Stats 318: Lecture # 14

Agenda: MCMC

- Exact simulation
- The Propp-Wilson algorithm
- Coupling from the past
- Monotone systems

The Propp-Wilson algorithm

- In 1996, Jim Propp & David Wilson proposed a remarkable algorithm

  (i) Producing an output whose distribution is exactly the equil. dist. π

  (ii) And which stops in finite time (time is however random)

The basic idea

- P transition matrix with invariant dist. π
- Equivalent description: $X_{t+1} = f_{\theta_{t+1}}(X_t)$

$$\therefore \qquad X_t = f_{\theta_t} \circ f_{\theta_{t-1}} \ldots \circ f_{\theta_1}(X_0)$$

- Now consider

$$\begin{aligned} Y_t &= f_{\theta_1} \circ \ldots \circ f_{\theta_t}(X_0) \\ D_t &= f_{\theta_1} \circ \ldots \circ f_{\theta_t} \end{aligned}$$

- Classical MCMC algorithms run $\{X_t\}$
- Propp-Wilson algorithm runs D_t (or Y_t with different values of X_0)
  $\longrightarrow$ This is the main original feature

Propp-Wilson algorithm

Notation: CST is the set of constant mappings from $\mathcal{X}$ to $\mathcal{X}$

$$f \in \text{CST} \Leftrightarrow \quad f(x) = f(y) \quad \forall\, x, y \in \mathcal{X}$$

- Initialization: $D_0(x) = x \quad \forall\, x \in \mathcal{X}$
- While $D_t \notin$ CST
  - Draw θ_{t+1}
  - $D_{t+1} = D_t \circ f_{\theta_{t+1}}$

When algorithm stops (if it stops) we hold $Y_\tau = D_\tau(x) \;\; \forall\, x$

$$\tau = \inf \{t \geq 1 \;:\; D_t \in \text{CST}\}$$

- τ called time of coalescence
- Y is a random variable

Exact sampling

Theorem

Assume P irreducible & $\mathbb{P}(\tau < \infty) = 1$. Then Y is distributed as π

Proof: Key point is this:

$$\text{For } t \geq \tau, D_t = D_\tau, \text{ i.e. } D_t \longrightarrow D_\tau \text{ a.s.}$$

And by dominated convergence

$$\lim_{t\to\infty} \mathbb{P}(D_t(x) = y) = \mathbb{P}(D_\tau(x) = y) = \mathbb{P}(Y = y)$$

However

$$X_t = f_{\theta_t} \circ \ldots \circ f_{\theta_1}(X_0) \stackrel{d}{=} Y_t = f_{\theta_1} \circ \ldots \circ f_{\theta_t}(X_0)$$

$$\therefore \quad \mathbb{P}(Y = y) = \lim_{t\to\infty} \mathbb{P}(Y_t = y) = \lim_{t\to\infty} \mathbb{P}(X_t = y) = \pi(y)$$

Conclusion: $\mathbb{P}(Y = y) = \pi(y)$!

Coupling from the past

Equivalent viewpoint:
Run several copies of a chain (different initial values) from some time in the past up to time $t = 0$

- Scheduling: $n_1, n_2, \ldots$, increasing integer sequence
  e.g. $1, 2, 4, 8, \ldots$
- Numbers $-n_1, -n_2, \ldots$ are used as starting times
- Sequence $\theta'_0, \theta'_{-1}, \ldots, \theta'_{-n}$ iid sequence $\sim U[0, 1]$

PWA (alternative formulation)

(1) Set $k = 1$

(2) For each $x \in \mathcal{X}$, simulate chain starting at time $-n_k$ in state x & running up to time $t = 0$ with update f & rand. numbers $\theta'_{-n_{k+1}}, \ldots, \theta'_0$ (order is important)

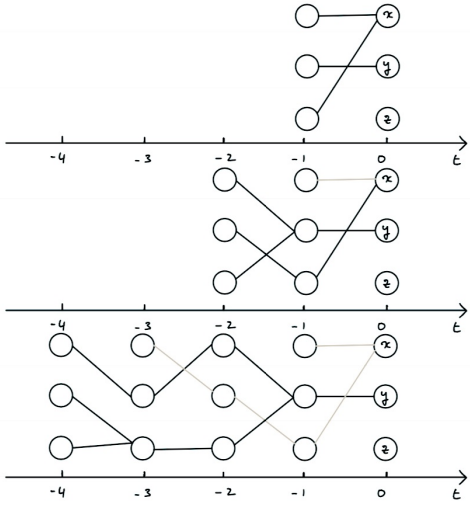
$$X_0^x = f_{\theta'_0} \circ \ldots \circ f_{\theta'_{-n_{k+1}}}(x)$$

(3) If all chains end up in same state y, output y

(4) Otherwise, increase k & go to (2)

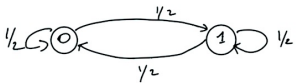
Example

- 3 states $\{x, y, z\}$ & algorithm outputs y

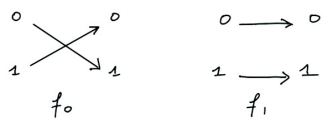


About time of coalescence

- Strong irreducibility does not necessarily imply that $\mathbb{P}(\tau < \infty) = 1$
- Example: $\mathcal{X} = \{0, 1\}$ & $P(x, y) = \frac{1}{2}$ $\forall\, x, y$



$\theta \sim \text{BER}(1/2)$



Then for each t, D_t is either f_0 or $f_1 \implies \mathbb{P}(\tau < \infty) = 0$!

In fact $\mathbb{P}(\tau < \infty) \in \{0, 1\}$ always (Kolmogorov $0 - 1$ law)

Choose f & θ (transition matrix is P) s.t.

$$\forall\ A \subset \mathcal{X} \qquad |A| > 1 \Longrightarrow \mathbb{P}(|f_\theta(A)| < |A|) \geq \alpha > 0$$

Under this hypothesis

$$\forall\ E \subset \mathcal{X} \qquad \mathbb{P}(|f_{\theta_1}(E)| \leq |\mathcal{X}| - 1) \geq \alpha$$

And

$$\begin{aligned}
\mathbb{P}(|f_{\theta_1} \circ f_{\theta_2}(E)| \leq |\mathcal{X}| - 2) &= \sum_{A \subset S} \mathbb{P}(|f_{\theta_1}(A)| \geq |\mathcal{X}| - 2)\mathbb{P}(f_{\theta_2}(E) = A) \\
&\geq \sum_{A \subset S : |A| \leq |\mathcal{X}| - 1} \mathbb{P}(|f_{\theta_1}(A)| \geq |\mathcal{X}| - 2) \quad \text{———} \\
&\geq \alpha^2
\end{aligned}$$

By induction

$$\mathbb{P}(D_t \in \ \mathsf{CST}) \geq \alpha^t > 0 \Leftrightarrow \mathbb{P}(D_t \notin \ \mathsf{CST}) \leq 1 - \alpha^t < 1$$

Thus

$$\mathbb{P}(\tau \geq kn) \leq (1-\alpha^n)^k \underset{k\to\infty}{\longrightarrow} 0 \quad \Longrightarrow \mathbb{P}(\tau < \infty) = 1$$

An important objection

- Major drawback: compute $D_t(x) \quad \forall\, x \in \mathcal{X}$
- This is impossible in practice
  e.g. Ising model: $|\mathcal{X}| = 2^{|V|}$
- Importance of algorithm comes from the fact that in some cases, we do not need to run many chains
- Important spacial cases: chains with monotonicity properties

Monotone systems

Suppose there is a partial order on $\mathcal{X}$ denoted by $\leq$

(i) $x \leq x$ [reflexivity]

(ii) $x \leq y \;\&\; y \leq x \Longrightarrow x = y$ [antisymmetry]

(iii) $x \leq y \;\&\; y \leq z \Longrightarrow x \leq z$ [transitivity]

⚠ We may not be able to compare all elements of $\mathcal{X}$

Definition: Chain is monotone if wp 1 on θ

$$x \leq y \Longrightarrow f_\theta(x) \leq f_\theta(y)$$

Suppose we have two elements $\underline{x}$, $\overline{x}$ for which

$$\underline{x} \leq x \leq \overline{x} \quad \forall\, x \in \mathcal{X}$$

By monotonicity: $\tau = \inf\{t \geq 1 : D_t \in \mathsf{CST}\} = \inf\{t \geq 1 : D_t(\underline{x}) = D_t(\overline{x})\}$

Fundamental example: Ising model (ferromagnetic)

- $x \in \{\pm 1\}^{|V|}$, $\pi(x) = \frac{1}{Z} e^{-\beta E(x)}$, $E(x) = -\sum_{v \sim v'} x_v x_{v'}$

Partial order $x \leq x' \iff x_v \leq x_{v'} \quad \forall\, v \in V$ [check (i) - (iii)]

- Wish to find an update s.t.

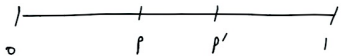
$$x \leq x' \implies f_\theta(x) \leq f_\theta(x')$$

Gibbs sampler

- v at random
- $\theta \sim U[0,1]$
- If $\theta < \frac{1}{1+e^{-2\beta g_v}}$ $\quad x_v = +1$ else $x_v = -1$ $\quad g_v = \sum_{v' : v' \sim v} x_{v'}$

Claim f_θ is monotone

Why? $x \leq x' \Longrightarrow g_v \leq g'_v \Leftrightarrow p = \dfrac{1}{1+e^{-2\beta g_v}} \leq \dfrac{1}{1+e^{-2\beta g'_v}} = p'$



$$\theta \in [0, p) \Longrightarrow x_v = +1,\ x'_v = +1$$
$$\theta \in [p, p') \Longrightarrow x_v = -1,\ x'_v = +1$$
$$\theta \in [p', 1] \Longrightarrow x_v = -1,\ x'_v = -1$$

In all cases $x_v \leq x'_v$

Propp-Wilson algorithm for Ising model

- $\underline{x} = (-1, \ldots, -1), \quad \overline{x} = (1, \ldots, 1)$
- Starting times $-n_1, -n_2, \ldots$
- Sequence θ'_{-n} iid $U[0,1]$

(1) Set $k = 1$

(2) Simulate chain starting at $-n_k$ in state $\underline{x}$ & $\overline{x}$ & running up to time $t = 0$
Random numbers $\theta'_{-n_k+1}, \ldots, \theta'_0$ [in this order]

(3) If two chains end up in same state y, output y

(4) Else increase k & go to (2)

Numerical example

Exact sampling from Ising model (click here)

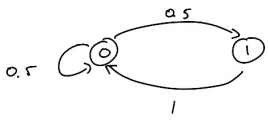
Intuitive variations do not work

A] Coupling into the future?

Run chains forward in time & stop when they all meet

$$\tau = \inf\left\{t \geq 1 : f_{\theta_{t+1}} \circ \ldots \circ f_{\theta_1} \in \text{CST}\right\}$$

Example

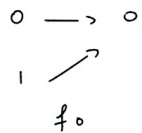


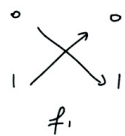
$$\pi = (2/3, 1/3)$$

$$\begin{aligned} \tau - 1 &\Longrightarrow \neq \text{ states} \\ &\Longrightarrow \text{time } \tau \text{ chain have to meet at } 0 \\ &\Longrightarrow Y = 0 \text{ always!} \end{aligned}$$

B⌋ Why reuse same numbers?

- Suppose 'fresh' numbers
- Same example: $n_1 = 1$, $n_2 = 2$, $n_3 = 4$, $n_4 = 8$, $\ldots$





- Two possibilities for f_θ : $\theta \sim \text{BER}(1/2)$
- $\mathbb{P}(Y = 0) = \sum_{j=0}^{\infty} \mathbb{P}(\tau = n_j, Y = 0)$

$$\geq \mathbb{P}(Y = 0 \mid \tau = n_1)\mathbb{P}(\tau = n_1) + \mathbb{P}(Y = 0 \mid \tau = n_2)\mathbb{P}(\tau = n_2)$$

(a) $\mathbb{P}(Y = 0 \mid \tau = 1) = 1$

$\mathbb{P}(\tau = 1) = 1/2$

(b) $\mathbb{P}(\tau = 2) = \mathbb{P}(\tau = 2 \mid \tau > 1)\mathbb{P}(\tau > 1) = \mathbb{P}(\tau = 2 \mid \tau > 1)\frac{1}{2}$

$\mathbb{P}(\tau_2 = 2 \mid \tau > 1) = \mathbb{P}\{f_0 \circ f_1, f_1 \circ f_0, f_0 \circ f_0\} = \frac{3}{4}$

$$\mathbb{P}(Y = 0 \mid \tau = 2) = \frac{2}{3}$$

$$f_0 \circ f_0 \longrightarrow 0$$
$$f_0 \circ f_1 \longrightarrow 0$$
$$f_1 \circ f_0 \longrightarrow 1$$

$\therefore \quad \mathbb{P}(Y = 0) \geq \frac{1}{2} + \frac{2}{3} \times \frac{3}{8} = 3/4 > 2/3$