Activity: Intro to Monte Carlo Integration

Group members:

Monte Carlo Integration

Suppose we wish to estimate the quantity $\theta = \int\limits_{\mathcal{X}} g(x)f(x)dx$, where f is some density function. Then, we recognize that

$$\theta = \mathbb{E}[g(X)]$$

where $X \sim f$ is a random variable with density f.

Monte Carlo integration estimates θ by generating a sample from f, and using the sample mean to approximate the true mean. In particular:

- Sample $X_1, ..., X_n \overset{iid}{\sim} f$
- **Monte Carlo estimate:** $\widehat{\theta} = \frac{1}{n}\sum\limits_{i=1}^{n} g(X_i)$

As shown in the slides, $\mathbb{E}[\widehat{\theta}] = \theta$ and so

$$MSE(\widehat{\theta}) = Var(\widehat{\theta}) = \frac{1}{n}Var(g(X))$$

As the sample size n increases, the variability (i.e., the error) in our estimate $\widehat{\theta}$ decreases.

Part 1

Suppose we wish to calculate the quantity $\theta = \int\limits_0^1 \frac{e^{-x}}{1+x^2}dx$

1. Find a pdf f and function g such that $\theta = \int\limits_0^1 g(x)f(x)dx$.

2. Sample $n = 10$ observations $X_1, ..., X_{10}$ from the distribution with pdf f, and report the Monte Carlo estimate $\widehat{\theta}$.

3. Repeat question 2 many times to approximate $MSE(\widehat{\theta})$ when $n = 10$. What is the approximate MSE?

4. Now repeat question 3 with different values of n, and plot $MSE(\widehat{\theta})$ agains n.

Part 2

Suppose we wish to calculate the quantity

$$\theta = \int_0^1 \frac{e^{-x}}{1+x^2}dx = \int_0^1 g(x)f(x)dx$$

As discussed in the slides, here are two possible options for f and g:

- $f_1(x) = 1$, $g_1(x) = \dfrac{e^{-x}}{1+x^2}$
- $f_2(x) = \dfrac{4}{\pi(1+x^2)}$, $g_2(x) = \frac{\pi}{4}e^{-x}$

5. The distribution with pdf $f_2(x) = \dfrac{4}{\pi(1+x^2)}$ has cdf $F_2(t) = \frac{4}{\pi}\text{atan}(t)$ for $t \in [0,1]$. Explain how to use the inverse transform method to sample $X \sim f_2$; that is, if $U \sim Uniform(0,1)$, find $F_2^{-1}(U)$ as a function of U.

6. Using the inverse transform method, sample $n = 10$ observations $X_1, ..., X_{10}$ from the distribution with pdf f_2, and report the Monte Carlo estimate $\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} g_2(X_i)$.

7. Repeat question 6 many times to approximate $MSE(\widehat{\theta}_2)$ when $n = 10$.

8. How does $MSE(\widehat{\theta}_2)$ compare to the MSE for the Monte Carlo estimate with f_1 and g_1?

9. Plot $\dfrac{e^{-x}}{1+x^2}$ for $x \in (0, 1)$, and add plots of $f_1(x)$ and $f_2(x)$. Why do you think using f_2 gives a Monte Carlo estimate with lower variability?