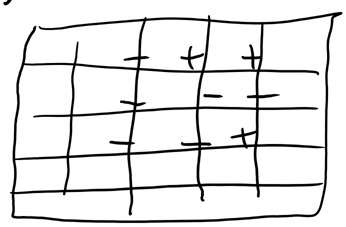
# MCMC

Sunday, October 22, 2017 7:14 PM

In many fields e.g. statistics, physics we want to understand complicated high dimensional distributions.

E.G. Ising Model



$\sigma \in \{+1, -1\}^V$

$\mu(\sigma) = \frac{1}{Z} \exp(\beta \sum_{u \sim v} \sigma_u \sigma_v)$

Random Colouring $\sigma \in C^V$

$$\mu(\sigma) = \frac{1}{Z} \cdot \prod_{i \sim j} I(\sigma_i \neq \sigma_j)$$

$Z = \#$ colourings

Q: Can we sample from $\mu$, compute $\int f \, d\mu$ or evaluate $Z$?

- Often efficient methods use Markov chains.

## MC Law of Large Numbers

If $X_n$ is a finite-state ergodic M.C. with stationary distribution $\pi$ then for any $x$, given $X_0 = x$,

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i) \longrightarrow \int f \, d\pi. \quad \text{in probability.}$$

Proof: We will calculate

$$Var\left(\sum_{i=1}^{n} f(X_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(f(X_i), f(X_j))$$

By our coupling argument,

$$|P_{xy}^{\ell} - \pi_y| \leq c_1 e^{-c_2 \ell}.$$

$$\mathbb{E} f(X_n) = \sum_a f(a) P_{xa}^n = \sum_a f(a)\pi_a + \sum_a f(a)(P_{xa}^n - \pi_a)$$

$$= \int f \, d\pi + O(e^{-cn})$$

So $\mathbb{E} S_n = n\int f(a)\, d\pi + O(1)$.

Now if $i<j$, $f(X_i) \sim \mu$ then

$$Cov(f(X_i), f(X_j)) = \mathbb{E} f(X_i)f(X_j) - \mathbb{E} f(X_i) f(X_j)$$

$$= \sum_{a,b} f(a) f(b)\ \mathbb{P}[X_i = a, X_j = b \mid X_0 = x]$$

$$- \left(\sum_a f(a)\,\mathbb{P}[X_i = a \mid X_0 = x]\right)\left(\sum_b f(b)\,\mathbb{P}[X_j = b \mid X_0 = x]\right)$$

$$= \sum_a f(a) P_{xa}^i \left(\sum_b f(b)\left(P_{ab}^{j-i} - P_{xb}^j\right)\right)$$

$$\leq \sum_a |f(a)| P_{xa}^i \ R\ \|f\|_\infty 2c_1 (e^{-c_2|j-i|} \qquad R = \#\text{states}$$

$$\leq \|f\|_\infty^2 \cdot 2Rc_1 e^{-c_2|j-i|}$$

$$Var(S_n) \leq 2c_1 R \|f\|_\infty^2 \sum_{i=1}^{n}\sum_{k=0}^{n} 2\cdot e^{-c_2 k}$$

$$\leq C n.$$

So $\mathbb{P}[|\frac{1}{n}S_n - \mathbb{E}\frac{1}{n}S_n| > \varepsilon]$

$$\leq \frac{\mathrm{Var}(\frac{1}{n}S_n)}{\varepsilon^2} \leq \frac{\frac{1}{n^2}Cn}{\varepsilon^2} = \frac{C}{n\varepsilon^2} \to 0.$$

Since $\mathbb{E}\frac{1}{n}S_n \to \int f\,d\pi$, $\frac{1}{n}S_n \xrightarrow{P} \int f\,d\pi$.

---

A Markov Chain $X_n$, started from its stationary distribution is <u>reversible</u> if $\forall n$,

$$(X_0, X_1, \dots, X_n) \overset{d}{=} (X_n, X_{n-1}, \dots, X_0).$$

Check case $n=1$.

$$\mathbb{P}[X_0 = x, X_1 = y] = \mathbb{P}[X_0 = y, X_1 = x]$$

so $\boxed{\pi_x P_{xy} = \pi_y P_{yx}}$ called <u>Detailed Balance Equations</u>

<u>Lemma</u> DBE $\Rightarrow$ $\pi$ is stationary

<u>Proof:</u>

$$(\pi P)_y = \sum_x \pi_x P_{xy} = \sum_x \pi_y P_{yx} = \pi_y \checkmark$$

<u>Lemma</u> DBE $\Rightarrow$ $\mathbb{P}[X_0 = x_0, \dots, X_n = x_n] = \mathbb{P}[X_0 = x_n, \dots, X_n = x_0]$

<u>Proof:</u>

$$\mathbb{P}[X_0 = x_0, \dots, X_n = x_n] = \pi_{x_0} \prod_{i=1}^n P_{x_{i-1} x_i}$$

$$= \pi_{x_0} \prod_{i=1}^n P_{x_i x_{i-1}} \cdot \frac{\pi_{x_i}}{\pi_{x_{i-1}}}$$

$$= \pi_{x_n} \prod_{i=1}^n P_{x_i x_{i-1}}$$

$$= \mathbb{P}[X_0 = x_n, \dots, X_n = x_0]$$

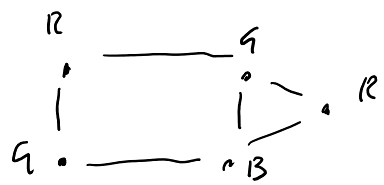
## Gibbs Sampler / Heat Bath / Glauber dynamics.

Let $\pi$ be a probability measure over $\sigma \in C^V$

Markov transition Rule. $X \mapsto X'$

- Pick $I \in V$ u.a.r.
- Choose $Y \sim \pi(\sigma_I \mid \sigma_{V \setminus \{I\}} = X_{V \setminus \{I\}})$
- Set $X'_I = Y$, $X'_j = X_j$ for $j \neq i$.

<u>Example</u> $\pi$ random colouring




- pick $I$ from $V$
- delete $X(I)$
- Pick new $X(I)$ uniformly given neighbours

<u>Example</u> Ising $\mathbb{P}[\sigma] = \frac{1}{Z} \exp(\beta \sum_{i \sim j} \sigma_i \sigma_j)$

Let $\sigma^{i,\pm}$ - assign $i$ to $\pm$.

$$\pi(\sigma^{i,+} \mid \sigma(V \setminus \{i\}))$$

$$= \frac{\pi(\sigma^{i,+})}{\pi(\sigma^{(i,+)}) + \pi(\sigma^{i,-})}$$

$$= \frac{\frac{1}{Z} \exp(\beta \sum_{u \sim v} \sigma_u^{i,+} \sigma_v^{i,-})}{\text{''} \qquad + \frac{1}{Z} \exp(\beta \sum \sigma_u^{i,-} \sigma_v^{i,-})}$$

$$= \frac{\exp(\beta \sum_{j \sim i} \sigma_j)}{\exp(\beta \sum_{j \sim i} \sigma_j) + \exp(-\beta \sum_{j \sim i} \sigma_j)}$$

$$= \frac{1}{2} + \frac{1}{2} \tanh[\beta \sum_{j \sim i} \sigma_i]$$

Simple function of the local neighbourhood.

---

Coupling: If $G = (V, E)$, $|V| = n$, $\pi$ Ising measure $d = \max_{i \in V} \deg(i)$. If $\beta < 1/d$ and $X_n$ is the Glauber dynamics then, $\exists C_\beta > 0$

$$d_{TV}(X_t, \pi) \leq n \exp(-(1-\beta d) t/n)$$

Proof: Let $Y_0 \sim \pi$ & couple $X_t, Y_t$ as follows

- pick same vertex $v_t$ to update at time $t$
- let $U_t \sim \mathrm{Unif}[0,1]$. IID
- Set $X_{t+1}(v_t) = \begin{cases} - & \text{if } U_t \leq \frac{1}{2} - \frac{1}{2}\tanh(\beta \sum_{u \sim v_t} X_t(u)) \\ + & \text{if } U_t \geq \frac{1}{2} - \frac{1}{2}\tanh(\beta \sum_{u \sim v_t} X_t(u)) \end{cases}$

$$Y_{t+1}(v_t) = \begin{cases} - & \text{if } U_t \leq \frac{1}{2} - \frac{1}{2}\tanh \beta \sum_{u \sim v_t} (Y_t(u)) \\ + & \text{o.w.} \end{cases}$$

$$\mathbb{P}[X_{t+1}(v_t) \neq Y_{t+1}(v_t) \mid v_t, X_t, Y_t]$$

$$= \frac{1}{2} \left| \tanh(\beta \sum_{u \sim v_t} X_t(u)) - \tanh(\beta \sum_{u \sim v_t} Y_t(u)) \right|$$

$$\leq \frac{1}{2} \beta \left| \sum_{u \sim v_t} X_t(u) - \sum_{u \sim v_t} Y_t(u) \right|$$

$$\leq \tfrac{1}{2}\beta \Big| \sum_{u\sim v_t} X_t(u) - \sum_{u\sim v_t} Y_t(u) \Big|$$

$$\leq \beta \,\#\{u \sim v_t : X_t(u) \neq Y_t(u)\}.$$

Let $D_t = \{u : X_t(u) \neq Y_t(u)\}$ & $\rho_t = |D_t|$.

Then $\mathbb{E}[\rho_{t+1} - \rho_t]$

$$= \mathbb{E}\left[ I(v_t \in D_{t+1}) - I(v_t \in D_t) \right]$$

$$= \mathbb{E}\, \beta \sum_{u\sim v_t} I(X_t(u) \neq Y_t(u)) - \mathbb{P}[v_t \in D_t]$$

$$= \beta\, \mathbb{E} \sum_{u \in D_t} I(v_t \sim u) - \frac{\mathbb{E}|D_t|}{n}$$

$$\leq \frac{\beta d}{n} \mathbb{E}|D_t| - \mathbb{E}|D_t| = \frac{\beta d - 1}{n} \mathbb{E}\rho_t$$

So $\mathbb{E}\rho_{t+1} \leq \left(1 - \frac{1-\beta d}{n}\right) \mathbb{E}\rho_t$

So $\mathbb{E}\rho_t \leq \left(1 - \frac{1-\beta d}{n}\right)^t \rho_0$

$$\leq n \exp(-(1-\beta d)t/n).$$

## Metropolis Hastings

Sometimes it is not easy to sample from the conditional probability.

Let $Q_{xy}$ be a Markov transition matrix. The idea is that is should be easy to simulate.

Set

$$A_{xy} = \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}} \wedge 1 \quad \text{acceptance probability}$$

and

$$P_{xy} = A_{xy} Q_{xy}$$

$$P_{xx} = 1 - \sum_{y \neq x} P_{xy}.$$

<u>Check DBE:</u> Assume $\pi_x Q_{xy} \geq \pi_y Q_{yx}$

$$\pi_x P_{xy} = \pi_x A_{xy} Q_{xy} = \pi_x Q_{xy} \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}}$$

$$= \pi_y Q_{yx}$$

$$= \pi_y A_{yx} Q_{yx} = \pi_y P_{yx} \checkmark$$

---

<u>Example:</u> Hidden Markov Model

$X_n$ a Markov Chain t.m. $P_{xx'}$

$$\begin{array}{ccccccc} X_1 & - & X_2 & - & X_3 & - \cdots - & X_n \\ | & & | & & & & | \\ Y_1 & & Y_2 & & \cdots & & Y_n \end{array}$$

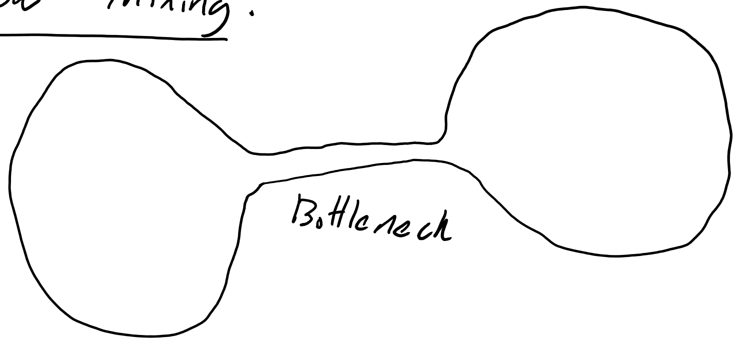
Markov transition $X_i$ to $Y_i$: $R_{xy}$.

Sample $(X_1, \ldots, X_n)$ given $\underline{Y} = (Y_1, \ldots, Y_n)$, estimate $\mathbb{P}[X_n = \cdot \mid \underline{Y} = \underline{y}]$.

---

Slow Mixing:

## Slow Mixing:




## Bottleneck Ratio:

$$\Phi = \min_{S : \pi(S) \leq \frac{1}{2}} \frac{\sum_{x \in S, y \in S^c} \pi(x) P_{xy}}{\pi(S)}$$

**Theorem:** If $X_0 \sim \pi|_S$ then

$$d_{TV}(X_n, \pi) \geq \frac{1}{2} - n\,\Phi.$$

**Proof:** Let $Y_0 \sim \pi$.

$$\begin{aligned}
\mathbb{P}[X_n \in S^c] &\leq \sum_{i=1}^{n} \mathbb{P}[X_{i-1} \in S, X \in S^c] \\
&= \sum_{i=1}^{n} \mathbb{P}[Y_{i-1} \in S, Y_i \in S^c \mid Y_0 \in S] \\
&\leq \sum_{i=1}^{n} \frac{\mathbb{P}[Y_{i-1} \in S, Y_i \in S^c, Y_0 \in S]}{\mathbb{P}[Y_0 \in S]} \\
&\leq \sum_{i=1}^{n} \frac{\sum_{x \in S, y \in S^c} \mathbb{P}[Y_{i-1} \in S, Y_i \in S^c]}{\pi(S)} \\
&= \sum_{i=1}^{n} \frac{\sum \pi_x P_{xy}}{\pi(S)} = n\,\Phi
\end{aligned}$$

So $d_{TV}(X_n, \pi) \leq \pi(S^c) - \mathbb{P}[X_n \in S^c]$

$$\leq \frac{1}{2} - n\,\Phi.$$