DSA5202 Lec 10 $\longrightarrow$ $\boxed{\text{Sampling Method}}$

Recap : Bayesian Approach on Uncertainty Quantification

$$1. \quad \underbrace{P(\beta \mid \mathcal{D})}_{\boxed{\text{posterior}}} \propto \underbrace{P(\mathcal{D} \mid \beta)}_{\boxed{\text{likelihood}}} \cdot \underbrace{P(\beta)}_{\boxed{\text{prior}}}$$

2. our interest is :

Here : $y = X\beta + \boxed{\varepsilon}$ $\rightarrow$ randomness

① $y \mid x_{new}, \mathcal{D}$

$$P(y \mid x_{new}, \mathcal{D}) \propto \int_\beta P(y, \beta \mid x_{new}, \mathcal{D})\, d\beta$$

$$= \int_\beta P(y \mid \beta, x_{new}) \cdot P(\beta \mid \mathcal{D})\, d\beta$$

$\rightarrow$ utilize all $P(\beta \mid \mathcal{D})$ information

Bayesian LR example : $P(y \mid x_{new}, \mathcal{D}) = N(\mu_p^T x_{new}, \beta^{-1} + x_{new}^T \Sigma_p x_{new})$

$\Downarrow$

$\boxed{\text{gives confidence interval}}$

② More common case :

$\Longrightarrow$ Model of prediction : $y = f_\beta(x)$

then, our interest is : distribution of our prediction $\boxed{f_\beta(x_{new}) \mid \mathcal{D}}$

$\longrightarrow$ Bayesian LR example : $f_\beta(x) = \beta^T x$

$\Rightarrow f_\beta(x_{new}) \mid \mathcal{D} = x_{new}^T \beta \mid \mathcal{D} \sim N(x_{new}^T \mu_p, x_{new}^T \Sigma_p x_{new})$

→ However, in most cases:

we have no access to exact distribution for: $\boxed{f_\beta(x_{new}) \mid \mathcal{D}}$

$\Rightarrow$ cannot compute $\begin{cases} \mathbb{E}[\cdot] \\ Var[\cdot] \end{cases}$ directly

☆

→ $\boxed{\text{what we have}}$ is: $p(\beta \mid \mathcal{D}) \propto p(\mathcal{D} \mid \beta) \cdot p(\beta)$

$\Rightarrow$ $\boxed{\text{if we can sampling from } p(\beta \mid \mathcal{D}),}$ $\Rightarrow$ Assumption

then
$$\boxed{\begin{cases} \mathbb{E}[f_\beta(x_{new}) \mid \mathcal{D}] \approx \hat{f}_m = \frac{1}{m} \sum_{i=1}^{m} f_{\beta^i}(x_{new}) \\ Var[f_\beta(x_{new}) \mid \mathcal{D}] \approx \frac{1}{m} \sum_{i=1}^{m} \left(f_{\beta^i}(x_{new}) - \hat{f}_m \right)^2 \end{cases}}$$

↳ Monte Carlo Approximation $[\beta^i \sim p(\beta \mid \mathcal{D}) \quad i \in [m]]$

↳ $\boxed{\text{Approximation of Confidence Interval}}$

Question: How to sample from $p(\beta \mid \mathcal{D})$?

↓

$\boxed{\text{Sampling Method}}$

→ generally speaking, it is hard to sample from:

$$\boxed{\underbrace{\boxed{p(\beta \mid \mathcal{D})}}_{\text{posterior density}} = \propto \frac{\boxed{p(\mathcal{D} \mid \beta) \cdot p(\beta)}}{\underbrace{\int_\beta p(\mathcal{D} \mid \beta) \cdot p(\beta) \, d\beta}_{\text{expensive to compute}}}}$$

$\Rightarrow$ we only have this term

Today's Lecture : sample from $\pi(\beta) := p(\beta \mid \mathcal{D})$

$$\Rightarrow \pi(\beta) \propto p(\mathcal{D} \mid \beta) \cdot p(\beta)$$

1. The distribution we can sample directly :

① Bernoulli $B \sim \mathrm{Bern}(p)$

② Uniform $U \sim \mathrm{Unif}(0, 1)$

③ Gaussian $N \sim N(0, \mathbb{1}_{d \times d})$

④ Linear Transform
$$\begin{cases} a \circ U + b \sim \prod_{i=1}^{d} \mathrm{Unif}(b_i, b_i + a_i) \\ AN + b \sim N(b, AA^T) \\ X_1 B + X_2 (1 - B) \sim \pi_1 \cdot p + \pi_2 (1-p) \end{cases}$$

→ However, it is insufficient for us to sample generally from $X \sim \pi$ based on the RV we can directly sample!

2. Reject Sampling Technique

→ goal : sample from $X \sim \pi \propto \boxed{\tilde{\pi}}$ → we only focus on this

→ approach : ① $q \Rightarrow$ the distribution we can sample from

② assume there exists C s.t

$\tilde{\pi}(x) \le C \cdot q(x)$ for all $x \in$ domain

③ $x_1', \ldots\ldots, x_n' \sim q$

accept x_i' with probability $\dfrac{\tilde{\pi}(x_i')}{C \cdot q(x_i')}$

implementation

Sample $U \sim \mathrm{Unif}(0,1)$.

$\begin{cases} \text{if } U \le \dfrac{\tilde{\pi}(x_i')}{C \cdot q(x_i')}, & \text{accept } x_i' \\ \text{else}, & \text{reject } x_i' \end{cases}$

☆ ④ Final samples : $x_1, x_2, \ldots, x_k$

$\boxed{x_i \sim \pi}$

i-th accepted sample

Claim : $\boxed{X \mid X \text{ is accepted} \sim \pi}$.

Here $X \sim q$

Pf :

① CDF of $X \mid X$ is accepted

$$\begin{aligned}
&\mathbb{P}(X \le x \mid X \text{ is accepted}) \\
&= \frac{P(X \text{ is accepted} \mid X \le x)\, \mathbb{P}(X \le x)}{P(X \text{ is accepted})} \quad \to \text{const} \\
&= \frac{\int_{-\infty}^{x} \frac{\pi(u)}{c \cdot q(u)} \frac{q(u)}{\mathbb{P}(X \le x)}\, du \cdot \mathbb{P}(X \le x)}{\int_{-\infty}^{+\infty} \frac{\pi(x)}{c \cdot q(x)} q(x)\, dx} \\
&= \int_{-\infty}^{x} \pi(u)\, du
\end{aligned}$$

② PDF of $X \mid X$ is accepted

$$\begin{aligned}
&\mathbb{P}(x \le X \le x + \Delta x \mid X \text{ is accepted}) \approx p(x \mid X \text{ is accepted})\, \Delta x \\
&\propto \mathbb{P}(X \text{ is accepted} \mid x \le X \le x + \Delta x) \cdot \mathbb{P}(x \le X \le x + \Delta x) \\
&\approx \frac{\pi(x)}{c \cdot \cancel{q(x)}} \cdot \cancel{q(x)}\, \Delta x \\
&\Rightarrow p(x \mid X \text{ is accepted}) \propto \pi(x)
\end{aligned}$$

$$\mathbb{E}\left[\frac{\#\text{ of accepted samples}}{\#\text{ of proposed samples}}\right] = \frac{1}{c}$$

Note : $\longrightarrow \boxed{\mathbb{P}(X \text{ is accepted}) = c^{-1}}$ if $X \sim \pi$ is normalized

$$\longrightarrow P(X \text{ is accepted}) = \frac{M}{c}$$

Normalization constant

Assume that $\tilde{\pi}(x) \propto \pi(x) \iff \tilde{\pi}(x) = M \cdot \pi(x)$

Now, if we have $p(\beta \mid \mathcal{D}) \propto p(\mathcal{D} \mid \beta) \cdot p(\beta)$

and our interest is : $\mathbb{E}[f_\beta(x_{new}) \mid \mathcal{D}]$

or $Var[f_\beta(x_{new}) \mid \mathcal{D}]$

$\boxed{\beta^1, \beta^2, \dots, \beta^n \sim \pi \propto p(\beta \mid \mathcal{D})} \longrightarrow$ reject sampling

then we can achieve the Monte Carlo Approximation via rejection sampling technique

---

Rmk : 1. "proposed dist." $q \longrightarrow$ support domain D_q

"target dist." $\pi \longrightarrow$ support domain D_π

we must have $D_q \supseteq D_\pi$

2. $\mathbb{E}\left[\dfrac{\#\text{ of accepted}}{\#\text{ of proposed}}\right] = c^{-1}$ if $\pi(\cdot)$ is normalized

Pf sketch : suppose we sample N times,

then ① $(\#\text{ of proposed}) = N$

② $(\#\text{ of accepted}) = \sum_{i=1}^{N} \mathbb{1}\left\{u_i \le \dfrac{\pi(x_i)}{c\cdot q(x_i)}\right\}$

$$\rightarrow \mathbb{E}_{x,u}\left[\mathbb{1}\left\{u \le \frac{\pi(x)}{c\cdot q(x)}\right\}\right]$$

$$= \mathbb{E}_x\left[\mathbb{E}_u\left[\mathbb{1}\left\{u \le \frac{\pi(x)}{c\cdot q(x)}\right\} \,\middle|\, x\right]\right]$$

$$= \mathbb{E}_x\left[\frac{\pi(x)}{c\cdot q(x)}\right] = \int_x \frac{\pi(x)}{c\cdot q(x)}\cdot q(x)\,dx = \frac{1}{c}$$

$$\Rightarrow \mathbb{E}\left[\frac{\#\text{ of accepted}}{\#\text{ of proposed}}\right] = \frac{②}{①} = \frac{1}{c}$$

$\Rightarrow$ Trade-off (w.r.t c).

- $c \rightarrow$ infinity $\rightarrow$
  - guarantee $c\cdot q(x) \ge \pi(x)$
  - accepted rate $= \frac{1}{c} \rightarrow 0$
- $c \rightarrow$ small $\rightarrow$
  - hard to guarantee $c\cdot q(x) \ge \pi(x)$
  - accepted rate $\frac{1}{c}$ big

$\Rightarrow$ Therefore : optimal $\hat{c} = \max_x \dfrac{\pi(x)}{q(x)}$ for $\boxed{\text{given proposal } q}$

$\Downarrow$

$\rightarrow$ Generally speaking, hard to solve the maximize problem

(actually we can choose the best proposal distribution q)

Intuitively, c measures the difference between proposed dist. q and target dist π $\Rightarrow$ bigger c means more different

3. Curse of dimensionality

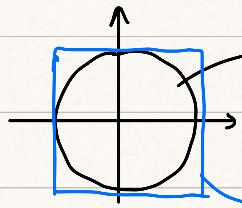
[e.g.] consider $\pi(x) := \dfrac{\mathbb{1}\{\|x\|_2 \le 1\}}{V_p}$ $\quad \boxed{x \in \mathbb{R}^p}$

$$\boxed{V_p = \int_{\mathbb{R}^p} \mathbb{1}\{\|x\|_2 \le 1\}\, dx}$$

Question: ① How can we estimate V_p ?

② How to apply reject sampling technique ?

→ Reject Sampling Framework :



target $\pi(x) = \dfrac{1}{V_p} \cdot \mathbb{1}\{\|x\|_2 \le 1\}$

proposal $q(x) = \dfrac{1}{2^p} \mathbb{1}\{x \in [-1,1]^p\}$

$\Rightarrow$ select c : $\quad c \cdot q(x) \ge \pi(x) \quad x \in \mathbb{R}^p$

$$\Leftrightarrow \quad c \cdot \frac{1}{2^p} \mathbb{1}\{x \in [-1,1]^p\} \ge \frac{1}{V_p} \mathbb{1}\{x \in B_1^2\}$$

$$\Leftrightarrow \quad c \cdot \frac{1}{2^p} \ge \frac{1}{V_p}$$

Two choice of c : ① $\Rightarrow$ $\hat{c} = \dfrac{2^p}{V^p}$ if we know V_p exactly

② $\Rightarrow$ $V_p \ge 2^{\frac{1}{2}p}$

$\Rightarrow$ one choice of c is $\boxed{c = \dfrac{2^p}{2^{\frac{1}{2}p}} = 2^{\frac{1}{2}p}}$

$\boxed{\text{sufficient condition}}$

find the upper bound of $\dfrac{1}{V_p}$

Curse of dimensionality : as $p \to +\infty$, $c \to +\infty$ exponentially

reweighted estimate schema

3. $\boxed{\text{Importance sampling}}$ [Normalized]

not generate the target dist.

but approximate expectation with weights

① $\boxed{\text{target : } \pi}$ $\longrightarrow$ $\boxed{\text{Normalized}}$

② proposal : q

③ if our interest is : $\mathbb{E}_{\beta \sim \pi} [f_\beta (x_{new})]$

then we estimate it through $\frac{1}{n} \sum_{i=1}^{n} \frac{\pi(\beta^i)}{q(\beta^i)} f_{\beta^i} (x_{new})$

Here, $\beta^i \sim q \quad i = 1, 2, 3, \dots, n$

Calculation : $\mathbb{E}_{\beta \sim q} \left[\frac{\pi(\beta)}{q(\beta)} \cdot f_\beta (x_{new}) \right]$

$\boxed{\text{unbiased estimator}}$

$$= \int_x \frac{\pi(x)}{q(x)} \cdot f_x (x_{new}) \cdot q(x) \, dx$$

$$= \int_x f_x (x_{new}) \cdot \pi(x) \, dx$$

$$= \mathbb{E}_{\beta \sim \pi} [f_\beta (x_{new})]$$

$\boxed{\text{importance}}$ of i-th sample $\beta^i \sim q$

From SLLN, $\frac{1}{n} \sum_{i=1}^{n} \boxed{\frac{\pi(\beta^i)}{q(\beta^i)}} f_{\beta^i} (x_{new}) \xrightarrow{a.s.} \mathbb{E}_{\beta \sim q} \left[\frac{\pi(\beta)}{q(\beta)} \cdot f_\beta (x_{new}) \right]$

$$= \mathbb{E}_{\beta \sim \pi} [f_\beta (x_{new})]$$

$\boxed{\text{importance - sampling estimator}}$

$\boxed{\text{support of } q \supseteq \text{support of } \pi}$

$\rightarrow$ one condition of proposal

4. Importance Sampling [Un-normalized]

normalized constant for target $\tilde{\pi}$

① target : $\tilde{\pi}(x) \propto \pi(x) \longrightarrow \tilde{\pi}(x) = C \cdot \pi(x)$

② proposal : $q(x)$

③ interest is still : $\boxed{\mathbb{E}_{\beta\sim\pi}[f_\beta(x_{new})]}$ unbiased

Now, estimator is : $\boxed{\dfrac{\sum_{i=1}^{n} w_i f_{\beta^i}(x_{new})}{\sum_{i=1}^{n} w_i}}$ $\beta^i \sim q \quad i \in [n]$

Here, $w_i = \dfrac{\boxed{\tilde{\pi}(\beta^i)}}{q(\beta^i)}$ → un-normalized target

Note :

$$\left[\begin{array}{l} \frac{1}{n}\sum_{i=1}^{n} w_i \approx C \\ \frac{1}{n}\sum_{i=1}^{n} w_i f_{\beta^i}(x_{new}) \approx C \cdot \underbrace{\mathbb{E}_{\beta\sim\pi}[f_\beta(x_{new})]}_{\text{our interest}} \end{array}\right.$$

5. Effective Sampling Size (ESS)

→ For Standard MC $\frac{1}{n}\sum_{i=1}^{n} f(x_i)$ $x_i \sim \pi(\cdot)$

reject sampling

the variance of estimator $= \frac{1}{n} \mathrm{var}_{X\sim\pi}[f(X)]$

n → $\boxed{ESS}$

→ For Importance Sampling MC $\frac{1}{n}\sum_{i=1}^{n} w(x_i)\cdot f(x_i)$ $x_i \sim q$.

the variance of estimator $= \frac{1}{n} \mathrm{Var}_{x\sim q}[w(x) f(x)]$

$\boxed{\text{Assume : } f(x) \le M_f \text{ for all } x}$

$$\le \frac{1}{\dfrac{n}{\mathbb{E}_{x\sim q}[w(x)^2]}} \cdot M_f$$

$\dfrac{n}{\mathbb{E}_{x\sim q}[w(x)^2]}$ → $\boxed{ESS}$

Rmk : 1. $\dfrac{\text{ESS}}{\text{\# of proposals}} = \dfrac{1}{\boxed{\mathbb{E}[\omega(x)^2]}}$

plays the role of "$\boxed{C}$" in

$\boxed{\text{reject sampling}}$ framework

$\boxed{\text{proposal}} \approx \boxed{\text{target}}$

$\Rightarrow$ if $q(x) \approx \pi(x)$, then $\mathbb{E}[\omega(x)^2] \approx 1$

$\Downarrow$

$\boxed{\text{large ESS}}$

$\Downarrow$

$\boxed{\text{small variance}}$

$\boxed{\text{very different}}$

$\Rightarrow$ if $q(x) \neq \pi(x)$, then $\mathbb{E}[\omega(x)^2] >> 1$

since $\boxed{\omega(x) = \dfrac{\pi(x)}{q(x)}}$ will have

some large value

$\Rightarrow$ $\boxed{\text{small ESS}}$ $\Rightarrow$ large variance