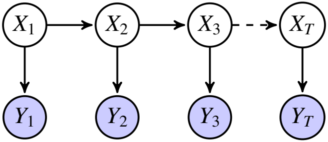
where $\mathcal{P}$ is the set of stochastic matrices on $\mathcal{S} \times \mathcal{S}$ and $\mu(\cdot | X_1)$ on $\mathcal{P}$ is the mixing measure.

SHGP

- Explicitly defined prior μ; 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 $K \to \infty$.

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

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

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

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

$X_t \in \{1, \ldots, K\}$ - *hidden* state sequence.
E - emission matrix
$Y_t, t = 1, \ldots, T$ - observed sequence with observation model $F(\cdot|E)$

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

$\{E_k, k = 1, \cdots, K\}$ state emission parameters. F; multinomial, Poisson and Gaussian observation models

Experiments

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 data from neural stem cells

- 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.

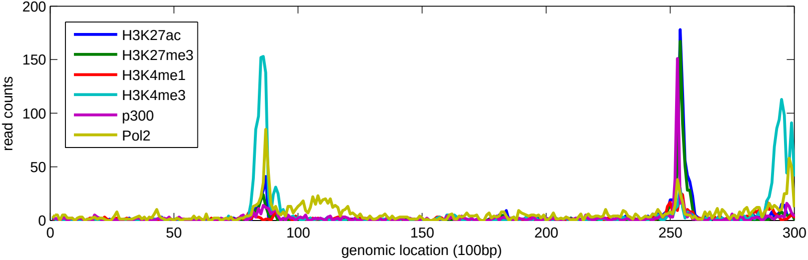


Figure: ChipSeq data for a small section of length 300 of the whole chromosome region, along with the $L = 6$ identifiers (proteins of interest)

CHIP-SEQ DATA FROM NEURAL STEM CELLS

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.

Model	Alogirthm	Train error	Test error	Train log likelihood	Test log likelihood
Reversible	HMC	$\mathbf{0.9122 \pm 0.0032}$	$\mathbf{1.1158 \pm 0.0097}$	$\mathbf{-1.0488 \pm 0.0009}$	$\mathbf{-3.2422 \pm 0.0023}$
Non-rev		0.9127 ± 0.0033	1.1167 ± 0.0095	-1.0494 ± 0.0009	-3.2478 ± 0.0022
iHMM	Beam Sampler	0.9383 ± 0.0061	1.1365 ± 0.0107	-1.0727 ± 0.0041	-3.3047 ± 0.0027

CHIP-SEQ DATA FROM NEURAL STEM CELLS

SHGP recovers known types of regulatory regions

- *promoters*.
- *enhancers*.

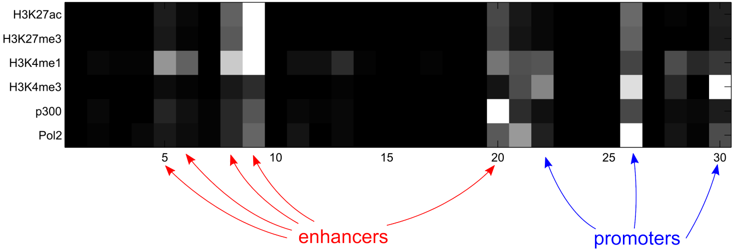


Figure: Learnt emission matrix $L \times K$ for ChIP-seq dataset. Element E_{lk} is the Poisson rate parameter for protein l in state k. Brighter indicates higher values

.

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.

Model	Alogirthm	Train error	Test error	Train log likelihood	Test log likelihood
Reversible	HMC	0.023 ± 0.001	0.030 ± 0.002	$\mathbf{2.204 \pm 0.055}$	$\mathbf{2.034 \pm 0.058}$
Non-reversible	HMC	0.027 ± 0.007	0.033 ± 0.007	2.108 ± 0.084	1.970 ± 0.078
iHMM	Beam sampler	0.038 ± 0.005	0.045 ± 0.004	2.134 ± 0.070	2.008 ± 0.058

Single ion channel recordings dataset

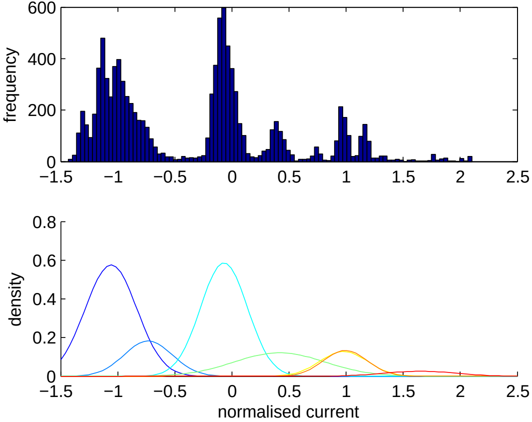


Figure: Clusters found by the SHGP-HMM for the ion channel dataset, shown relative to a histogram of levels across the recording. The smaller clusters at higher currents are often merged in the model.

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!

BIBLIOGRAPHY

Sergio Bacallado, Stefano Favaro, and Lorenzo Trippa. Bayesian nonparametric analysis of reversible markov chains. *The Annals of Statistics*, 41(2):870–896, 2013.

Perci Diaconis and David Freedman. De Finetti's theorem for Markov chains. *The Annals of Probability*, 8(1):115–130, 1980.

Persi Diaconis and Silke W. W. Rolles. Bayesian analysis for reversible markov chains. *The Annals of Statistics*, 34(3):pp. 1270–1292, 2006.

S. Favaro and Y. W. Teh. MCMC for normalized random measure mixture models. *Statistical Science*, 28(3):335–359, 2013.

David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

Appendix A

Gamma process $\Gamma\mathrm{P}(\alpha_0 H)$

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 α_0 the concentration parameter.

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

Countably infinite collection of pairs $\{X_i, w_i\}_{i=1}^{\infty}$ sampled from a Poisson process with intensity ν.

APPENDIX B - RELATION TO HIERARCHICAL DIRICHLET AND HIERARCHICAL GAMMA PROCESS

HDP	HGP	SHGP
$G'_0 \sim \mathrm{DP}(\alpha_0 \mu_0)$	$G_0 \sim \Gamma\mathrm{P}(\alpha_0, \mu_0)$	$G_0 \sim \Gamma\mathrm{P}(\alpha_0, \mu_0)$
$P_j \sim \mathrm{DP}(\alpha' G'_0)$	$\tilde{J}_j \sim \Gamma\mathrm{P}(\tilde{\alpha}, G_0)$	$J_j \sim \Gamma\mathrm{P}(\alpha w_j, G_0)$

Table: HDP, HGP and SHGP. P_j & J_j refer to the jth 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;

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

APPENDIX C

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