Monte Carlo Integration: Expected Values and Simulations

Introduction to Expected Value and Monte Carlo Integration

- We are often interested in computing $E(h(X))$, the expected value of a function $h(X)$, where X is a random variable.
- In some cases, $E(h(X))$ is easy to compute when the probability distribution $p_X(x)$ is known and simple.
- However, in more complex situations, direct computation of $E(h(X))$ becomes difficult or even impossible.

Motivating Example: Loading the Car

- Scenario: My family needs to load into the car quickly, involving tasks like putting on shoes, using the restroom, etc.
- Let X be the time it takes for the last of my five kids to get in the car.
- Define $X = \max\{Y_1, Y_2, \ldots, Y_5\}$, where Y_i are independent exponentially distributed random variables with rates $0.5, 0.75, 1.0, 0.75$, and 2.0, respectively.
- Computing $E(X)$ analytically is challenging due to the maximum of multiple exponential random variables.

Why is this Difficult?

- The distribution $p_X(x)$ of $X = \max\{Y_1, Y_2, \ldots, Y_5\}$ is not straightforward to calculate.
- Direct computation involves complex convolution of exponential distributions, which is difficult to express in a closed form.
- This motivates the use of Monte Carlo methods to estimate $E(X)$ through simulation.

Monte Carlo Integration for $E(h(X))$

- Monte Carlo integration allows us to estimate the expected value of $h(X)$ by simulating random values of X from its distribution $p_X(x)$.
- For independent samples $X_1, X_2, \ldots, X_n \sim p_X(x)$, the Monte Carlo estimate of $E(h(X))$ is:

$$\hat{E}(h(X)) = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

- As the sample size n increases, the estimate $\hat{E}(h(X))$ converges to the true expected value $E(h(X))$ (Law of Large Numbers).

Monte Carlo and Simulation Studies

- Monte Carlo integration is closely linked to simulations. In practice, you simulate samples from the distribution of interest and apply the function $h(X)$.
- In the loading car example, we simulate times $X = \max\{Y_1, \ldots, Y_5\}$, where the Y_i follow different exponential distributions.
- This technique provides a way to estimate complex integrals and expected values when direct computation is infeasible.

Formal Applications of Monte Carlo Integration

- Monte Carlo methods are widely used in formal settings such as:
  - Hypothesis testing: Generating random samples under the null hypothesis to evaluate the test statistic's distribution.
  - Model validation: Using simulations to assess the accuracy of a statistical or machine learning model.
  - Bayesian inference: Estimating posterior distributions through sampling (e.g., Markov Chain Monte Carlo).
- These methods are critical for solving complex, high-dimensional problems in science, finance, and engineering.

Probability as an Expected Value

- Consider the same family loading scenario: The time X it takes for the last of my five kids to get in the car is:

$$X = \max\{Y_1, Y_2, \ldots, Y_5\}$$

  where $Y_i \sim \text{Exp}(\lambda_i)$ with rates $0.5, 0.75, 1.0, 0.75$, and 2.0.
- We are interested in the probability that the car will be loaded in less than 3 minutes:

$$P(X \leq 3)$$

- This can be expressed as an expected value:

$$P(X \leq 3) = E[\mathbb{I}(X \leq 3)]$$

  where $\mathbb{I}(\cdot)$ is the indicator function.

Monte Carlo Estimation for Probabilities

- To estimate $P(X \leq 3)$, we use Monte Carlo integration:

$$\hat{P}(X \leq 3) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq 3)$$

  where X_i are independent samples of $X = \max\{Y_1, Y_2, \ldots, Y_5\}$.

- As n increases, the Monte Carlo estimate converges to the true probability $P(X \leq 3)$ due to the Law of Large Numbers.

Variance of the Monte Carlo Estimator

- The accuracy of a Monte Carlo estimate depends on its variance. For the general case of estimating $E(h(X))$, the variance of the Monte Carlo estimator is:

$$\text{Var}\left(\hat{E}(h(X))\right) = \frac{\text{Var}(h(X))}{n}$$

- For probabilities, where $h(X) = \mathbb{I}(X \leq 3)$, we have:

$$\text{Var}\left(\hat{P}(X \leq 3)\right) = \frac{P(X \leq 3)(1 - P(X \leq 3))}{n}$$

- The variance decreases as the sample size n increases.

Confidence Interval for Probability Estimate

From the variance we can compute the **Monte Carlo confidence interval** as

$$\hat{E}(h(X)) \pm 1.96 \times \sqrt{\text{Var}\left(\hat{E}(h(X))\right)}$$

- In the loading car example, we estimated $\hat{P}(X \leq 3) = 0.85$ with variance $\hat{\text{Var}} = 0.00001275$.
- The standard error is:

$$\text{SE} = \sqrt{\frac{0.85(1 - 0.85)}{10,000}} = 0.00357$$

- The 95% confidence interval is:

$$0.85 \pm 1.96 \times 0.00357 = [0.842, 0.858]$$

Step-by-Step Monte Carlo Integration

Step 1: Define the Problem

- We want to compute $E(h(X))$, where $h(X)$ is a function of a random variable X.
- Example: The time X it takes for the last of five kids to load the car.

Step 2: Simulate Random Samples

- Simulate independent random samples $X_1, X_2, \ldots, X_n$ from the distribution of X.
- In our case, simulate $X = \max\{Y_1, Y_2, \ldots, Y_5\}$, where $Y_i \sim \text{Exp}(\lambda_i)$.

Step-by-Step Monte Carlo Integration (cont.)

Step 3: Apply the Function

- For each sample X_i, compute $h(X_i)$.
- In the loading car example, for estimating the probability $P(X \leq 3)$, use $h(X) = \mathbb{I}(X \leq 3)$.

Step 4: Compute the Monte Carlo Estimate

- The Monte Carlo estimate of $E(h(X))$ is:

$$\hat{E}(h(X)) = \frac{1}{n} \sum_{i=1}^{n} h(X_i)$$

Step-by-Step Monte Carlo Variance

Step 5: Estimate the Variance and Confidence Interval

- The variance of the Monte Carlo estimator is given by:

$$\mathrm{Var}\left(\hat{E}(h(X))\right) = \frac{\mathrm{Var}(h(X))}{n}$$

- For probabilities, $h(X) = \mathbb{I}(X \leq 3)$, the variance becomes:

$$\mathrm{Var}\left(\hat{P}(X \leq 3)\right) = \frac{P(X \leq 3)(1 - P(X \leq 3))}{n}$$

- Find the Monte Carlo confidence interval using

$$\hat{E}(h(X)) \pm 1.96 \times \sqrt{\mathrm{Var}\left(\hat{E}(h(X))\right)}$$

Step-by-Step Reporting Monte Carlo Results

Step 6: Reporting Results in Context

- Report the Monte Carlo estimate $\hat{E}(h(X))$ along with the sample size n and the variance of the estimator or the Monte Carlo confidence interval.
- Example: After simulating 10,000 samples, the estimated probability that the car will be loaded within 3 minutes is $\hat{P}(X \leq 3) = 0.85$ with variance:

$$\hat{\text{Var}}\left(\hat{P}(X \leq 3)\right) = \frac{0.85(1 - 0.85)}{10,000} = 0.00001275$$

- Interpret the results in context: "The Monte Carlo estimate for the probability the the car will be loaded in 3 minutes is 85%, with a 95% Monte Carlo confidence interval of [0.842, 0.858]"