1 Monte Carlo Integration

In a simple Monte Carlo problem, we express the quantity we want to know as the expected value of a random variable Y, such as $\mu = E(Y)$. Then we generate values $Y_1, \ldots, Y_n$ independently from the distribution of Y and take their average

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

as an estimate of μ.

Sometimes, we are interested to approximate $\mu = \int h(x)f(x)dx$, where $f(\cdot)$ is a probability density function and $h(\cdot)$ is a real valued function. It is clear that $\mu = E(h(Y))$, where $Y \sim f(\cdot)$. Now, if $Y_1, Y_2, \ldots, Y_n$ are random number generated form the distribution of Y, then μ can be approximated by

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} h(Y_i).$$

Example 1. An important special case arises when $h(x) = 1_A(x)$, where 1_A is the indicator function. It is clear that $E(h(X)) = P(X \in A)$. Therefore, $P(X \in A)$ can be approximated using Monte Carlo method as follows.

1. Generate $X_1, X_2, \ldots, X_n$ from the distribution of X.
2. Find $N_n = \#\{i : X_i \in A\}$.
3. Approximate $P(X \in A)$ by $\frac{1}{n}\sum_{i=1}^{n} h(X_i) = \frac{1}{n}\sum_{i=1}^{n} 1_A(X_i) = \frac{N_n}{n}$. ||

1.1 Justification for Simple Monte Carlo

Let Y be a random variable for which $\mu = E(Y)$ exists, and suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed with the same distribution as Y. Then under the weak law of large numbers,

$$\lim_{n\to\infty} \mathbb{P}\left(|\widehat{\mu}_n - \mu| \le \epsilon\right) = 1,$$

holds for any $\epsilon > 0$. The weak law tells us that our chance of missing by more than ϵ goes to zero. The strong law of large numbers tells us a bit more. The absolute error $|\widehat{\mu}_n - \mu|$ will eventually get below ϵ and then stay there forever:

$$\mathbb{P}\left(\lim_{n\to\infty} |\widehat{\mu}_n - \mu| = 0\right) = 1.$$

1.2 Error and its Estimation

While both laws of large numbers tell us that Monte Carlo will eventually produce an error as small as we like, neither tells us how large n has to be for this to happen. They also do not say for a given sample $Y_1, \ldots, Y_n$ whether the error is likely to be small. The situation improves markedly when Y has a finite variance. Suppose that $Var(Y) = \sigma^2 < \infty$. In IID sampling, $\widehat{\mu}_n$ is a random variable and it has its own mean and variance. The mean of $\widehat{\mu}_n$ is

$$\mathbb{E}(\widehat{\mu}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \mu.$$

Because the expected value of $\widehat{\mu}_n$ is equal to μ, we say that simple Monte Carlo is unbiased. The variance of $\widehat{\mu}_n$ is

$$Var(\widehat{\mu}_n) = \mathbb{E}\left((\widehat{\mu}_n - \mu)^2\right) = \frac{\sigma^2}{n}.$$

As $\widehat{\mu}_n$ is unbiased, the variance of $\widehat{\mu}_n$ can be interpreted as average squared error. It is clear that for fixed n, the variance increases as σ increases. Similarly, the variance decreases as n increases for fixed σ. The root mean squared error (RMSE) of $\widehat{\mu}_n$ is defined by

$$\sqrt{Var(\widehat{\mu}_n)} = \frac{\sigma}{\sqrt{n}}.$$

The average squared error in Monte Carlo sampling is σ^2/n. We seldom know σ^2 but it is easy to estimate it from the sample values. The most commonly used estimates of σ^2 are

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \widehat{\mu}_n\right)^2,$$

and

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{\mu}_n\right)^2.$$

Monte Carlo sampling usually uses such large values of n, and hence s^2 and $\widehat{\sigma}^2$ will be much closer to each other than either of them is to actual variance σ^2. The familiar motivation for s^2 is that it is unbiased, *i.e.*, $\mathbb{E}(s^2) = \sigma^2$, for $n \geq 2$. Finally, an estimator of variance of $\widehat{\mu}_n$ is

$$\frac{s_n^2}{n} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(Y_i - \widehat{\mu}_n\right)^2 \quad \text{or} \quad \frac{\widehat{\sigma}_n^2}{n} = \frac{1}{n^2}\sum_{i=1}^{n}\left(Y_i - \widehat{\mu}_n\right)^2.$$

The formula for s is simple enough but, perhaps surprisingly, it can lead to numerical difficulties specially when n is large and when $\sigma \ll |\mu|$. There is a way to obtain good numerical stability in a one-pass algorithm. Let $S_n = \sum_{i=1}^{n}\left(y_i - \widehat{\mu}_n\right)^2$. Starting with $\widehat{\mu}_1 = y_1$ and $S_1 = 0$, make the updates

$$\begin{aligned}
\delta_i &= y_i - \widehat{\mu}_{i-1} \\
\widehat{\mu}_i &= \widehat{\mu}_{i-1} + \frac{\delta_i}{i} \\
S_i &= S_{i-1} + \frac{i-1}{i}\delta_i^2
\end{aligned}$$

for $i = 2, \ldots, n$. Then approximate s_n^2 by $\frac{S_n}{(n-1)}$ and $\widehat{\sigma}^2$ by $\frac{S_n}{n}$.

1.3 Confidence Interval

Let $Y_1, \ldots, Y_n$ be a random sample. Let L and U be two functions having domain that includes the sample space of $Y_1, \ldots, Y_n$ and $L(\underline{y}) \le U(\underline{y})$ for all $\underline{y}$. Then the random interval $(L(Y_1, \ldots, Y_n), U(Y_1, \ldots, Y_n))$ is called a $100(1-\alpha)\%$ confidence interval for a parameter θ if $P_\theta\left(L(Y_1, \ldots, Y_n) \le \theta \le U(Y_1, \ldots, Y_n)\right) \ge 1-\alpha$. Here $\alpha \in (0, 1)$. A small value of α is useful.

An asymptotic confidence interval for μ can be computed using the central limit theorem. The central limit theorem (CLT) states the following: Let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and finite variance $\sigma^2 > 0$. Then

$$\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} Z \sim N(0,1),$$

i.e., for all $z \in \mathbb{R}$

$$P\left(\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\sigma} \le z\right) \to \Phi(z),$$

as $n \to \infty$, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

CLT can be used to get asymptotic confidence interval for μ, but it requires that we know σ. However, if σ is unknown, we can proceed as follows. Note that using weak law of large numbers, $s_n^2 \xrightarrow{P} \sigma^2$. Now, using Slutsky's Theorem,

$$\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{s_n} \xrightarrow{\mathcal{D}} Z \sim N(0,\, 1).$$

In other words, for all $z \in \mathbb{R}$,

$$\mathbb{P}\left(\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{s_n} \le z\right) \to \Phi(z),$$

as $n \to \infty$. Hence, for $\Delta > 0$

$$\begin{aligned}
\mathbb{P}\left(\widehat{\mu}_n - \frac{\Delta s_n}{\sqrt{n}} \le \mu \le \widehat{\mu}_n + \frac{\Delta s_n}{\sqrt{n}}\right) &= \mathbb{P}\left(-\Delta \le \sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{s_n} \le \Delta\right) \\
&\to \Phi(\Delta) - \Phi(-\Delta) \\
&= 2\Phi(\Delta) - 1.
\end{aligned}$$

For a 95% confidence interval, set $2\Phi(\Delta) - 1 = 0.95$. Then $\Delta = \Phi^{-1}(0.975) \approx 1.96$, yielding the familiar 95% confidence interval $\left(\widehat{\mu}_n - 1.96\frac{s}{\sqrt{n}},\, \widehat{\mu}_n + 1.96\frac{s}{\sqrt{n}}\right)$. Similarly 99% confidence interval can be found as $\left(\widehat{\mu}_n - 2.58\frac{s}{\sqrt{n}},\, \widehat{\mu}_n + 2.58\frac{s}{\sqrt{n}}\right)$.