Reversibility and Time Reversals

Detailed Balance Equations:

A distribution π over S satisfies the detailed balance equations if:

$$\forall x,y \in S \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

Theorem: Any distribution π that satisfies the detailed balance equations for a Markov chain with transition matrix P, is a stationary distribution for P.

$$\text{Detailed Balance Equations} \implies \pi P = \pi$$

Proof:

ith column

$$\pi P(i) = \pi P^{i} = \sum_j \pi(j) P(j,i) \overset{DBE}{=} \sum_j \pi(i) P(i,j) = \pi(i)$$

☒

→ If π satisfies DBE for P, then seeing a $x \to y$ transition is as likely as seeing a $y \to x$ transition; if we start from π.

→ Evenmore, if a chain (X_t) has the initial distribution π, then $(X_0, \dots, X_n)$ has the same distribution as $(X_n, \dots, X_0)$. More Concretely,

$$\forall x_0, x_1, \dots, x_n, \quad \Pr_\pi[X_0 = x_0, \dots, X_n = x_n] = \Pr_\pi[X_0 = x_n, \dots, X_n = x_0].$$

→ This means that we can't distinguish the chain going forward or backward in time! Thus, such a chain is called reversible.

Time Reversal

Definition: The time reversal of an irreducible Markov chain with transition matrix P and stationary distribution π is the chain with matrix:

$$\hat{P}(x,y) := \frac{\pi(y)}{\pi(x)} P(y,x)$$

Theorem: Let $(X_t)_t$ be an irreducible chain with transition matrix P and stationary distribution π. Write $(\hat{X}_t)_t$ for the time-reversed chain with transition matrix $\hat{P}$. Then a) π is stationary for $\hat{P}$ and b) for any $x_0, \dots, x_t \in S$,

$$\mathbb{P}r_\pi[X_0 = x_0, \dots, X_t = x_t] = \mathbb{P}r_\pi[\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0].$$

Proof:

a) $\pi\hat{P}(x) = \sum_y \pi(y)\hat{P}(y,x) = \sum_y \pi(y) \frac{\pi(x)}{\pi(y)} P(x,y) = \pi(x)$

b) To show that the trajectory probabilities are equal, note that

$$\begin{aligned}\mathbb{P}r_\pi[X_0 = x_0, \dots, X_t = x_t] &= \pi(x_0)P(x_0,x_1)\dots P(x_{t-1},x_t)\\ &= \pi(x_t)\hat{P}(x_t,x_{t-1})\dots\hat{P}(x_1,x_0) = \mathbb{P}r_\pi[\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0]\end{aligned}$$

(Because $\pi(x_i)P(x_i,x_{i+1}) = \pi(x_{i+1})\hat{P}(x_{i+1},x_i)$)

$\boxtimes$

$\rightarrow$ Observe that if a chain is reversible (i.e., satisfies DBE) then $P = \hat{P}$.

Markov Chain Monte Carlo

Let's say we have a distribution π over a set S that we would like to sample from. One way of doing this is to create a Markov chain, P, over S with stationary distribution π. If this chain converges to its stationary distribution, then after running the chain for some initial steps, it will have distribution $\sim \pi$.

Metropolis Chains:

Suppose Q is a symmetric, irreducible transition matrix over the set of states S. The uniform distribution satisfies the detailed balance equations for Q:

$$\pi_Q \sim \text{Uniform}(S) \Rightarrow \forall x, y \in S \quad \pi_Q(x) Q(x,y) = \pi_Q(y) Q(y,x).$$

Thus, the uniform distribution is the unique stationary distribution of Q. The Metropolis algorithm suggests a method for "modifying" the dynamics of Q such that the stationary distribution becomes π.

This is how the modified dynamics are:

→ At each state x, generate a candidate next state $y \sim Q(x,\cdot)$

→ Accept the new candidate state with probability $a(x,y)$; o.w. stay in x.

This new chain has the following transition dynamics P:

$$P(x,y) = \begin{cases} Q(x,y)\, a(x,y), & y \neq x \\ 1 - \sum_{z \neq x} Q(x,z)\, a(x,z), & y = x \end{cases}$$

Metropolis Chains

We would like $a(x,y)$'s be such that the stationary distribution of P be π. So we find them so that π satisfies the detailed balance equations:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \Rightarrow \pi(x)Q(x,y)a(x,y) = \pi(y)Q(y,x)a(y,x)$$

$$\Rightarrow \underbrace{\pi(x)a(x,y)}_{\alpha} = \underbrace{\pi(y)a(y,x)}_{\beta}.$$

Because $a(\cdot,\cdot)$ is a probability, we must have $\alpha = \beta \leq \pi(x), \pi(y)$. Thus, $\alpha \leq \min\{\pi(x), \pi(y)\}$. We would like to reject transitions as hardly as possible; so we maximize the acceptance probability by setting:

$$\alpha = \min\{\pi(x), \pi(y)\} \Rightarrow \pi(x)a(x,y) = \min\{\pi(x), \pi(y)\}$$

$$\Rightarrow a(x,y) = \min\{1, \frac{\pi(y)}{\pi(x)}\}.$$

So the new dynamics is as follows:

$$P(x,y) = \begin{cases} Q(x,y)\min\{1, \frac{\pi(y)}{\pi(x)}\}, & \text{if } x \neq y \\ 1 - \sum_{z \neq x} Q(x,z)\min\{1, \frac{\pi(z)}{\pi(x)}\}, & \text{if } x = y \end{cases}$$

→ Basically, if $\pi(y) \geq \pi(x)$ we always accept the transition from x to y. However, if $\pi(y) < \pi(x)$ we reject the transition with probability $1 - \frac{\pi(y)}{\pi(x)}$.

Metropolis - Hastings Chains

Hastings extended Metropolis chains so that the initial transition dynamics (Q) need not be symmetric, only irreducible.

The general idea is the same: we sample a transition from Q and accept it with probability $a(x,y)$. However, here for the new chain to be time-reversable with stationary distribution π, $a(x,y)$ must be different.

Writing the detailed balance equations for the new chain, P, and π we get:

$$\pi(x) P(x,y) = \pi(y) P(y,x) \Rightarrow \pi(x) Q(x,y) a(x,y) = \pi(y) Q(y,x) a(y,x)$$

$$\Rightarrow a(x,y) \leqslant \frac{\pi(y) Q(y,x)}{\pi(x) Q(x,y)}$$

This gives rise to the following chain:

$$P(x,y) = \begin{cases} Q(x,y) \min\left\{ \frac{\pi(y) Q(y,x)}{\pi(x) Q(x,y)}, 1 \right\} & \text{if } y \neq x \\ 1 - \sum_{z \neq x} Q(x,z) \min\left\{ \frac{\pi(z) Q(z,x)}{\pi(x) Q(x,z)}, 1 \right\} \end{cases}$$

Glauber Dynamics

Often, we study chains whose state space is contained in a set of the form S^V where S is a finite label set and V is the vertex set of a graph. We can think of S^V as the set of labelings of the vertices of the graph with S. Elements of S^V are called configurations.

Definition: [Glauber Dynamics]

Given a distribution π on a space of configurations, the Glauber dynamics for π is a Markov chain with stationary distribution π.

In statistics, it is often referred to as the Gibbs sampler.

Intuition:

At a high level, the Glauber chain moves as follows:

when at a state x, select a vertex v uniformly from V. The new state is chosen according to the measure π, conditioned on the set of states that are equal to x except possibly at v.

Glauber Dynamics (Cont.)

Definition: [Glauber Dynamics]

Let S and V be finite sets and let $\mathcal{X} \subseteq S^V$. Let π be a probability distribution over $\mathcal{X}$. The Glauber dynamics for π is a reversible Markov chain with state space $\mathcal{X}$, stationary distribution π and transition probabilities as described below:

For a configuration $x \in \mathcal{X}$ and vertex $v \in V$, let

$$\mathcal{X}(x,v) = \{ y \in \mathcal{X} : y(w) = x(w) \text{ for all } w \neq v \}$$

be the set of states agreeing with x everywhere except possibly at v.

Define

$$\pi^{x,v}(y) = \pi(y \mid \mathcal{X}(x,v)) = \begin{cases} \dfrac{\pi(y)}{\pi(\mathcal{X}(x,v))}, & \text{if } y \in \mathcal{X}(x,v) \\ 0, & \text{if } y \notin \mathcal{X}(x,v) \end{cases}$$

to be the distribution of π conditioned on $\mathcal{X}(x,v)$. The rule for updating a configuration is as follows: pick a vertex v uniformly from V, and choose a new configuration according to $\pi^{x,v}$.

→ π is always stationary and reversible for Glauber dynamics.