

# 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  $Y_1, \dots, Y_n$  independently from the distribution of  $Y$  and take their average

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

as an estimate of  $\mu$ .

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  $Y_1, Y_2, \dots, Y_n$  are random number generated from the distribution of  $Y$ , then  $\mu$  can be approximated by

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

## 1.1 Justification for Simple Monte Carlo

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\mu}_n - \mu| \leq \epsilon) = 1,$$

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

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

## 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  $Y_1, \dots, Y_n$  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,  $\hat{\mu}_n$  is a random variable and it has its own mean and variance. The mean of  $\hat{\mu}_n$  is

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

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

$$\text{Var}(\hat{\mu}_n) = \mathbb{E}((\hat{\mu}_n - \mu)^2) = \frac{\sigma^2}{n}.$$

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

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

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

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

and

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

Monte Carlo sampling usually uses such large values of  $n$ , and hence  $s^2$  and  $\hat{\sigma}^2$  will be much closer to each other than either of them is to actual variance  $\sigma^2$ . 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  $\hat{\mu}_n$  is

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

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  $S_n = \sum_{i=1}^n (y_i - \hat{\mu}_n)^2$ . Starting with  $\hat{\mu}_1 = y_1$  and  $S_1 = 0$ , make the updates

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

for  $i = 2, \dots, n$ . Then use  $s_n^2 = S_n/(n-1)$  in approximate confidence intervals.

### 1.3 Confidence Interval

Let  $Y_1, \dots, Y_n$  be a random sample. Let  $L$  and  $U$  be two functions having domain that includes the sample space of  $Y_1, \dots, Y_n$  and  $L(y) \leq U(y)$  for all  $y$ . Then the random interval  $(L(Y_1, \dots, Y_n), U(Y_1, \dots, Y_n))$  is called a  $100(1 - \alpha)\%$  confidence interval for a parameter  $\theta$  if  $P_\theta(L(Y_1, \dots, Y_n) \leq \theta \leq U(Y_1, \dots, Y_n)) \geq 1 - \alpha$ . Here  $\alpha \in (0, 1)$ . A small value of  $\alpha$  is useful.

An asymptotic confidence interval for  $\mu$  can be computed using central limit theorem. The

central limit theorem (CLT) states the following: Let  $Y_1, \dots, Y_n$  be independent and identically distributed random variables with mean  $\mu$  and finite variance  $\sigma^2 > 0$ . Then

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

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

$$P\left(\sqrt{n} \frac{\hat{\mu}_n - \mu}{\sigma} \leq z\right) \rightarrow \Phi(z),$$

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

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

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

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

$$\mathbb{P}\left(\sqrt{n} \frac{\hat{\mu}_n - \mu}{s_n} \leq z\right) \rightarrow \Phi(z),$$

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

$$\begin{aligned} \mathbb{P}\left(\hat{\mu}_n - \frac{\Delta s}{\sqrt{n}} \leq \mu \leq \hat{\mu}_n + \frac{\Delta s}{\sqrt{n}}\right) &= \mathbb{P}\left(-\Delta \leq \sqrt{n} \frac{\hat{\mu}_n - \mu}{s_n} \leq \Delta\right) \\ &\rightarrow \Phi(\Delta) - \Phi(-\Delta) \\ &= 2\Phi(\Delta) - 1. \end{aligned}$$

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  $\left(\hat{\mu}_n - 1.96 \frac{s}{\sqrt{n}}, \hat{\mu}_n + 1.96 \frac{s}{\sqrt{n}}\right)$ . Similarly 99% confidence interval can be found as  $\left(\hat{\mu}_n - 2.58 \frac{s}{\sqrt{n}}, \hat{\mu}_n + 2.58 \frac{s}{\sqrt{n}}\right)$ .

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