Monte Carlo and coupling

▷ Monte Carlo

Suppose we want to approximate $\mathbb{E}_\pi f$, one way is to sample $X_1, \cdots X_n \overset{iid}{\sim} \pi$ and use

$$\sum_{i=1}^{n} f(X_i) \Big/ n \quad \text{as our estimator.}$$

LLN: $\sum_{i=1}^{n} f(X_i) \Big/ n \xrightarrow{a.s.} \mathbb{E}_\pi f$

(More samples become more accurate)

CLT: $\sqrt{n}\left(\frac{\sum_{i=1}^{n} f(X_i)}{n} - \mathbb{E}_\pi f\right) \to N(0, \operatorname{Var} f(X_1))$

(Convergence rate $= \frac{1}{\sqrt{n}}$)

Curse of dimensionality:

Def $e(n,f) := \sqrt{\mathbb{E}\left[\left(\frac{\sum_{i=1}^{n} f(X_i)}{n} - \mathbb{E}_\pi f\right)^2\right]}$ be the rooted Mean squared error (RMSE), then

$$e(n,f) = \sqrt{\frac{\operatorname{Var} f(X_1)}{n}}$$

~~If we fix f, then the rate is $1/\sqrt{n}$.~~

Therefore if we want

$$e(n,f) \le \varepsilon$$

we can choose $n \ge \mathrm{Var} f(X_1) \cdot \varepsilon^{-2}$.

▷ Dependency on ε does not depend on the dimensionality of f's domain!

For other deterministic methods, typically we need to partition each dimension into pieces with width ε, with costs $\Omega\left(\left(\frac{1}{\varepsilon}\right)^d\right)$.

▷ MCMC

Sampling from π can be difficult, thus we implement a MC $(X_0, X_1, \cdots X_n, \cdots)$ and estimate $\mathbb{E}_\pi f$ by

$$\sum_{i=1}^{n} f(X_i) \Big/ n.$$

MC LLN $\quad \dfrac{\sum_{i=1}^{n} f(X_i)}{n} \xrightarrow{a.s.} \mathbb{E}_\pi f$

MC CLT $\quad \sqrt{n}\left(\dfrac{\sum_{i=1}^{n} f(X_i)}{n} - \mathbb{E}_\pi f\right) \to N(0, \sigma_f^2)$

σ_f^2 is called the asymptotic variance.

One difference between iid and MCMC is that MCMC is generally biased, i.e.

$$\mathbb{E}\left[\frac{\sum_{i=1}^{n} f(X_i)}{n}\right] \text{ typically } \neq \mathbb{E}_{\pi}[f(X_1)]$$

, unless we start from stationary.

We introduce an algorithm that eliminates the bias.

Suppose we have a coupling that is faithful,

we start two chains $Y_0 \sim \mu_0$, $Y_1 \sim P(Y_0, \cdot)$

$Z_0 \sim \mu_0$

Then we apply our coupling on (Y_1, Z_0) to get (Y_2, Z_1)

then (Y_3, Z_2), ...

We have $$\mathbb{E}_{\pi}[f] = \lim_{n\to\infty} \mathbb{E}_{\mu_0}[f(X_n)]$$

$$= \sum_{n=1}^{\infty} \mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y_{n-1})] + \mathbb{E}[f(Y$$

$$= \sum_{n=1}^{\infty} \mathbb{E}[f(Y_n)] - \mathbb{E}[f(Z_{n-1})] + \mathbb{E}[f(Y_0$$

$$= \mathbb{E}\left[\sum_{n=1}^{\tau} f(Y_n) - f(Z_{n-1})\right] + \mathbb{E}[f(Y_0)]$$

This suggest the following estimator

① Run the algorithm until two chain meets

② calculate $f(Y_0) + \sum_{n=1}^{\tau} f(Y_n) - f(Z_{n-1})$

Benefits : ① Remove the bias

② Implement on many processers

Applications : ① Bayesian inference / learning

② MCMC convergence diagonis

③ Asymptotic variance estimation