Lecture 10: Sampling

Soufiane Hayou

Friday 9^{th} June, 2023

Recall: Monte Carlo

How to compute

$$\mathbb{E}[f(\beta)|S] = \int f(\beta)p(\beta|S)d\beta?$$

Generate samples $\beta^i \sim \pi(\beta) = p(\beta|S)$, use

$$\hat{f}_m = \frac{1}{m}\sum_{i=1}^{m} f(\beta^i)$$

- $\hat{f}_m$ is an unbiased estimator
- If β^i are i.i.d. from p, then

$$\mathrm{var}(\hat{f}_m) = \frac{1}{m}\mathrm{var}(f)$$

- We want to sample the posterior density

$$\pi(x) = \frac{p_0(x)p(S|x)dx}{\int p_0(x)p(S|x)dx}$$

- How do we generate samples from $\pi(x)$?

- We want to sample the posterior density

$$\pi(x) = \frac{p_0(x)p(S|x)dx}{\int p_0(x)p(S|x)dx}$$

- How do we generate samples from $\pi(x)$?
- Additional challenge: $\int p_0(x)p(S|x)dx$ is not known
- We only have $\pi(x) \propto p_0(x)p(S|x)$
- Reference "Machine Learning: A Bayesian and Optimization Perspective"

Simple distributions

- Bernoulli: $\pi(B=1) = q, \pi(B=0) = 1-q$ (np.random.binomial)
- Uniform: $\pi(u) = 1_{u \in [0,1]}$ (np.random.rand)
- Gaussian: $\pi(z) = \frac{1}{\sqrt{2\pi}^d} \exp(-\frac{1}{2}|z|^2)$ (np.random.randn).

Linear transformation

- $\mathbf{a} \circ U + \mathbf{b} \sim \prod_{i=1}^{d} U[b_i, b_i + a_i]$
- $AZ + \mathbf{b} \sim \mathcal{N}(\mathbf{b}, AA^T)$
- $X_1 B + X_2(1-B) \sim q\pi_1 + (1-q)\pi_2$

Rejection sampling

Rejection Sampling

- We want to sample $\pi(x)$
- Assume that there exists $c > 0$ and a function q such that $c\,q(x) \geq \pi(x)$
- Proposal from q: $x'_1, \ldots, x'_n$
- Accept x'_i with probability $\frac{\pi(x_i)}{cq(x_i)}$
- Bayes formula $\mathbb{E}[f(x'_i)|x'_i \text{ is accepted}] = \mathbb{E}_\pi f(x)$.

Input: Number of samples N

Output: N samples X_i where each follows density $\pi(x)$.

1. For $i = 1, \cdots, N$ do step 2-6.
2. Flag=1;   %whether the i-th sample has been done
3. While (Flag) do step 4-6
4.   Sample X' from $q(x)$.
5.   Sample U from uniform $[0, 1]$.
6.   If $U < \frac{\pi(X')}{cq(X')}$, set $X_i = X'$ and Flag=0.

Properties of rejection sampling

- The support of π should be inside the support of q.
- The accepted samples can be seen as from density π.
- Efficiency:
$$\mathbb{E}\frac{\#\text{ accepted samples}}{\#\text{ proposal samples}} = \frac{1}{c}$$
- We want to use proposal densities similar to π
- c in general increase exponentially with d

Example

Consider the uniform distribution on unit-ball:

$$\pi(x) = \frac{1_{\|x\| \leq 1}}{V_p}$$

Formulate the rejection sampling method with proposal being uniform in $[-1, 1]^p$. Can you find a way to estimate V_p?

Comic from SMBC.

- Proposal from q: $x_1, \ldots, x_n$
- Assign weights $w_i = \frac{\pi(x_i)}{q(x_i)}$
- Estimator: $\frac{1}{n}\sum_{i=1}^{n} w_i f(x_i)$
- Justification:

$$\mathbb{E}_q \frac{\pi(X_i)}{q(X_i)} f(X_i) = \int \frac{\pi(x)}{q(x)} f(x) q(x) dx = \mathbb{E}_\pi f(X).$$

Importance Sampling

- Sometimes we only know $\pi(x) = Cg(x)$ with unknown C
- Proposal from q: $x_1, \ldots, x_n$
- Assign weights $w_i = \frac{g(x_i)}{q(x_i)}$
- C can be approximated as $(\frac{1}{n}\sum_{i=1}^{n} w_i)^{-1}$
- Estimator: $\frac{\sum_{i=1}^{n} w_i f(x_i)}{\sum_{i=1} w_i}$

Effective sample size

- Suppose sample size is n, $|f(x)| \leq M_f$.
- Variance of $f(x) \leq M_f^2$.
- Standard Monte Carlo variance $\leq \frac{1}{n} M_f^2$
- Importance sampling single sample variance is

$$\mathrm{var}(w(x)f(x)) \leq \mathbb{E}w(x)^2 f(x)^2 \leq M_f^2 \mathbb{E}w(x)^2.$$

- Importance MC variance $\leq \frac{1}{n} M_f^2 \mathbb{E}w(x)^2$
- The effective sample size is $n/(\mathbb{E}w(x)^2)$

Estimating effective sample size

- How to estimate $n/(\mathbb{E}w(x)^2) = n/(\mathbb{E}(\frac{\pi(x)}{q(x)})^2)$?
- Note that $\frac{w_i}{\frac{1}{n}\sum_{i=1}^n w_i} \approx \frac{\pi(x_i)}{q(x_i)}$.
- The estimated effective sample size is

$$\hat{n} \approx \frac{(\sum w_i)^2}{\sum w_i^2}$$

- If $w_i \equiv 1$, $\hat{n} = n$.

Example

Consider the uniform distribution on unit-ball:

$$\pi(x) = \frac{1_{\|x\| \leq 1}}{V_p}$$

Formulate the importance sampling method with proposal being uniform in $[-1, 1]^p$.

Final Test ($\sim$ Exam)

- Final Test: 14/06/2023, 19:00 - 21:00 (2 hours)
- Format: Online Quizz (General Questions + Problems)
- Contents: Everything we have seen!