

Winter 2020 Math 126
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 7: Monte Carlo

7.1 Monte Carlo Integration

For a function $f(x)$ which is integrable over $[0, 1]^d$, we want to calculate the mean value of f

$$I[f] = \int_{[0,1]^d} f(x) dx = \bar{f}.$$

Setting: We assume that we know how to evaluate $f(x)$ but there is not simple formula for the antiderivative of $f(x)$.

- ▶ If we use a grid-based methods, the **convergence rate** is $\mathcal{O}(n^{-k/d})$ where k is the order of the grid-based method.
- ▶ The Monte Carlo integration draws a sample $\{x_i\}$ from the inform distribution on $[0, 1]^d$ and estimate the integral

$$I[f] \approx \hat{I}_n[f] = \frac{1}{n} \sum_i f(x_i).$$

- ▶ The convergence rate of the Monte Carlo integration is $\mathcal{O}(n^{-1/2})$.

7.1 Monte Carlo Integration

- ▶ The probabilistic interpretation of $I[f]$ is that $I[f]$ is an expected value of $f(x)$ where x has the uniform density in $[0, 1]^d$

$$I[f] = E[f] = \int_{[0,1]^d} f(x) dx$$

- ▶ From the law of large numbers,

$$\hat{I}_n[f] \rightarrow I[f].$$

- ▶ Also, $\hat{I}_n[f]$ is unbiased

$$E[\hat{I}_n[f]] = I[f].$$

Exercise Prove that $\hat{I}_n[f]$ is unbiased.

7.1 Monte Carlo Integration

- ▶ Let $e_n[f]$ be the error of the Monte Carlo estimator

$$e_n[f] = I_n[f] - I[f].$$

- ▶ From the Central limit theorem, for a large n , we have

$$e_n[f] \approx \sigma n^{-1/2} \nu \quad (1)$$

where ν is a standard normal random variable and the constant σ^2 is the variance of f , that is,

$$\sigma^2 = \left(\frac{1}{n} \int (f - I[f])^2 \right).$$

Exercise Prove (1) (you need to show the variance of $\hat{I}_n[f]$ first).

7.1 Monte Carlo Integration

Now we are interested in

$$I[fp] = \int_{[0,1]^d} f(x)p(x)dx$$

where $p(x) \geq 0$, $\int_{[0,1]^d} p(x)dx = 1$.

There are two approaches for this problem

1. Draw a sample $\{x_i\}$ of size n from the uniform density of $[0, 1]^d$ and

$$I[fp] \approx \frac{1}{n} \sum_i f(x_i)p(x_i).$$

2. Or draw a sample $\{x_i\}$ of size n from the density $p(x)$ and

$$I[fp] = I_p[f] \approx \frac{1}{n} \sum_i f(x_i).$$

7.1 Monte Carlo Integration

How do you decide which method to use?

7.1 Monte Carlo Integration

How do you decide which method to use?

Check the variances

$$\sigma_1^2 = \int (fp - I[fp])^2 dx$$

and

$$\sigma_2^2 = \int (f - I_p[f])^2 p dx.$$

Choose the method with a smaller variance.

7.2 Sampling Methods

Now we are concerned with a sampling method to generate a sample from a given density $p(x)$.

- ▶ Transformation method
- ▶ Acceptance-rejection method

7.2.1 Transformation method

Let Y be a uniform random variable and look for a transformation $X = f(Y)$ such that the density of X is $p(x)$.

Example. Cauchy density $p(x) = \frac{1}{\pi(1+x^2)}$.

Example. Gaussian density $p(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-(x)^2/2}$

7.2.1 Transformation method

Let Y be a uniform random variable and look for a transformation $X = f(Y)$ such that the density of X is $p(x)$.

Example. Cauchy density $p(x) = \frac{1}{\pi(1+x^2)}$.

Example. Gaussian density $p(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-(x)^2/2}$

$$\begin{aligned} P_X(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \\ &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}(x/\sqrt{2}) \end{aligned}$$

where $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, the error function.

$$\begin{aligned} y = P_Y(y) = P_X(x) &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}(x/\sqrt{2}) \\ x &= \sqrt{2} \operatorname{erf}^{-1}(2y - 1). \end{aligned}$$

7.2.1 Transformation method

Let Y be a uniform random variable and look for a transformation $X = f(Y)$ such that the density of X is $p(x)$.

Example. Cauchy density $p(x) = \frac{1}{\pi(1+x^2)}$.

Example. Gaussian density $p(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-(x)^2/2}$

$$\begin{aligned} P_X(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \\ &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}(x/\sqrt{2}) \end{aligned}$$

where $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$, the error function.

$$\begin{aligned} y = P_Y(y) = P_X(x) &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}(x/\sqrt{2}) \\ x &= \sqrt{2} \operatorname{erf}^{-1}(2y - 1). \end{aligned}$$

Another method: Box-Muller method.

7.2.2 Acceptance-rejection method

For a given density $p(x)$, suppose that we know a function $q(x)$ satisfying

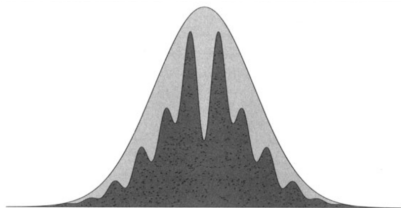
$$q(x) \geq p(x),$$

and that we have a way to sample from the density

$$\tilde{q}(x) = q(x)/I[q].$$

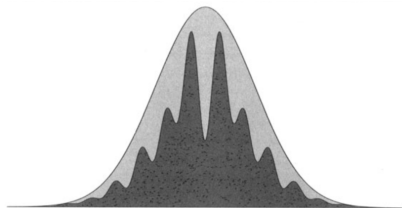
- ▶ Pick two random variables, x' and y , in which x' is a trial variable chosen according to $\tilde{q}(x')$, and y is a decision variable chosen according to the uniform density on $0 < y < 1$.
- ▶ Accept if $0 < y < p(x')/q(x')$
- ▶ Reject if $p(x')/q(x') < y < 1$.

7.2.2 Acceptance-rejection method



* black: density of interest, $p(x)$, * gray: Gaussian, $q(x)$

7.2.2 Acceptance-rejection method



* black: density of interest, $p(x)$, * gray: Gaussian, $q(x)$

Idea of Proof.

$$\begin{aligned} p(x) &= \frac{p(x)}{q(x)} \tilde{q}(x) I[q] \\ &= \int_0^1 I\left(\frac{p(x)}{q(x)} > y\right) dy \tilde{q}(x) I[q] \end{aligned}$$

where $I(\frac{p(x)}{q(x)} > y) = 1$ if $\frac{p(x)}{q(x)} > y$ and 0 otherwise. So,

$$\int f(x) p(x) dx = \int \int_0^1 f(x) I\left(\frac{p(x)}{q(x)} > y\right) dy \tilde{q}(x) I[q] dx$$

7.3 Accuracy and Improvements

In the Monte Carlo integration of $I[fp] = \int fpdx$, the error e and the number n of samples are related by

$$e = \mathcal{O}(\sigma n^{-1/2}),$$

$$n = \mathcal{O}((\sigma/e)^2).$$

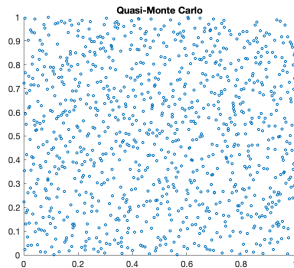
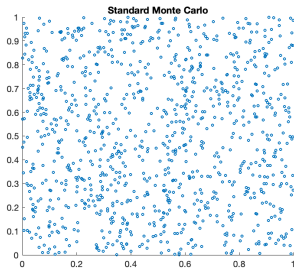
where σ is the variance of f (or fp).

There are two approaches to improve the accuracy

- ▶ Increase the convergence rate
- ▶ or decrease the variance, i.e., variance reduction.

7.3.1 Quasi-Monte Carlo

- ▶ A deterministic sequence, not random
- ▶ Maintains uniformity
- ▶ Convergence rate $\mathcal{O}((\ln n)^d n^{-1})$.



Standard Monte Carlo (left) and Quasi-Monte Carlo (right)
samples of the same size 1000.

7.3.2 Variance Reduction

Antithetic variates method For a sample value x where m is a mean of $p(x)$, also use the value $x' = m - (x - m)$.

That is, if $\{x_i\}$ is a sample of size n ,

$$I[fp] = I_p[f] \approx \frac{1}{2n} \sum_i^n (f(x_i) + f(m - (x_i - m))).$$

Motivation. If the standard deviation of $p(x)$, say std_p , is small,

$$f(x) = f(m) + f'(m)std_p\tilde{x} + \mathcal{O}(std_p^2)$$

where $\tilde{x} = \frac{x - m}{std_p}$.

7.3.2 Variance Reduction

Control variates method If there is a function $g(x)$ such that g is similar to f and $I_p[g] = \int g(x)p(x)dx$ is known,

$$\int f(x)p(x)dx = \int (f(x) - g(x))p(x)dx + \int g(x)p(x)dx.$$

That is, the control variates is effective if the variance of $(f - g)$ is smaller than the variance of $f(x)$.

One may try the following idea to reduce the variance further.

Introduce a multiplier λ

$$\int f(x)p(x)dx = \int (f(x) - \lambda g(x))p(x)dx + \lambda \int g(x)p(x)dx.$$

Use λ minimizing the variance of $f - \lambda g$.

7.3.2 Variance Reduction

Matching moments method Let m_1 and m_2 be the first and the second moments of $p(x)$. Also let α_1 and α_2 are the first and the second sample moments of a sample $\{x_i\}$.

Then, instead of $\{x_i\}$, use the following transformed sample $\{y_i\}$ that preserves the correct moments up to the second order

$$y_i = (x_i - \alpha_1)c + m_1$$

where $c = \sqrt{\frac{m_2 - m_1^2}{\alpha_2 - \alpha_1^2}}$.

Exercise. Show that the first two sample moments of $\{y_i\}$ are equal to the true moments m_1 and m_2 .

7.3.2 Variance Reduction

Stratification method For simplicity, let us consider an interval $\Omega = [0, 1]$ and a problem of

$$\int_{[0,1]} f(x) dx.$$

For a fixed $m > 0$, divide $[0, 1]$ into M equal subintervals $\Omega_k = [\frac{k-1}{M}, \frac{k}{M}]$.

Also for simplicity, assume that the sample size n is a multiple of m . Then, for each $k \leq m$, sample n/m points $\{x_i^k\}$ uniformly distributed in Ω_k .

$$\int_{[0,1]} f(x) dx \approx \frac{1}{n} \sum_k^m \sum_i^{n/m} f(x_i^k).$$

Then the error e is

$$e \approx n^{-1/2} \sigma_s$$

where $\sigma_s^2 = \sum_k^m \int_{\Omega_k} (f(x) - \bar{f}_k)^2 dx$ and $\bar{f}_k = \int_{\Omega_k} f(x) dx$.

7.3.2 Variance Reduction

Method: Stratification

Claim. The variance σ_{ξ}^2 is smaller than the variance without stratification $\sigma^2 = \int_{[0,1]} (f - \bar{f})^2 dx$ where $\bar{f} = \int_{[0,1]} f(x) dx$.

7.3.2 Variance Reduction

Method: Stratification

Claim. The variance σ_{ξ}^2 is smaller than the variance without stratification $\sigma^2 = \int_{[0,1]} (f - \bar{f})^2 dx$ where $\bar{f} = \int_{[0,1]} f(x) dx$.

Idea of proof. The minimizer c of $\int_{\Omega_k} (f(x) - c)^2 dx$ is $\bar{f}_k = \int_{\Omega_k} f(x) dx$.

7.4 Example

We want to calculate

$$p = \int_2^{\infty} \frac{1}{\pi(1+x^2)} dx = 0.15$$

- ▶ Estimator 1: $\hat{p}_1 = \frac{1}{n} \sum_i^n I(X_i > 2)$ where $\{X_i\}$ is from Cauchy
- ▶ Estimator 2:
 $\hat{p}_2 = \frac{1}{2n} \sum_i^n I(|X_i| > 2)$ where $\{X_i\}$ is from Cauchy
- ▶ Estimator 3: $\hat{p}_3 = \frac{1}{2} - \frac{1}{n} \sum_i^n \frac{1}{\pi(1+X_i)^2}$ where $\{X_i\}$ is from Uniform[0,2].
- ▶ Estimator 4: $\hat{p}_4 = \frac{1}{n} \sum_i^n \frac{X_i^{-2}}{\pi(1+X_i^{-2})}$ where $\{X_i\}$ is from Uniform[0,1/2].

Homework

1. Numerically calculate the variances of estimator 1,2,3 and 4 in the previous slide (show your work). Compare them with analytic results if possible.