A reversible infinite HMM using normalised random measures

Konstantina Palla [1] David A. Knowles [2] Zoubin Ghahramani [1]

[1]University of Cambridge kp376@cam.ac.uk [2]Standford University

Abstract

We use the Gamma process to construct a nonparametric prior over reversible Markov chains. We use the resulting reversible Markov chain as the hidden sequence in a Hidden Markov model and present experimental results on two real datasets: epigenomics and ion channel recording.

Motivation

Reversible Markov chain:

$$\mathrm{P}(X_1,\dots,X_t,\dots X_T)=\mathrm{P}(X_T,\dots,X_t,\dots X_1)$$

Applications

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

Tasks

- Find the transition matrix of the reversible Markov chain.
- *Put a prior on the transition matrix.*

Background

Gamma process $\Gamma\mathrm{P}(\alpha_0 H)$: Completely random measure on $\mathcal{X}$ with Lévy measure

$$\nu(dw,dx)=\rho(dw)H(dx)=\alpha_0 w^{-1}e^{-\alpha_0 w}dw\ H(dx).$$

on the space $\mathcal{X}\times[0,\infty)$. H; base measure, α_0: concentration parameter.

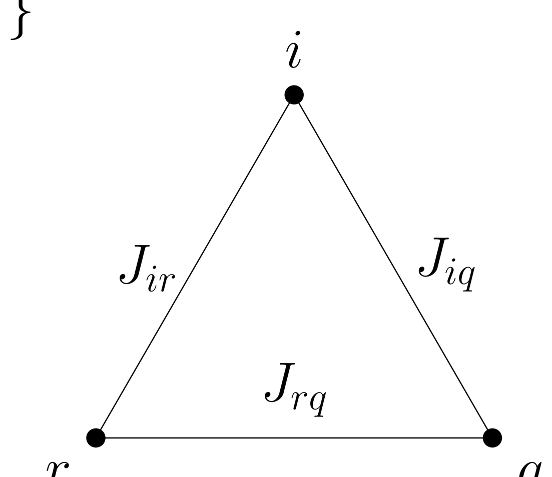
$$G_0:=\sum_{i=1}^{\infty} w_i\delta_{X_i}\sim\Gamma\mathrm{P}(\alpha_0 H)$$

Countably infinite collection of pairs $\{X_i,w_i\}_{i=1}^{\infty}$.

Random walk on a graph $\mathcal{G}$: Discrete-time *random walk* on $\mathcal{G}\rightarrow$ Markov chain with $X_t=k$, $k\in\{i,r,\dots\}$

& transition matrix

$$P(i,r):=\frac{J_{ir}}{\sum_k J_{ik}}$$



Put prior on the transition matrix P (or on weights Js).

Related work

- Edge Reinforced 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, no closed form for the prior.

Symmetric Hierarchical Gamma Process (SHGP)

Define a prior over the weights Js using the ΓP *hierarchically*.

1. ΓP over $\mathcal{X}$:

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

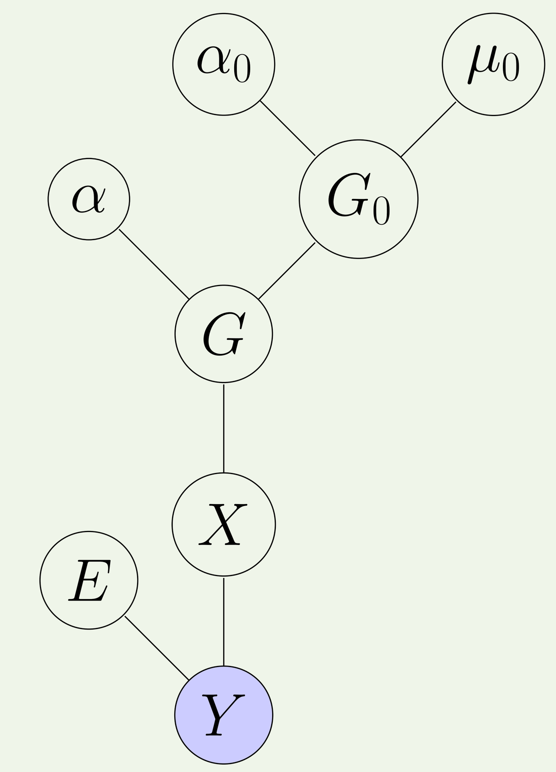
States $\mathcal{S}:=\{x_i; x_i\in\mathcal{X},\ i\in\mathbb{N}\}$, *countably infinite*.

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

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

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

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



Reversibility

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

Result: *detailed balance* holds $\rightarrow$ Reversible markov chain

$$\pi_i P(i,j)=\pi_j P(j,i)$$

where $\pi_i=\frac{\sum_k J_{ik}}{\sum_j\sum_k J_{ik}}$, $0<\sum_k J_{ik}<\infty$

Corollary: π is the invariant measure of the chain.

De Finetti Representation

[Representation Theorem, Diaconis & 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*

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

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.

- Explicitly defined prior μ (SHGP): hierarchical construction of J's.
- SHGP is a mixture of recurrent, reversible Markov chains.
- SHGP is recurrent, Markov exchangeable and reversible.

SHGP as part of a Hidden Markov Model

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

$$G_0=\sum_{i=1}^{K} w_i\delta_{x_i}\qquad\qquad G=\sum_{i=1}^{K}\sum_{j=1}^{K} J_{ij}\delta_{x_i,x_j}$$

$$w_i\sim\mathrm{Gamma}(\alpha_0\mu_0(x_i),\alpha_0)\qquad\qquad J_{ij}=J_{ji}\sim\mathrm{Gamma}(\alpha w_i w_j,\alpha)$$

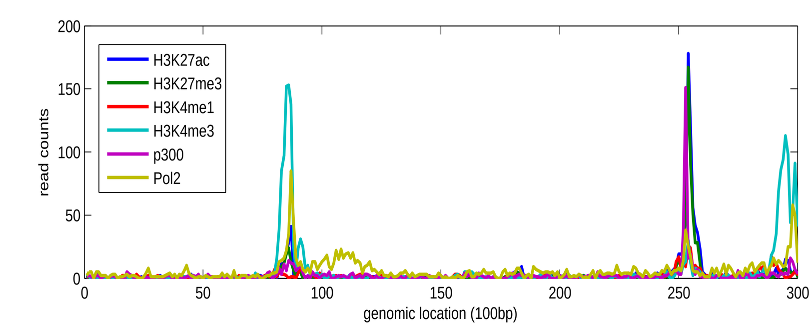
$$Y_t|X_t,E\sim^{iid}F(\cdot|E_{X_t})$$

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

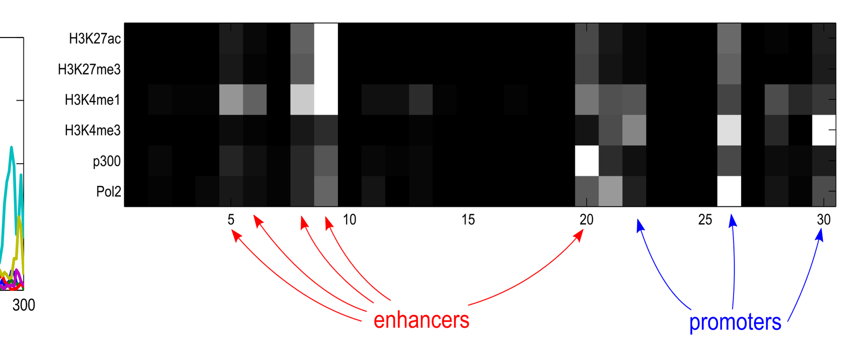
Experiments

ChIP-seq: measures what proteins are bound to DNA along the genome

- Y matrix $T\times L$, $T=20^4$ and $L=6$: counts, how many reads for the protein of interest l map to bin t
- Poisson (multivariate) likelihood model F



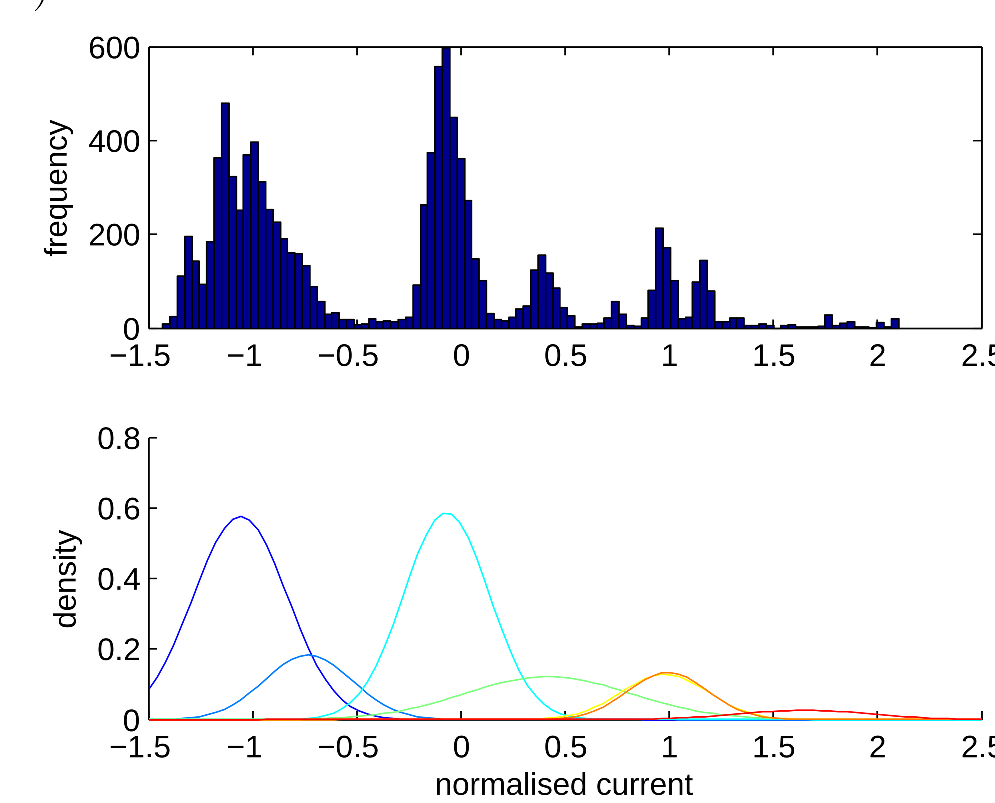
ChipSeq data for $L=6$ proteins



Learnt emission matrix

Patch clamp recordings: recordings of changes in electrical potential caused by conformational changes in ion channels.

- Y matrix $1\times T$, $T=10^4$: 10KHz recording of electrical potential measurements of a single alamethicin channel.
- Gaussian likelihood: $Y_t|X_t,E\sim N(Y_t;\mu,\sigma)$, $\mu=E(X_t,1)$, $\sigma=E(X_t,2)$ with $K\times 2$ emission matrix E.



Clusters found by SHGP shown relative to a histogram of levels across the recording

	ChIP-seq		Ion channel recording	
	Train log likelihood	Test log likelihood	Train log likelihood	Test log likelihood
Reversible	$\mathbf{-1.0488\pm0.0009}$	$\mathbf{-3.2422\pm0.0023}$	$\mathbf{2.204\pm0.055}$	$\mathbf{2.034\pm0.058}$
Non-rev	-1.0494 ± 0.0009	-3.2478 ± 0.0022	2.108 ± 0.084	1.970 ± 0.078
iHMM	-1.0727 ± 0.0041	-3.3047 ± 0.0027	2.134 ± 0.070	2.008 ± 0.058

References

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

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