Lecture 9: Big moves

- ☐ End-of-year agenda items
- ☐ Review of parallel tempering
- ☐ Multilevel Monte Carlo
- ☐ Preconditioned Crank-Nicolson

Recall In parallel tempering, we want to sample $\pi(x)$

- we introduce a vector of MCMC samplers $\vec{X} = (X^{(1)}, \dots, X^{(N)})$ w/ each sampler targeting

$$\pi_i(x) \propto \pi(x)^{\beta_i}$$

where $0 \le \beta_1 < \dots < \beta_N = 1$.

- In addition to standard MCMC moves, we propose swap moves

$$\begin{pmatrix} X^{(1)}_{t+1} \\ \vdots \\ X^{(i)}_{t+1} \\ X^{(i+1)}_{t+1} \\ \vdots \\ X^{(N)}_{t+1} \end{pmatrix} = \begin{pmatrix} X^{(1)}_{t} \\ \vdots \\ X^{(i+1)}_{t} \\ X^{(i)}_{t} \\ \vdots \\ X^{(N)}_{t} \end{pmatrix}$$

← swap these two positions in the vector

- we accept w/ probability

$$\min\left\{1, \left(\frac{\pi(x_+^{(i)})}{\pi(x_+^{(i+1)})}\right)^{\beta_{i+1}-\beta_i}\right\}$$

b/c we are targeting $\pi(\vec{x}) = \prod_{i=1}^{N} \pi_i(x^{(i)})$

and the MH acceptance ratio is

$$\frac{\pi_{i+1}(x_+^{(i)})\,\pi_i(x_+^{(i+1)})}{\pi_{i+1}(x_+^{(i+1)})\,\pi_i(x_+^{(i)})} = \left(\frac{\pi(x_+^{(i)})}{\pi(x_+^{(i+1)})}\right)^{\beta_{i+1}-\beta_i}$$

Ex π is Gaussian $\mathcal{N}(\mu, \Sigma)$ density

$$\pi_i(x) \propto \exp\left(-\frac{\beta_i}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right)$$

$\Rightarrow \pi_i$ is Gaussian $\mathcal{N}(\mu, \frac{1}{\beta_i}\Sigma)$ density

Acceptance rate is

$$\mathbb{E}\left[\left(\pi(x^{(i)})/\pi(x^{(i+1)})\right)^{\beta_{i+1}-\beta_i} \wedge 1 \;\middle|\; \begin{array}{l} x^{(i)} \sim \mathcal{N}(\mu, \frac{1}{\beta_i}\Sigma), \\ x^{(i+1)} \sim \mathcal{N}(\mu, \frac{1}{\beta_{i+1}}\Sigma), \text{ ind} \end{array}\right]$$

$$= \mathbb{E}\left[\left(\frac{\pi(\mu + (\Sigma/\beta_i)^{1/2} z^{(i)})}{\pi(\mu + (\Sigma/\beta_{i+1})^{1/2} z^{(i+1)})}\right)^{\beta_{i+1}-\beta_i} \wedge 1 \;\middle|\; z^{(i)}, z^{(i+1)} \text{ iid } \mathcal{N}(0, I)\right]$$

③

$$= \mathbb{E}\left[\exp\left(\frac{\beta_i - \beta_{i+1}}{2} \left[\frac{1}{\beta_i} \| z^{(i)} \|^2 - \frac{1}{\beta_{i+1}} \| z^{(i+1)} \|^2 \right] \right) \wedge 1 \right]$$

$$= \mathbb{E}\left[\exp\left(\frac{1}{2}\left(1 - \frac{\beta_{i+1}}{\beta_i}\right) \| z^{(i)} \|^2 + \frac{1}{2}\left(1 - \frac{\beta_i}{\beta_{i+1}}\right) \| z^{(i+1)} \|^2 \right) \wedge 1 \right]$$

= constant if we fix the ratio $\dfrac{\beta_{i+1}}{\beta_i}$

$\Rightarrow$ Best option is geometrically spaced temperatures, e.g. $\beta_i = 2^{i-N}$ for $i = 1, \ldots, N$.

Q: How can we calculate $M = \mathbb{E}\left[\max_{0 \le t \le 1} X(t)\right]$ where $X(t)$ is the solution to an SDE

$$dX = b(x)\, dt + \sigma(X)\, dW$$

where $W(t)$ is standard Brownian motion?

- For simplicity, $X(t) = W(t)$ is Brownian motion
- We know we can simulate $W(t)$ at times $t = 0, \Delta, 2\Delta, \ldots, 1-\Delta, 1$ b/c the increments $W(t+\Delta) - W(t)$ are iid $\mathcal{N}(0, \Delta)$ for $t = 0, \Delta, \ldots, 1-\Delta$.
- We can take $\hat{M} = \max_{0 \le i \le 1/\Delta} W(i\Delta)$
- Even better, we can take independent samples $\hat{M}^{(1)}, \ldots, \hat{M}^{(N)}$ and average

$$\hat{M} = \frac{1}{N} \sum_{i=1}^{N} \hat{M}^{(i)}.$$

- This requires $\mathcal{O}(N/\Delta)$ work and the mean square error $\mathbb{E}|\hat{M} - M|^2$

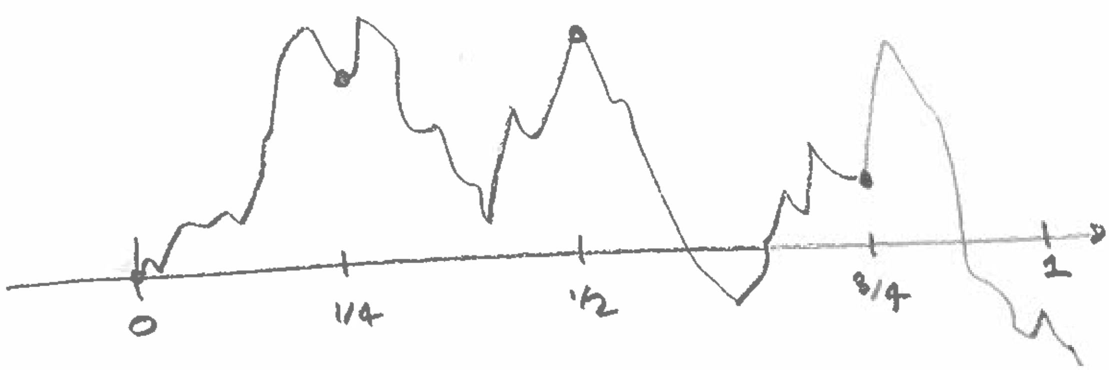
is asymptotically $$\mathcal{O}\Big(\underbrace{\tfrac{1}{\Delta^2}}_{\text{discretization error}} + \underbrace{\tfrac{1}{N}}_{\text{variance due to random sampling}}\Big)$$

as $\Delta \to 0$, $N \to \infty$.

$\Rightarrow$ To obtain accuracy ε, we need

$\Delta = \mathcal{O}(1/\varepsilon)$, $N = \mathcal{O}(1/\varepsilon^2) \Rightarrow \mathcal{O}(1/\varepsilon^3)$ cost!

<u>Q</u> Can we do any better?



<u>Idea</u> Set $\hat{M} = \hat{m}_0 + \sum_{\ell=1}^{L} \hat{\Delta}_\ell$

where $$\hat{M}_0 = \frac{1}{N_0}\sum_{i=1}^{N_0} \max_{0\le j\le 1} W^{(i)}(j)$$

and $$\hat{\Delta}_\ell = \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \underbrace{\left[\max_{0\le j\le 2^\ell} W^{(i)}(j/2^\ell) - \max_{0\le j\le 2^{\ell-1}} W^{(i)}(j/2^{\ell-1})\right]}_{\text{same path, two different timesteps}}$$

⑥

$$\Rightarrow \mathbb{E}\left[M - \hat{M}\right] = \mathbb{E}\left[M - \max_{0 \le j \le 2^L} w\left(j/2^L\right)\right] \le \frac{c_1}{2^L}$$

$$\Rightarrow \mathrm{Var}\left[\hat{M}\right] = \mathrm{Var}\left[\hat{M}_0\right] + \sum_{\ell=1}^{L} \mathrm{Var}\left[\hat{\Delta}_\ell\right]$$

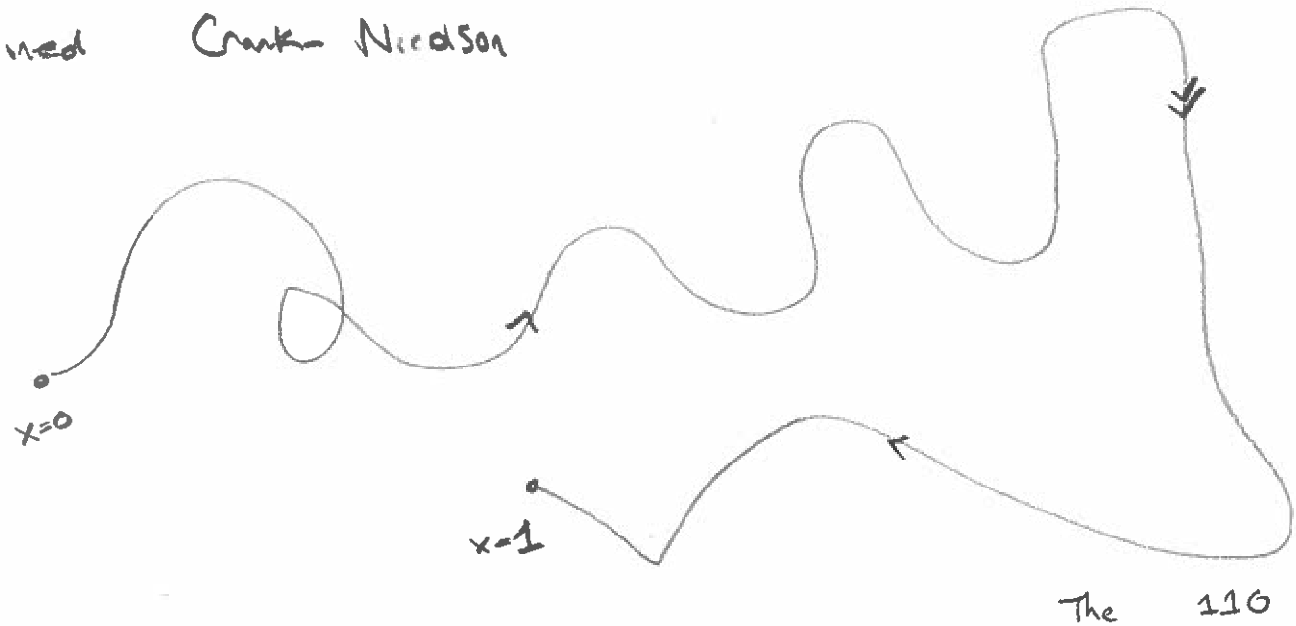
$$\le \sum_{\ell=0}^{L} \frac{c_2}{2^\ell N_\ell}$$

Setting $N_\ell = \Theta\left(\frac{L}{\varepsilon^2 2^\ell}\right)$ and $L = \Theta\left(\log\left(\frac{1}{\varepsilon}\right)\right)$

$\Rightarrow$ MSE $\mathbb{E}\,|M - \hat{M}|^2 \le \varepsilon^2$

Computational cost is $\Theta\left(\frac{\left(\log\left(\frac{1}{\varepsilon}\right)\right)^2}{\varepsilon^2}\right)$!

Preconditioned Crank Nicolson



- Density at time $t=0$ is $u(x)$
- Flow advects

$$(\partial_t + c(x)\partial_x)\, u(t,x) = 0$$

- $u(0,x) = e^{\xi(x)}$, $\underbrace{\xi \sim \mathcal{N}(0,\Sigma)}_{\text{Gaussian process}}$
- $c(x) = e^{\eta(x)}$, $\eta \sim \mathcal{N}(0,\Sigma)$ independent of ξ
- Observations V_i of density at times t_i and locations x_i w/ error $\mathcal{N}(0,\sigma)$

$\Rightarrow$ likelihood ratio $e^{-\Phi(u)}$ where

$$\Phi(u) = \frac{1}{2}\sum_i \left(V_i - u(t_i,x_i)\right)^2$$

requires $u_0(x)$, $c(x)$ and a PDE solve

Want to learn posterior distribution of (ξ, η)

Q How can we solve this problem using MCMC?

Easier Q How can we even solve this problem if we know $C = 1$ exactly?

A $(\partial_t + \partial_x)\, u(t,x) = 0$

$\Rightarrow \frac{d}{dt} v(t) = 0$, where $v(t) = u(t, x+t)$

$\Rightarrow v(t) = v(0) = u(0, x)$

$\Rightarrow u(t,x) = u(0, x-t)$

$\Rightarrow$ Likelihood ratio takes the form

$e^{-\Phi(u_0)}$ w/ $\Phi(u_0) = \frac{1}{2} \sum_i (V_i - u_0(x_i - t_i))^2$

Need to sample $\xi \sim \mathcal{N}(0, \Sigma)$ and weight by $e^{-\Phi(\xi)}$ w/ $\Phi(\xi) = \frac{1}{2} \sum_i \left(V_i - e^{\xi(x_i - t_i)}\right)^2$

Q How to sample?

Random walk method

- Propose $\xi' = \xi + \beta \Delta$, $\Delta \sim \mathcal{N}(0, \Sigma)$
- Accept w/ prob $\min\left\{1, \exp\left(\frac{1}{2}\|\Sigma^{-1/2}\xi\|^2 - \frac{1}{2}\|\Sigma^{-1/2}\xi'\|^2 + \Phi(\xi) - \Phi(\xi')\right)\right\}$

Need to shrink β with the grid refinement in order to maintain an acceptance probability bounded $>$ 0. $\ddot{\frown}$

Preconditioned Crank-Nicolson

- Propose $\xi' = \sqrt{1-\beta^2}\,\xi + \beta\Delta, \quad \Delta \sim \mathcal{N}(0,\Sigma)$
- Accept w/ prob $\min\left\{1, \exp\left(\Phi(\xi) - \Phi(\xi')\right)\right\}$

why?

MH ratio is

$$\frac{\pi(\xi')\,T(\xi',\xi)}{\pi(\xi)\,T(\xi,\xi')}$$

$$= \frac{\exp\left(-\frac{1}{2}\|\Sigma^{-1/2}\xi'\|^2 - \Phi(\xi')\right)\exp\left(-\frac{1}{2}\left\|\Sigma^{-1/2}\left(\frac{\xi - \sqrt{1-\beta^2}\,\xi'}{\beta}\right)\right\|^2\right)}{\exp\left(-\frac{1}{2}\|\Sigma^{-1/2}\xi\|^2 - \Phi(\xi)\right)\exp\Bigg(\underbrace{-\frac{1}{2}\left\|\Sigma^{-1/2}\left(\frac{\xi' - \sqrt{1-\beta^2}\,\xi}{\beta}\right)\right\|^2}_{-\frac{1}{2\beta^2}\|\Sigma^{-1/2}\xi'\|^2 + \frac{\sqrt{1-\beta^2}}{\beta^2}\langle\Sigma^{-1/2}\xi,\,\Sigma^{-1/2}\xi'\rangle - \frac{1-\beta^2}{2\beta^2}\|\Sigma^{-1/2}\xi\|^2}\Bigg)}$$

$$= \exp\left(\Phi(\xi) - \Phi(\xi')\right)$$

<u>Intuition</u> The moves are <u>ALWAYS</u> accepted when $\Phi(\bar{\xi}) = 0$ (the case with zero observations), regardless of β

$\Rightarrow$ This means we obtain a valid function space sampler even as we refine the grid, keeping β fixed.

$\Rightarrow$ With finite observations, we mantain an acceptance rate bounded > 0 with fixed β as we refine the grid.