Stats 318: Lecture # 12

Agenda: MCMC

- Simulated tempering
- Parallel tempering

Motivation

- In general, convergence of MCMC algorithm
- But convergence can be very slow!

  If chain is slowly mixing, may take astronomical amount of time

  e.g. Ising model at low temperature
- Mixing can be exponential in number of variables
- Possible improvement: parallel simulations of distributions at different temperatures to move freely in the state space: same spirit as *simulated annealing*
- Two related proposals
  - Simulated tempering
  - Parallel tempering

Simulated tempering (ST)

- Family of distributions $\Pi = \{\pi_i(x) \ : \ i \in I\}$
- i temperature parameter

$$\pi_i(x) \ \propto \ e^{-\frac{E(x)}{T_i}}$$

- Original target distribution π : member of this family with lowest temperature
- Other distributions are used to increase mixing
- Define new target distribution on augmented space $(x, i) \in \mathcal{X} \times I$

$$\pi_{\mathsf{ST}}(x, i) = c(i)\pi_i(x)$$

$c(i) > 0$ controlled by user : $\sum_{i \in I} c(i) = 1$

- Now use an MCMC algorithm to draw samples from π_{ST}

ST algorithm

- Initialize $I_0 = 0$ & $X_0 \in \mathcal{X}$
- Repeat: current state $(X_t, I_t) = (x, i)$
  - (a) Update x using a Metropolis-Hastings or Gibbs sampler s.t. π_i is invariant
  - (b) Set $X_{t+1} = x$ and propose a level transition $i \longrightarrow j$ from a transition matrix $Q(i,j)$ (e.g. random walk with reflecting boundary)
    Draw $U \sim U[0,1]$ and

$$\text{set } I_{t+1} = j \text{ if } U \leq \min(1, r) \qquad r = \frac{\pi_j(x)c(j)Q(j,i)}{\pi_i(x)c(i)Q(i,j)}$$

$$\text{and } I_{t+1} = i \text{ otherwise}$$

Intuition: By heating up the distribution repeatedly, new sampler can escape from local modes and increase its chance of reaching main body of the distribution

Equilibrium distribution

- Equilibrium distribution

$$\pi_{\mathsf{ST}}(x, i) = c(i)\pi(x)$$

Why?

(a) Leaves invariant $\pi_i(x)$ for each i
$\longrightarrow$ Leaves invariant $c(i)\pi_i(x)$

(b) Leaves invariant $c(i)\pi_i(x)$ for each x
$\longrightarrow$ Leaves invariant $c(i)\pi_i(x)$ QED

- Marginal distribution of i in equilibrium is

$$\sum_{x \in \mathcal{X}} c(i)\pi_i(x) = c(i)$$

Hence if $c(i) = \frac{1}{m}$, sampler spends $1/m$ of the time sampling each distribution and there is no temperature that is not visited frequently

Practical considerations

- Even in an ideal situation when sampler behaves like a symmetric RW along the temperature scale, expected time before traversing all distributions is $\Omega(m^2)$
  $\implies$ Limits number of distributions one can use

- For ST to work well, two adjacent distributions need to have significant overlap
  Otherwise, practically impossible to accept temperature transitions

- Given sequence of temperatures $T_1 < \ldots < T_m$, sim. temp. simulates from m tempered dist. $\propto e^{-\frac{1}{T_i}E(x)}$ simultaneously

- Hope is to borrow information across $\neq$ chains to achieve better sampling result

- Final remark: we need to know normalizing constant Z_{T_i} !

Parallel tempering (PT)

- Instead of augmenting $\mathcal{X}$ as in $\mathcal{X} \times I$, consider several copies of $\mathcal{X}$: $\mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_m$
- Family of distributions $\Pi = \{\pi_i\}_{i=1,\ldots,m}$
- Joint prob. dist. on product space

$$\pi_{\mathsf{PT}}(x_1, \ldots, x_m) = \prod_{i=1}^{m} \pi_i(x_i)$$

- Auxiliary distributions

$$\pi_i(x) = \frac{1}{Z_i} e^{-\frac{E(x)}{T_i}} \qquad \text{temperature scale : } T_1 < T_2 < \ldots < T_m$$

- Run parallel MCMC chains on all the $\mathcal{X}_i$'s
- Swap index from time to time

PT algorithm

PT step: Current state $x_1, \ldots, x_m$

(a) Parallel step: Update every x_i with their respective MCMC scheme

(b) Swapping step: Randomly choose a neighboring pair; say i & $i+1$ and propose swapping x_i & x_{i+1}. Accept swap wp

$$\min\left(1,\ \frac{\pi_i(x_{i+1})\pi_{i+1}(x_i)}{\pi_i(x_i)\pi_{i+1}(x_{i+1})}\right)$$

Remarks:

- Can utilize information in multiple MCMC chains
- Does not need fine tuning (i.e. knowledge of normalizing constants)
- Applied to simulate complex systems; e.g. polymers & other similar molecular structures

Invariance of $\pi_{\mathsf{PT}}(\ .\ )$

- π_{PT} invariant wrt parallel step
- Swapping step reversible wrt π_{PT} [for each $(i, i+1)$]

$$x = (x_1, \ldots, x_m) \longrightarrow x' = (x_1, \ldots, x_{i-1}, x_{i+1}, x_i, x_{i+2}, \ldots, x_m)$$

$$\begin{aligned}
\pi_{\mathsf{PT}}(x) P^{(i)}(x, x') &= \prod_j \pi_j(x_j) \min\left(1,\ \frac{\pi_{i+1}(x_i)\pi_i(x_{i+1})}{\pi_i(x_i)\pi_{i+1}(x_{i+1})}\right) \\
&= \prod_{j : j \notin \{i, i+1\}} \pi_j(x_j) \quad \min\left(\pi_i(x_i)\pi_{i+1}(x_{i+1}),\ \pi_{i+1}(x_i)\pi_i(x_{i+1})\right) \\
&= \pi_{\mathsf{PT}}(x') P^{(i)}(x, x')
\end{aligned}$$

- Swapping step reversible $\Longrightarrow \pi_1^{(t)} \longrightarrow \pi_1$ as $t \longrightarrow \infty$

How to choose the temperature scale?

- Swap acceptance rate r

$$\begin{aligned}\frac{\pi_i(x_{i+1})\pi_{i+1}(x_i)}{\pi_i(x_i)\pi_{i+1}(x_{i+1})} &= \exp\left(-\frac{1}{T_i}\left[E(x_{i+1}) - E(x_i)\right] - \frac{1}{T_{i+1}}\left[E(x_i) - E(x_{x_i+1})\right]\right) \\ &= \exp\left(-\left(\frac{1}{T_i} - \frac{1}{T_{i+1}}\right)\triangle E\right)\end{aligned}$$

- Rough guideline is to choose T_i's s.t.

$$\left(\frac{1}{T_i} - \frac{1}{T_{i+1}}\right)\triangle E = \underbrace{-\log p_a}_{\text{lower bound on targeted acceptance rate}}$$

$|\triangle E|$ "typical" energy difference
e.g. mean energy change under target distribution