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

\( X = (X^{(1)}, \ldots, X^{(n)}) \) \( \sim \) each sampler targeting

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

where \( 0 \leq \beta_1 \leq \cdots \leq \beta_n = 1 \).

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

\[
\begin{pmatrix}
X^{(1)}
& X^{(n)}
& \cdots
& X^{(n)}
& X^{(n)}
\end{pmatrix}
= \begin{pmatrix}
X^{(1)}
& X^{(1)}
& \cdots
& X^{(n)}
& X^{(n)}
\end{pmatrix}
\]

\( \leftrightarrow \) swap these two positions in the vector
we accept with probability
\[ \min \left\{ 1, \left( \frac{\prod(X_i)}{\prod(X_i^{(i+2)})} \right)^{\beta_{i+1} - \beta_i} \right\} \]

bc we are targeting \( \prod(X) = \prod_{i=1}^{N} \pi_i(X^{(i)}) \)

and the MH acceptance ratio is
\[ \frac{\prod_{i+1}(X_i^{(i+1)}) \prod_i(X_i^{(i+2)})}{\prod_{i+1}(X_i^{(i+2)}) \prod_i(X_i^{(i)})} = \left( \frac{\prod(X_i)}{\prod(X_i^{(i+1)})} \right)^{\beta_{i+1} - \beta_i} \]

Ex \( \prod \) is Gaussian \( \mathcal{N}(\mu, \Sigma) \) density
\[ \pi_i(x) \propto \exp \left( -\frac{\beta_i}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right) \]

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

Acceptance rate is
\[ E \left[ \left( \frac{\prod(X^{(i)})}{\prod(X^{(i+2)})} \right)^{\beta_{i+1} - \beta_i} \left\{ X^{(i)} \sim \mathcal{N}(\mu, \frac{1}{\beta_i} \Sigma), X^{(i+2)} \sim \mathcal{N}(\mu, \frac{1}{\beta_{i+1}} \Sigma), \text{iid} \right\} \right] \]
\[ = E \left[ \left( \frac{\prod(\mu + (\Sigma/\beta_i)^{1/2} Z^{(i)})}{\prod(\mu + (\Sigma/\beta_{i+1})^{1/2} Z^{(i)})} \right)^{\beta_{i+1} - \beta_i} \left\{ Z^{(i)}, Z^{(i+2)} \text{ iid } \mathcal{N}(0, I) \right\} \right] \]
\[ E \left[ \exp \left( \frac{\beta_i - \beta_{i+1}}{2} \left[ \frac{1}{\beta_i} \| z^{(i)} \|^2 - \frac{1}{\beta_{i+1}} \| z^{(i+1)} \|^2 \right] \right)^{\frac{1}{2}} \right] \]

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

= constant if we fix the ratio \( \frac{\beta_{i+1}}{\beta_i} \)

\[ \Rightarrow \text{Best option is geometrically spaced temperatures, e.g. } \beta_i = 2^{i-N} \text{ for } i=1, \ldots, N. \]
Q: How can we calculate \( M = \max_{0 \leq t \leq 1} X(t) \) 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, \( X(t) = W(t+1) \) is Brownian motion.

- We know we can simulate \( W(t) \) at times \( t=0, \Delta, 2\Delta, \ldots, 1-\Delta, 1 \) where the increments \( W(t+\Delta) - W(t) \) are i.i.d. \( \mathcal{N}(0, \Delta) \) for \( t=0, \Delta, \ldots, 1-\Delta. \)

- We can take \( \hat{M} = \max_{0 \leq i \leq 1/\Delta} W(i\Delta) \)

- 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 \( O(n/\Delta) \) work and
The mean square error \( \mathbb{E}(\hat{M} - M)^2 \) is asymptotically \( \Theta\left(\frac{1}{\Delta^2} + \frac{1}{N}\right) \) due to discretization error and variance due to random sampling as \( \Delta \to 0, \; N \to \infty \).

\( \Rightarrow \) To obtain accuracy \( \varepsilon_j \), we need

\[ \Delta = \Theta(\varepsilon_j^2), \quad N = \Theta(\varepsilon_j^{-2}) \Rightarrow \Theta(\varepsilon_j^{-3}) \text{ cost!} \]

Can we do any better?

\[ \text{Idea: Set } \hat{M} = \hat{M}_0 + \frac{1}{N_e} \sum_{l=1}^{N_e} \Delta l \]

where \( \hat{M}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} \max_j W(i, j) \) and

\[ \Delta l = \frac{1}{N_e} \sum_{i=1}^{N_e} \left[ \max_{0 \leq j \leq 2^l} W(i, j) - \max_{0 \leq j \leq 2^{l-1}} W(i, j) \right] \]
\[ \Rightarrow \mathbb{E}[M - \hat{M}] = \mathbb{E}[M - \max_{0 \leq j \leq 2^L} W(j/2^L)] \leq \frac{C_1}{2^L} \]

\[ \Rightarrow \text{Var}[\hat{M}] = \text{Var}[\hat{M}_o] + \sum_{e=1}^{L} \text{Var}[\Delta_e] \]

\[ \leq \frac{L}{2} \frac{C_2}{\epsilon^2 2^L N_e} \]

Setting \( N_e = \Theta\left(\frac{L}{\epsilon^2 2^L}\right) \) and \( L = \Theta(\log(1/\epsilon)) \)

\[ \Rightarrow \text{MSE} \quad \mathbb{E}|M - \hat{M}|^2 \leq \epsilon^2 \]

Computational cost is \( \Theta\left(\frac{(\log(1/\epsilon))^2}{\epsilon^2}\right) \)!
- Density at time $t=0$ is $u(x)$

- Flow advects

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

- $u(0,x) = \mathcal{N}(x), \quad \phi \sim \mathcal{N}(0,\Sigma)$

  - Gaussian process

- $c(x) = \mathcal{N}(x), \quad \sigma \sim \mathcal{N}(0,\Sigma)$ independent of $\phi$

- Observations $V_i$ of density at times $t_i$ and locations $x_i$ with error $\mathcal{N}(0,\sigma^2)$

  $\Rightarrow$ likelihood ratio $e^{-\phi(x)}$ where

\[\Phi_i(u) = \frac{1}{2} \sum_i \left( V_i - u(x_i) \right)^2\]

  requires $u_0(x), \partial_t c(x)$

Want to learn posterior distribution of $(\phi, \sigma)$
How can we solve this problem using MCMC?

This problem if we know $C = 1$ exactly?

$$\frac{d}{dt} V(t) = 0 \quad \text{where} \quad V(t) = u(t, x+t)$$

$$\Rightarrow \quad V(t) = V(0) = u(0, x)$$

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

$$\Rightarrow \quad \text{Likelihood ratio takes the form}$$

$$e^{\frac{1}{2} \sum_i (V_i - u_0(x_i-t))^2}$$

Need to sample $\mathbf{z} \sim \mathcal{N}(0, I)$ and weight by $e^{-\frac{1}{2} \mathbf{z}^T \mathbf{z}}$ with $\mathbf{z} = \frac{1}{2} \sum_i (V_i - e^{\frac{3}{2} (x_i-t)})^2$

How to sample?

**Random walk method**

- Propose $\mathbf{z}' = \mathbf{z} + \beta \Delta$, $\Delta \sim \mathcal{N}(0, I)$
- Accept with probability $\min\left\{ 1, \exp\left( \frac{1}{2} \| \mathbf{z}' \|^2 - \frac{1}{2} \| \mathbf{z} \|^2 \right)\right\}$
Need to shrink $\beta$ with the grid refinement in order to maintain an acceptance probability bounded $\geq 0$.

Preconditioned Crank-Nicolson

- Propose $\tilde{s}' = \sqrt{1 - \beta^2} \tilde{s} + \beta \Delta$, $\Delta \sim \mathcal{N}(0, \Sigma)$
- Accept with prob $\min\{1, \exp(\overline{\Phi}(\tilde{s}) - \overline{\Phi}(\tilde{s}'))\}$

Why?

MH ratio is:

\[
\frac{\Pi(\tilde{s}') \mathcal{T}(\tilde{s}', \tilde{s})}{\Pi(\tilde{s}) \mathcal{T}(\tilde{s}, \tilde{s}')}
\]

\[
= \frac{\exp\left(-\frac{1}{2} \| \Sigma^{-1/2} \tilde{s}' \|_2^2 - \overline{\Phi}(\tilde{s}')\right) \exp\left(-\frac{1}{2} \| \Sigma^{-1/2} \tilde{s} \|_2^2 - \overline{\Phi}(\tilde{s})\right)}{\exp\left(-\frac{1}{2} \| \Sigma^{-1/2} \tilde{s} \|_2^2 - \overline{\Phi}(\tilde{s})\right) \exp\left(-\frac{1}{2} \| \Sigma^{-1/2} \tilde{s}' \|_2^2 - \overline{\Phi}(\tilde{s}')\right)}
\]

\[
= \exp\left(\frac{1 - \beta^2}{2 \beta^2} \| \Sigma^{-1/2} \tilde{s}' \|_2^2 + \frac{1 - \beta^2}{\beta^2} \langle \Sigma^{-1/2} \tilde{s}', \Sigma^{-1/2} \tilde{s} \rangle - \frac{1 - \beta^2}{2 \beta^2} \| \Sigma^{-1/2} \tilde{s} \|_2^2\right)
\]

\[
= \exp\left(\overline{\Phi}(\tilde{s}) - \overline{\Phi}(\tilde{s}')\right)
\]
Intuition: The moves are always accepted when \( \Phi(\mathbf{s}) = 0 \) (the case with zero observations), regardless of \( \beta \).

This means we obtain a valid function space sampler even as we refine the grid, keeping \( \beta \) fixed.

With finite observations, we maintain an acceptance rate bounded to 0 with fixed \( \beta \) as we refine the grid.