Adaptive MCMC

Stochastic gradient:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, f(\theta) = \mathbb{E}_{\pi}\left(F(\theta, x)\right)$$

<u>gradient</u>: $\theta^{(k)} = \theta^{(k-1)} - \gamma \nabla f(\theta^{(k)})$

<u>stochastic</u> $X \sim \pi$

$$\theta^{(k)} = \theta^{(k-1)} - \gamma F(\theta^{(k)}, X)$$

But if $f(\theta) = \int F(x, \theta) \pi_\theta(x)\, dx$

$$\Rightarrow \begin{cases} X^{(k)} \mid \theta^{(k-1)}, X^{(k-1)} \sim \pi_{\theta^{(k-1)}} \\ \theta^{(k)} = \theta^{(k-1)} - \gamma F(\theta^{(k-1)}, X^{(k)}) \end{cases}$$

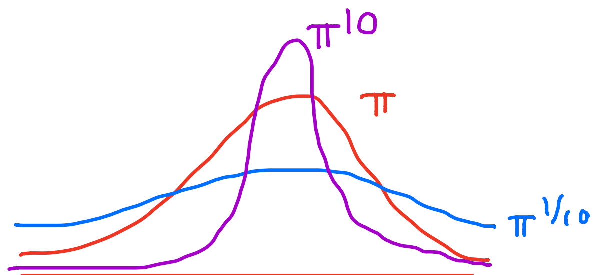
<u>For adaptive MCMC</u>: sample from π using $\{P_\theta\}$

Given $\theta^{(k-1)}, X^{(k-1)}$:

$$X^{(k)} \sim P_{\theta^{(k-1)}}\left(X^{(k-1)}, \cdot\right)$$

$$\theta^{(k)} \in \mathcal{F}\left(\theta^{(k-1)}, X^{(k)}\right)$$

Tempering

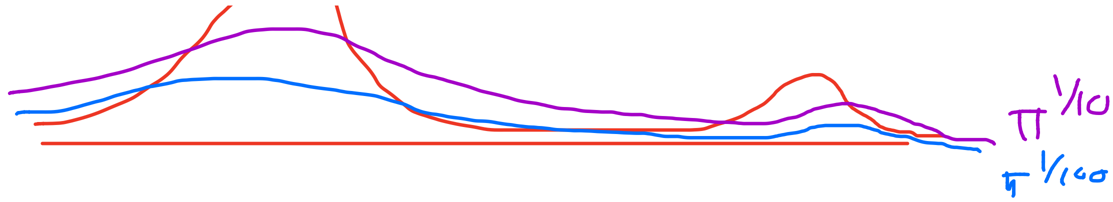


Ex: $f(x) = (2\pi)^{-1/2} \exp(-x^2/2) \longrightarrow N(0,1)$

$$f^{1/10}(x) \propto \exp\left(-x^2/20\right) \longrightarrow N(0,10)$$

$$f^{10}(x) \propto \exp\left(-x^2/2\right) \longrightarrow N(0, 1/10)$$

- higher powers less variability
- lower powers more variability



Apply tempering / inject variance.

much easier to explore related distributions $\pi^{1/100}$, $\pi^{1/10}$

$$\pi \implies \pi^{1/T}$$

$$\begin{cases} T > 1 & \text{“flatter” than } \pi \\ T < 1 & \text{“spikier” than } \pi \end{cases}$$

<u>Exercise</u>: Fix π a probability density

Define $\pi_t \propto \pi^{1/t}$.

Show that $\pi_t \xrightarrow[t \to 0]{w} U_\Omega$

$$\Omega = \underset{x}{\operatorname{argmax}}\, \pi(x)$$

Simulated Annealing

Let $U: \mathcal{X} \longrightarrow \mathbb{R}$ be a function

s.t. $\int \exp(-U(x))\,dx < +\infty$.

We want

$$\underset{x}{\operatorname{argmin}}\ U(x)$$

Choose $t_1 > t_2 > \cdots > t_n > 0$ define

$$\pi_i(x) \propto \exp(-U(x)/t_i)$$

Let P_i be a markov kernel such that

$$\pi_i P_i = \pi_i$$

Alg: Given $x^{(t)}$

$$X^{(k+1)} \sim P_{k+1}(x^{(k)}, \cdot)$$

This method changes the target dist. (by a factor of $1/t_i$) to attain the solutions to

$$\underset{x}{\operatorname{argmin}}\; U(x)$$

Simulated Tempering

Goal: Sampling from π on $\mathbb{R}^p$

Suppose $\pi(x) = \dfrac{e^{-U(x)}}{Z}$ ← normalizing constant.

$$\pi_t(x) = \frac{e^{-U(x)}}{Z_t} \leftarrow \text{normalizing constant.}$$

Z, Z_t unknown.

Fix $1 = t_1 < t_2 < t_3 < \cdots < T_K$ temperatures.

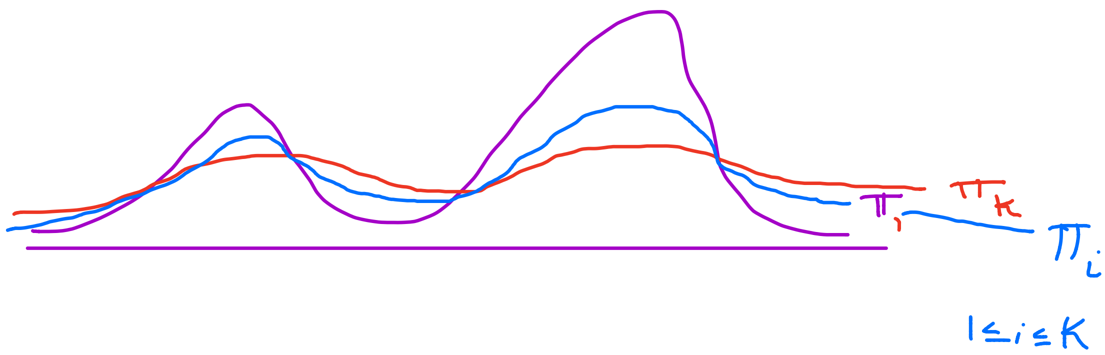
Set $\pi_k(x) = \pi_{t_k}(x) \propto e^{-\frac{U(x)}{t_k}}$

Consider

$$\bar{\pi}(k, x) = \frac{\frac{1}{c_k} e^{-U(x)/t_k}}{\sum_{k=1}^{K} \frac{1}{c_k} \int e^{-U(x)/t_k} dx}$$

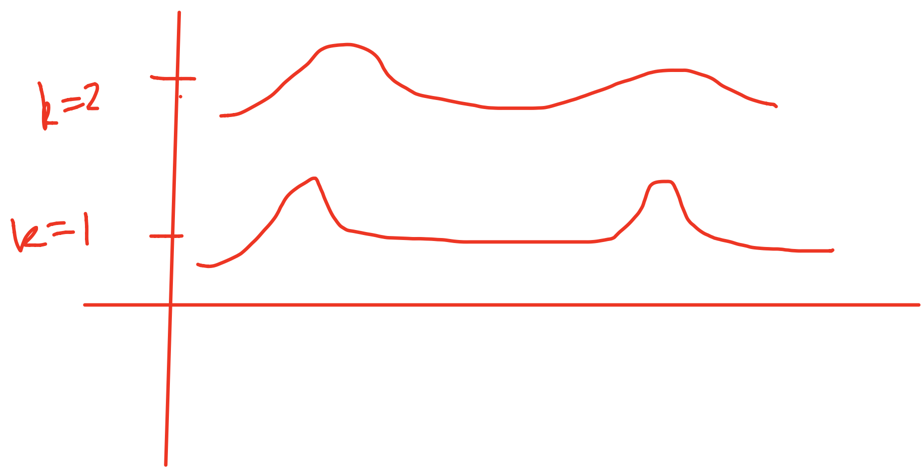
$\{c_k\}$ user chosen

$\bar{\pi}$ is a density on $\{1, 2, \ldots, K\} \times \mathbb{R}^p$.



<u>Idea</u>: Fixing k we can sample from the joint $X \sim \bar{\pi}(\cdot k, x)$

Then given X we can sample the new distribution (k).

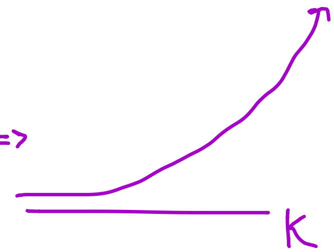


<u>Q</u>: What is the marginal $\bar{\pi}(k)$?

$$\bar{\pi}(k) = \frac{\frac{1}{c_k}\int e^{-u(x)/t_k}\,dx}{\sum \frac{1}{c_i}\int e^{-u(x)/t_i}\,dx}$$

If $\quad c \equiv 1$

$$\pi(k) = \frac{\int e^{-u(x)/t_k}\, dx}{\sum_i \int e^{-u(x)/t_i}\, dx} \implies$$



So the M.C. will spend most of its time on the most tempered distributions.