A model for reversible Markov chains

Konstantina Palla, David A. Knowles, Zoubin Ghahramani

University of Cambridge

ERCIM, London 2013
Motivation - Problem

Background on reversible Markov chains

Proposed Model

Experiments

Conclusion - Future Work
Assume a Markov chain $X_1, \ldots, X_t, \ldots X_T$, 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 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 $\mathcal{V} = \{i, j, \ldots\}$
- edge-set $\mathcal{E} = \{e_1, e_2, \ldots\}$
- weight-set $\mathcal{W} = \{w_{ij}, w_{iq}, \ldots\}$

Discrete-time *random walk* on $\mathcal{G} \rightarrow$ Markov chain with $X_t = k, k \in \mathcal{V}$ and transition matrix

$$P(i, j) := \frac{w_{ij}}{\sum_k w_{ik}} ,$$

Put priors on the transition matrix $P$ (or on weights).
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 $\mu$ (closed under sampling)
- Increase the edge weight by 1 each time an edge is crossed.
- $T \to \infty \Rightarrow [L_{ij}, L_{qi}, L_{iq}] \sim \mu$
- $T$ - 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, no closed form for the prior.
**A model for reversible Markov chains**

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

**Gamma process** \( \Gamma(\alpha_0 H) \)
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.

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

*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

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

**Model**

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

   $$G_0 = \sum_{i=1}^{\infty} w_i^0 \delta_{\theta_i} \sim \Gamma\mathcal{P}(\alpha_0 H)$$

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

2. **Second level**: $\Gamma\mathcal{P}$ over space $\mathcal{S} \times \mathcal{S}$.

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

   $$J_{ij} | \alpha, w_i^0, w_j^0 \sim \Gamma(\alpha w_i^0 w_j^0, \alpha)$$

   Base measure atomic on $\mathcal{S} \times \mathcal{S}$:

   $$\mu(\theta_i, \theta_j) = G_0(\theta_i) G_0(\theta_j)$$
1. Put weights on the nodes:
\[ G_0 = \sum_{i=1}^{\infty} w_i^0 \delta_{\theta_i} \sim \Gamma P(\alpha_0 H) \]

2. Put weights on the edges:
\[ G = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} J_{ij} \delta_{\theta_i \theta_j} \sim \Gamma P(\alpha \mu), \]
\[ J_{ij} | \alpha, w_i^0, w_j^0 \sim \Gamma(\alpha w_i^0 w_j^0, \alpha) \]

**Non-reversible**: Directed edges, \( J_{ij} \neq J_{ji} \)
Reversibility

Impose symmetry \( J_{ij} = J_{ji} \sim \Gamma(\alpha w_i^0, w_j^0, \alpha) \)

Proof: Sufficient to prove **detailed balance**

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

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

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

- \( G \) is not a completely random measure anymore.
- Each row \( G_i \) is a completely random measure marginally.

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. \( P_{ij}^t > 0 \), \( \forall i, j \in S \)

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

- **Aperiodicity**
  Period of a state \( X_i = \gcd(T_{ii}), T_{ii} := \{t \geq 1 : P_t(i, i) > 0\} \)

  SHGP is aperiodic:
  \( P(i, i)^t > 0 \rightarrow \gcd(T_{ii}) = 1, \forall i \in S \)
A model for reversible Markov chains

Properties

- **Recurrence** A state $i$ is positive recurrent if $E(\tau_{ii}) < \infty$, $\tau_{ij} := \min\{t > 1 : X_t = j | X_1 = i\}$

  SHGP is positive recurrent

Theorem (Levin et al. [2006])

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

- **Convergence**

  SHGP converges to the stationary distribution $\pi$

Theorem (Levin et al. [2006])

Irreducibility, aperiodicity and positive recurrence ensure that the invariant distribution $\pi$ is unique and $\forall i \in S$, $\lim_{t \to \infty} ||P^t(i, \cdot) - \pi||_{TV} = 0$

$TV$ denotes the total variation distance between the two distributions.
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

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

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.
**Finite Approximation**

**Infinite Divisibility**
For a $\Gamma P$ $G_0$ on $\mathcal{X}$, for each $K = 1, 2, \ldots$, there exists a sequence of i.i.d random variables $G_0(A_1) + \cdots + G_0(A_K)$ such that

$$G_0(\mathcal{X}) = G_0(A_1) + \cdots + G_0(A_K)$$

where $A_1, \ldots, A_K$ disjoint and measurable and $\mathcal{X} = \bigcup_{i=1}^{K} A_i$.

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

$$G_0 = \sum_{i=1}^{K} w^0_i \delta_{\theta_i}$$

$$w^0_i \sim \text{Gamma}(\alpha_0 \mu_0(\theta_i), \alpha_0)$$

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

$$J_{ij} = J_{ji} \sim \text{Gamma}(\alpha w^0_i w^0_j, \alpha)$$
**Finite Approximation**

**Hidden Markov model**

![Graph of Hidden Markov Model]

**Inference**

- *Y* - observed sequence. *X* - hidden state sequence
  - Hybrid Monte Carlo to sample the weights $J_{ij}$
  - Forward filtering, backward sampling, to sample state sequence $X_1, \ldots, X_T$. 

Konstantina Palla
ChIP-seq allows us to measure what proteins, with what chemical modifications, are bound to DNA along the genome

- $Y$ matrix $T \times L$, $T = 10^5$ and $L = 6$
- Poisson (multivariate) likelihood
SHGP recovers known types of regulatory regions

- **promoters**: Regulatory elements close to genes, with more H3K4me3 than H3K4me1.
- **enhancers**: Regulatory elements further from genes, with more H3K4me1 than H3K4me3.

The associated level of H3K27ac and PolII is indicative of how active these regions are.
CONCLUSION AND FUTURE WORK

• Constructed non-parametric prior for reversible Markov chains
• Presented a finite approximation
• Initial experimental results

Future Work

• Construct sampler for the infinite case. Use of sampling process proposed by Favaro and Teh [2013].
• Comparison to other models e.g infinite Hidden Markov Model (iHMM) by Beal et al. [2002] (non reversible, quantitative comparison).
• Look at the corresponding edge reinforcement schema (?)
Thank you!


