MCMC → Markov Chain Monte Carlo

① Monte Carlo Algo ⟵ Guarantee by $\begin{cases} \text{SLLN} \to \text{convergence guarantee} \\ \text{CLT} \to \text{convergence rate} \end{cases}$

a) SLLN $X_1, \dots, X_n$ i.i.d distributed

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} E[X]$$

⇒ If we want to calculate $\int_0^1 f(x)dx$

1. $X_i \sim \text{Uniform}(0,1)$

then $\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{a.s.} E[f(X)] = \int_0^1 f(x)\cdot 1\, dx$

2. $X_i \sim$ Some distribution $p(x)$

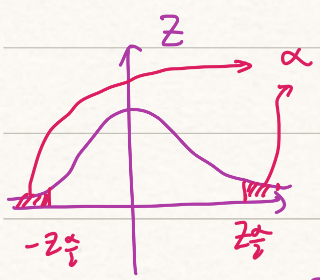
then $\frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{p(X_i)} \xrightarrow{a.s.} E\left[\frac{f(x)}{p(x)}\right] = \int_0^1 \frac{f(x)}{p(x)} dP(X \le x)$

$= \int_0^1 f(x)\, dx$

b) CLT denote $\boxed{f(x_i) \begin{cases} \to \mu \\ \to \sigma^2 \end{cases}}$, then $\frac{\sum_{i=1}^{n} f(x_i) - n\mu}{\sqrt{n\sigma^2}} \sim Z(0,1)$

$\Rightarrow \frac{\sum_{i=1}^{n} f(x_i)}{n} - \mu \approx \sqrt{\frac{\sigma^2}{n}}\, Z(0,1)$

$\frac{\sum_{i=1}^{n} f(x_i)}{n} \to \boxed{S_n}$



$\frac{S_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx Z(0,1)$

⇒ $\frac{1}{n}\sum f(x_i)$ converges to $E[f(x)]$

$\Rightarrow P\left(S_n - \sqrt{\frac{\sigma^2}{n}}\cdot Z_{\frac{\alpha}{2}} \le \mu \le S_n + \sqrt{\frac{\sigma^2}{n}}\, Z_{\frac{\alpha}{2}}\right) = 1-\alpha$ with rate $\sqrt{\frac{1}{n}}$

⇒ interval length $= 2\sqrt{\frac{\sigma^2}{n}}\cdot Z_{\frac{\alpha}{2}}$ will decrease at rate $\sqrt{\frac{1}{n}}$ for fixed confidence α

---

② Question: How to generate SAMPLE for a given $p(\underline{x})$?

a) inverse method b) reject sampling c) Box-Muller

Answer: for simple distribution $p(\underline{x})$, we can generate sampling !

(Example) Uniform $\xrightarrow{a}$ Any distribution with closed form (computationally expensive)

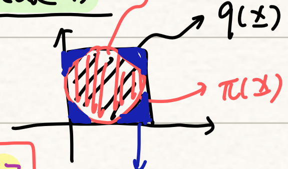
$\xrightarrow{b}$ Any distribution with closed form (waste)

$\xrightarrow{c}$ ☆ Gaussian (directly from uniform sampling)

Gaussian → { chi-square, t, F }

an extreme case i):

accept w.p.1



accept w.p.0

[Reject Sampling]

	r.v	pdf
target distribution	X	$\sim\!\to \pi$
proposal distribution	Y	$\sim\!\to q$
Uniform distribution	u	

accept $\Leftrightarrow$ $u \le \dfrac{\pi(X_i)}{Mq(X_i)}$

☆

流程

① Sample y_i from $q(\cdot)$

② sample u_i from Unif $(0,1)$

③ accept y_i

$\Leftrightarrow$ $u_i \le \dfrac{\pi(y_i)}{Mq(y_i)}$

we want to show : $p(x|\text{accept}) = \pi(x)$

① $p(x|\text{accept}) = \dfrac{p(x,\text{accept})}{\int p(x,\text{accept})\,dx}$

② $p(x,\text{accept}) = \int p(\text{accept}|x,u)\cdot p(u|x)\cdot q(x)\,du$

$= q(x)\cdot \int p(\text{accept}|x,u)\,p(u)\,du$

$$p(\text{accept}|x,u) = \begin{cases} 1, & u \le \dfrac{\pi(x)}{Mq(x)} \\ 0, & o/w \end{cases}$$

$= q(x)\cdot \int_0^{\frac{\pi(x)}{Mq(x)}} 1\,du$

$= \dfrac{\pi(x)}{M}$

③ $p(x|\text{accept}) = \dfrac{\frac{\pi(x)}{M}}{\frac{1}{M}\int \pi(x)\,dx} = \pi(x)$

③ MCMC

1) Reject $\boxed{\text{Sampling}}$ limits:

→ for some high-dimension distribution, it is difficult to achieve $q(x)$ & M

need to solve

→ waste samples which are rejected

→ but compared with Numerical Integration, Monte Carlo Integration works well in high-dimension cases

2) MC result

{ recurrent MC, stationary measure uniquely exits [e.g.]

{ $\boxed{\text{transient MC}}$, stationary measure is not unique → asymmetric random walk on $\mathbb{Z}$

*) for irreducible, positive recurrent MC $\quad \Pi^n(x,y) \to \mathbb{P}(X_n=y) \ \forall x$

$$\mathbb{P}(X_n=y) \longrightarrow \mu(y)$$

**) for irreducible, null recurrent / transient MC $\quad \Pi^n(x,y) \to 0$

$$\mathbb{P}(X_n=y) \to 0$$

***) [Reversible Measure] (detailed balance condition)

$\mu(x)\,\Pi(x,y) = \mu(y)\,\Pi(y,x)$ if $\mu(\cdot)$, $\Pi(\cdot,\cdot)$ satisfy

☆ then $\mu(\cdot)$ must be a stationary measure

⟹ Our aim is to construct a MC with long-term probability $\mu(\cdot)$

{ 1) irreducible, recurrent

{ 2) $\Pi(\cdot,\cdot)$ satisfy detailed balance condition for the given $\mu(\cdot)$.

⇒ then sample $X_1, \dots, \underbrace{X_N, \dots, X_{N+n}}_{\text{can be used as the sampling from } \mu(\cdot)}$ from MC

3) M-H Algorithm (Metropolis-Hastings Algorithm)

$\boxed{\text{irreducible}}$ ← origin trans accept → $\boxed{\text{aperiodic}}$

a) Construction : $\Pi(x,y) = Q(x,y)\,\alpha(x,y)$

$$\mu(x)\,\Pi(x,y) = \mu(y)\,\Pi(y,x)$$

$$\Longleftrightarrow \quad \mu(x)\,Q(x,y)\,\alpha(x,y) = \mu(y)\,Q(y,x)\,\alpha(y,x)$$

$$\Longleftarrow \quad \begin{cases} \alpha(x,y) = Q(y,x)\cdot\mu(y)\cdot C \\ \alpha(y,x) = Q(x,y)\,\mu(x)\cdot C \end{cases} \quad \text{(natural choice)} \quad \forall C.$$

$$\Longrightarrow \quad \begin{cases} \alpha^*(x,y) = \min\left\{1, \dfrac{Q(y,x)\cdot\mu(y)}{Q(x,y)\,\mu(x)}\right\} \\ \alpha^*(y,x) = \min\left\{\dfrac{Q(x,y)\,\mu(x)}{Q(y,x)\,\mu(y)}, 1\right\} \end{cases} \quad \text{(maximize } \alpha(\cdot,\cdot)\text{)}$$

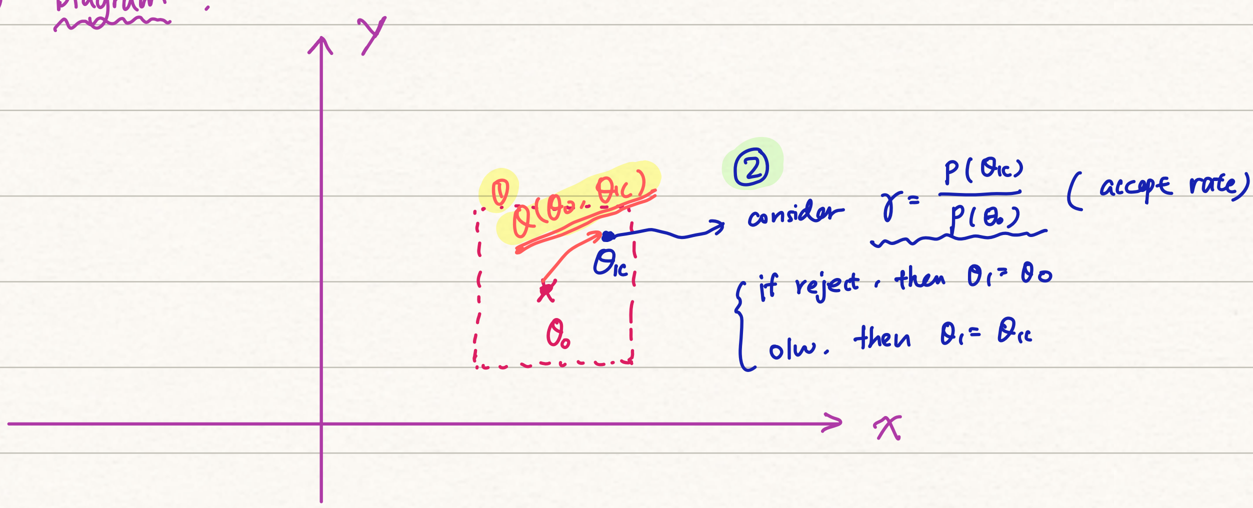
b) limitation:

1) difficult to compute $\boxed{\dfrac{Q(y,x)\,\mu(y)}{Q(x,y)\,\mu(x)}}$ term when $x \in \mathbb{R}^d$ d big

2) our interest may have good form of $p(x_t \mid x_{-t})$

Conditional probability given other coordinates

c) Diagram :



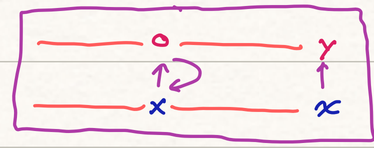
4) Gibbs Sampling

focus on high-dimension case

① idea: a) conditional probability can be easier to deal with

b) still use $\mu(x)\,\Pi(x,y) = \mu(y)\,\Pi(y,x)$

$\boxed{Q(x,y)\ \alpha(x,y)}$



② Formulation

$\underset{\sim}{x} = (x^1, \dots, x^d)$ $\qquad$ $\underset{\sim}{x}_n = (x_n^1, \dots, x_n^d)$

choose $\quad \Pi(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{1}{d} \boxed{\mu_{x^i | x^{-i}}(y^i | x^{-i})} \rightarrow \boxed{\dfrac{\mu(y)}{\sum_{x^i \in x} \mu(x)}}$

$\underset{\sim}{x} = (x^1, \dots, x^d)$

$\underset{\sim}{y} = (y^1, \dots, y^d)$ $\quad$ where there exists $i \in [d]$ s.t.

$x^i \neq y^i$, and others all the same

obviously, $\mu(x)\, \Pi(x,y) = \mu(y)\, \Pi(y,x)$

$\boxed{\text{Note}}$ this can be viewed as the special case of M-H Algo:

Recall: M-H Algo : $\quad \Pi(x,y) = Q(x,y)\, \alpha(x,y)$

$$\alpha(x,y) = \min\left\{\frac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)},\ 1\right\}$$

Here, $Q(x,y) = \frac{1}{d} \mu_{x^i | x^{-i}}(y^i | x^{-i})$

and $\mu(y)\, Q(y,x) = \mu(x)\, Q(x,y) \Rightarrow \alpha(x,y) \equiv 1 \quad \forall x, y$

③ Diagram