Consensus Monte Carlo

Scott et. al (2013)

C. Li et al. (2017)

Consider the Hierachical Model

$$\begin{tikzcd}
 & & \mu,\Sigma \arrow[dll, no head] \arrow[dl, no head] \arrow[dr, no head] & \\
\beta_1 \arrow[d, no head] & \beta_2 \arrow[d, no head] & \cdots & \beta_n \arrow[d, no head] \\
x_1 & x_2 & & x_n
\end{tikzcd}$$

<u>data:</u> $x_1 \quad x_2 \quad x_n$ $\qquad x_i \sim f_{\beta_i} | \beta_i \quad \beta_i \sim f_{\mu,\Sigma}$

$\pi(\beta_1, \beta_2, \dots, \beta_n, \mu, \Sigma \mid X_{1:n})$

Idea send $\beta_1, \dots, \beta_{50}$ to the first core

$\beta_{51}, \dots, \beta_{100}$ to the second

$\vdots \qquad \vdots$

<u>Rmk:</u> Parallel Gibbs with communication is still inefficient.
with respect to the # of cores

Consensus Monte Carlo

Take $\bar{x} = x_1, \dots, x_n$ and split into $\bar{x}_j = x_{j,1}, \dots, x_{j,m}$ for $j \in [k]$ s.t. $mk = n$.

<u>model:</u> $f_\theta(x)$

<u>posterior:</u> $\pi_j(\theta | \bar{x}_j) \propto p(\theta)^{1/k} \prod_{s=1}^{m} f_\theta(x_{js})$

<u>prior:</u> $p(\theta)$

$\bar{\pi}(\theta | \bar{x}) \propto p(\theta) \prod_{s=1}^{n} f_\theta(x_i)$

$$\propto p(\theta) \prod_{k=1}^{K} f_\theta(X_{k,1:m})$$

$$\propto \prod_{k=1}^{K} p(\theta)^{1/k} f_\theta(\bar{X}_k)$$

$$= \prod_{k=1}^{K} \pi_k(\theta \mid \bar{X}_k).$$

Clearly the assumption here is the small subsetting of the data.

<u>Question</u>: Can we sample from $\bar{\pi}(\cdot \mid \bar{X})$ given samples from $\pi_j(\cdot \mid \bar{X}_j)$.

Main idea: let $f_i \sim N(\mu_i, C_i)$ $\quad i=1,2$ and consider the density

$$\bar{f}(x) \propto f_1(x) f_2(x) \propto \exp\left[-\frac{1}{2}(x-\mu_1)^T C_1^{-1}(x-\mu_1) - \frac{1}{2}(x-\mu_2)^T C_2^{-1}(x-\mu_2)\right]$$

$$\propto \exp\left\{-\frac{1}{2}(x-\bar{\mu})^T \bar{C}^{-1}(x-\mu)\right]$$

$$\bar{C}^{-1} = (C_1^{-1} + \bar{C}_2^{-1})^{-1} \quad \bar{\mu} = (C_1^{-1} + C_2^{-1})^{-1}[C_1^{-1}\mu_1 + C_2^{-1}\mu_2]$$

Furthermore if $X_1 \sim N(\mu_1, C_1)$ $\quad X_2 \sim N(\mu_2, C_2)$

$$\bar{X} = (C_1^{-1} + C_2^{-1})^{-1}[C_1^{-1}X_1 + C_2^{-1}X_2] \sim N(\bar{\mu}, \bar{C})$$

<u>Consensus MC:</u>

(*) In parallel: draw $\{\theta_k^{(s)}; s=1,\dots,S\}$ for each $k \in [K]$

- MCMC sample from $\pi_k(\cdot \mid \bar{X}_k)$

(**) For $s=1,\dots,S$ combine $\theta_1^{(s)} \dots \theta_K^{(s)}$

$$\theta^{(s)} = \left(\sum_{k=1}^{K} \hat{C}_k^{-1}\right)\left(\sum_{k=1}^{K} \hat{C}_k^{-1} \theta_k^{(s)}\right)$$

when $\hat{C}_k$ is covariance of $\overline{\pi}_k(\cdot|\overline{x}_k)$

<u>Simple, Accurate, & Scalable Post. Interval Estimation</u>

Setup: $\overline{X} = X_{1:n}$, split $\overline{x}_1, \ldots, \overline{x}_K$ $\quad x_k = x_{k,1:m}$

$$\overline{\pi}_k(\theta|\overline{x}_k) \propto p(\theta) \underbrace{\left[\prod_{s=1}^{m} f_\theta(x_{k,s})\right]^K}_{\text{data cloning}}$$

Q: How to use samples from $\overline{\pi}_k(\cdot|\overline{x}_k)$ to sample from $\pi(\cdot|\overline{x})$

<u>Idea:</u> take average? <u>Q:</u> How?

Take barycente in a metric space

$$P_2(\Theta) = \left\{\mu \text{ on } \Theta : \int \|x\|_2^2 \mu(dx) < \infty\right\}$$

$$\|\mu - \nu\|_{wass} = \left\{\inf_{\gamma \in \Gamma(\mu,\nu)} \int \|x-y\|_{wass.}^2 \gamma(dx, dy)\right\}^{1/2} \overset{def}{=} W_2(\mu, \nu)$$

$\Gamma(\mu_1, \mu_2)$ set of all measures on $\Theta \times \Theta$ with marginals (μ_1, μ_2).

Then the average is given by

$$\tilde{\pi}(\cdot|\overline{x}) \overset{def}{=} \underset{\mu}{\operatorname{argmin}} \sum_{k=1}^{K} W_2^2(\mu, \overline{\pi}(\cdot|\overline{x}))$$

Remember given $x_1, \ldots, x_n$ $\quad \overline{x} = \underset{b \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (x_i - b)^2$

In $\mathbb{R}$: $W_2(\mu,\nu) = \left(\int |F_\mu^{-1}(t) - F_\nu^{-1}(t)|^2 dt\right)^{1/2} = \| F_\mu^{-1} - F_\nu^{-1} \|_2$

$$\overset{?}{\neq} \left(\int | F_\mu(x) - F_\nu(x)|^2 dx\right)^{1/2}$$

As $\|\cdot\|_2$ has an associated inner product,

$$\bar{\pi}(\cdot \mid \bar{x}) = \left(\frac{1}{K} \sum_{k=1}^{K} \pi_j^{-1}(\cdot \mid \bar{x}_j)\right)^{-1}$$