Exercise – Understanding the theory of Monte Carlo integration

1. Aim of the exercise

The aim of the exercise is to understand the intuition behind Monte Carlo integration. Instead of computing a definite integral directly, we reinterpret it as the expected value of a function with respect to a chosen probability distribution. By sampling from this distribution and averaging the function values, we approximate the integral. This approach is especially powerful in high dimensions, where traditional methods struggle. Monte Carlo's strength lies in its simplicity and dimension-independent convergence rate.

2. Theory

For the general case, consider the following definite integral which we wish to approximate.

$$I = \int_A \phi(x)\, dx \tag{1}$$

where $A \subseteq \mathbb{R}$, and $\phi : \mathbb{R} \to \mathbb{R}$ is an integrable function.

The key concept behind Monte Carlo integration is to represent I as an expectation which we can then approximate by a sample average. To achieve this, first we choose a random variable X which has probability density function (pdf) $f_X : \mathbb{R} \to \mathbb{R}$ with $supp(f_X) = A$. Next, we rewrite ϕ as the product of f_X and some function $g : \mathbb{R} \to \mathbb{R}$. That is, $\phi(x) = g(x)f_X(x)$. In practice one usually chooses the distribution of X, $g(x)$ then becomes $\frac{\phi(x)}{f_X(x)}$. Then it follows by the law of the unconscious statistician (LOTUS) that

$$\int_A \phi(x)\, dx = \int_A \frac{\phi(x)}{f_X(x)} \cdot f_X(x)\, dx = \int_A g(x) \cdot f_X(x)\, dx = \mathbb{E}_X[g(X)]. \tag{2}$$

The subscript in $\mathbb{E}_X$ denotes that we are taking the expected value with respect to X.

When choosing the distribution of X, it is necessary that $supp(f_X) = A$. Because of this, when I has finite bounds of integration, a uniform distribution is often a convenient choice for the distribution of X. Since if $X \sim$ Unif(a,b) then $supp(f_X) = [a, b]$ allowing for any choice of finite integration bounds.

When I has infinite bound(s) of integration, the distribution of X is usually chosen so that it is easy to sample from and so that $f_X(x)$ is a function similar to $\phi(x)$. For example, when integrating an exponential function over the positive real numbers an exponential distribution for X would be a natural choice.

Now we will estimate I by estimating $\mathbb{E}_X[g(X)]$ using Monte Carlo methods (Metropolis and Ulam, 1949). We do this by first taking N random samples, $(X_1, ..., X_N)$. Where $X_i \overset{\text{i.i.d.}}{\sim} X \quad \forall i \in (1, \cdots, N)$. We then calculate the Monte Carlo estimator.

$$\bar{g}_N := \frac{1}{N} \sum_{i=1}^{N} g(X_i). \tag{3}$$

Next, we will show that g_N is both consistent and unbiased, properties that are desirable for an estimator (Robert and Casella, 2010). To prove this we will be using the strong law of large numbers (SLLN).

Theorem 1 (Strong Law of Large Numbers)**.** *Let $Y_1, Y_2, \ldots$ be pairwise i.i.d. random variables with $\mathbb{E}[|Y_i|] < \infty$, $\forall i \in \mathbb{N}$. Then:*

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \xrightarrow{a.s.} \mathbb{E}[Y_i] \quad \Longleftrightarrow \quad \mathbb{P}\left(\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} Y_i = \mathbb{E}[Y_i]\right) = 1. \tag{4}$$

This theorem informs us that for an i.i.d. sample with finite expectation, the empirical average converges to the true mean as $N \to \infty$. Since a function of a random variable is still a random variable, the SSLN informs us that when the sample size N goes to infinity

$$\lim_{N\to\infty} \bar{g}_N = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} g(X_i) \xrightarrow{a.s.} \mathbb{E}_X[g(X)] = \int_A \phi(x)\, dx = I \tag{5}$$

which shows the consistency of the Monte Carlo estimator (Durrett, 2019). Next, we derive the bias and variance.

Theorem 2. *$\bar{g}_N$ is an unbiased estimator, with variance:*

$$\mathrm{Var}_{f_X}[\bar{g}_N] = \frac{\mathrm{Var}_{f_X}[g(X)]}{N}, \tag{6}$$

assuming $\mathrm{Var}_{f_X}[g(X)]$ exists.

Proof. For unbiasedness we have

$$\mathbb{E}_X[\bar{g}_N] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_X[g(X_i)] \overset{i.d.}{=} \frac{1}{N} \cdot N \cdot \mathbb{E}_X[g(X)] = \mathbb{E}_X[g(X)] = I \tag{7}$$

where the first equality holds by linearity of expectation.

For variance we have

$$\mathrm{Var}_X[\bar{g}_N] \overset{ind.}{=} \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}_X[g(X_i)] \overset{i.d.}{=} \frac{1}{N^2} \cdot N \cdot \mathrm{Var}_X[g(X)] = \frac{\mathrm{Var}_X[g(X)]}{N}, \tag{8}$$

where for both derivations we used that $X_i \overset{\text{i.i.d.}}{\sim} X$.

Furthermore, we can derive the asymptotic distribution. By Robbins central limit theorem (Robbins, 1943), we obtain:

$$\begin{aligned} \frac{\bar{g}_N - \mathbb{E}_X[\bar{g}_N]}{\sqrt{\mathrm{Var}_{f_X}[\bar{g}_N]}} \xrightarrow{d} \mathcal{N}(0,1) \quad \Longleftrightarrow \quad \bar{g}_N &\xrightarrow{d} \mathcal{N}\left(\mathbb{E}_X[\bar{g}_N], \mathrm{Var}_{f_X}[\bar{g}_N]\right) \\ &\overset{d}{=} \mathcal{N}\left(\int_A f(x)\, dx, \left[\frac{\mathrm{Var}_{f_X}[g(X)]}{\sqrt{N}}\right]^2\right) \end{aligned} \tag{9}$$

where $\mathbb{E}_X[\bar{g}_N] < \infty$ by assumption. ■

An important consequence of the above result is that the standard deviation of the Monte Carlo integration estimator $\bar{g}_N$ is of order of $\mathcal{O}(N^{-1/2})$.

The variance of $\bar{g}_N$ depends on $\mathrm{Var}_X[g(X)]$, which may or may not be known. We can estimate it, and consequentially $\mathrm{Var}_X[\bar{g}_N]$, by the sample variance:

$$\widehat{\mathrm{Var}_X[g(X)]} = \frac{1}{N-1}\sum_{i=1}^{N}(g(x_i) - \bar{g}_N)^2$$

$$\widehat{\mathrm{Var}_X[\bar{g}_N]} = \frac{1}{N}\widehat{\mathrm{Var}_X[g(X)]} = \frac{1}{N}\cdot\left[\frac{1}{N-1}\sum_{i=1}^{N}(g(x_i) - \bar{g}_N)^2\right] = \frac{1}{N(N-1)}\sum_{i=1}^{N}(g(x_i) - \bar{g}_N)^2, \tag{10}$$

where x_i denotes a realization of X_i.

The estimator $\widehat{\mathrm{Var}_{f_X}[\bar{g}_N]}$ is unbiased, since the sample variance is an unbiased estimator and

$$\mathbb{E}\left[\widehat{\mathrm{Var}_{f_X}[\bar{g}_N]}\right] = \frac{1}{N}\mathrm{Var}_{f_X}[g(X)] = \mathrm{Var}_{f_X}[\bar{g}_N], \tag{11}$$

A more general measure for the accuracy of an estimator is the mean squared error (MSE),

$$MSE(\bar{g}_N) := \mathbb{E}_X\left[\left(\int_A f(x)\,dx - \bar{g}_N\right)^2\right] = \mathrm{Var}_{f_X}[\bar{g}_N] + \mathrm{Bias}(\bar{g}_N)^2, \tag{12}$$

where $\mathrm{Bias}(\bar{g}_N) := \mathbb{E}_X[\bar{g}_N] - \int_A f(x)\,dx$. Therefore, when the estimator is unbiased, the MSE equals the variance of the estimator. Since $\bar{g}_N$ is unbiased, the MSE and variance are equivalent.

Now that we have examined the one dimensional case, we will look at Monte Carlo estimation in higher dimensions. Generalizing Monte Carlo integration to a d-dimensional integral over a d-dimensional space is relatively straightforward. Suppose we wish to estimate the following definite integral

$$I = \int_A \phi(\boldsymbol{x})\,d\boldsymbol{x} = \int_{A_d}\cdots\int_{A_1}\phi(x_1, ..., x_d)\,dx_1\cdots dx_d, \tag{13}$$

where $A := A_1 \times ... \times A_d \subseteq \mathbb{R}^d$ is our domain of integration and $\phi : A^d \to \mathbb{R}$ is an integrable function.

Then, we consider a vector of random variables $\boldsymbol{X} = (X_1, ..., X_d)'$ where the support of $X_i = A_i \quad \forall i \in (1, \ldots, d)$. Like in the one-dimensional case, we wish to rewrite ϕ as the joint pdf of $\boldsymbol{X}$, $f_{\boldsymbol{X}} : A \to \mathbb{R}$ and some function $g : \mathbb{R}^d \to \mathbb{R}$ chosen such that $\phi(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x})g(\boldsymbol{x})$.

We can then proceed by using the same ideas as the one-dimensional case, namely

$$I = \int_A \phi(\boldsymbol{x})\,d\boldsymbol{x} = \int_A g(\boldsymbol{x})\cdot f_{\boldsymbol{X}}(\boldsymbol{x})\,d\boldsymbol{x} = \mathbb{E}_{\boldsymbol{X}}[g(\boldsymbol{X})] \tag{14}$$

which again motivates our estimator

$$\bar{g}_N := \frac{1}{n}\sum_{i=1}^{N} g(\boldsymbol{X}_i), \tag{15}$$

where $\boldsymbol{X}_i = (X_{i,1}, ..., X_{i,d})'$ is part of a sequence of random vectors $(\boldsymbol{X}_1, ..., \boldsymbol{X}_N)'$ where $\boldsymbol{X_i} \overset{\text{i.i.d.}}{\sim} \boldsymbol{X} \quad \forall i \in (1, \ldots, N)$.

The SLLN in Theorem 1 also holds for random vectors. Hence, it holds that

$$\bar{g}_N = \frac{1}{n}\sum_{i=1}^{N} g(\boldsymbol{X}_i) \xrightarrow{a.s.} \mathbb{E}_{f_{\boldsymbol{X}}}[g(\boldsymbol{X})]. \tag{16}$$

A generalized version of Theorem 2 also holds in the case of multiple integrals.

Furthermore, since all the random variables X_j in the random vector $\boldsymbol{X}$ are independent, the joint PDF $f_{\boldsymbol{X}}$ can be written as a product of univariate PDFs:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) := f_{X_1,...,X_d}(x_1, ..., x_d) = f_{X_1}(x_1) \cdot ... \cdot f_{X_d}(x_d) = \prod_{j=1}^{d} f_{X_j}(x_j), \tag{17}$$

where $f_{X_j}(x_j)$ denotes the marginal/univariate PDF of the jth entry of the random vector $\boldsymbol{X}$.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}}\left[g(\boldsymbol{X})\right] &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(\boldsymbol{x}) \cdot \left[f_{X_1}(x_1) \cdot ... \cdot f_{X_d}(x_d)\right] \, dx_1 \cdots dx_d \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{X_d}(x_d) \cdot ... \cdot f_{X_2}(x_2) \int_{\mathbb{R}} g(\boldsymbol{x}) \cdot f_{X_1}(x_1) \, dx_1 \cdots dx_d \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{X_d}(x_d) \cdot ... \cdot f_{X_3}(x_3) \cdot \int_{\mathbb{R}} \mathbb{E}_{f_{X_1}}\left[g(x_1, x_2, ..., x_d)\right] \cdot f_{X_2}(x_2) \, dx_2 \cdots dx_d \\
&= \mathbb{E}_{X_d} \circ ... \circ \mathbb{E}_{X_2} \circ \mathbb{E}_{X_1}\left[g(\boldsymbol{X})\right].
\end{aligned} \tag{18}$$

When g can be written in the form $g(\boldsymbol{x}) = \prod_{j=1}^{d} g(x_j)$, then we can further simplify equation (18) to

$$\begin{aligned}
\mathbb{E}_{f_{\boldsymbol{X}}}\left[g(\boldsymbol{X})\right] &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{j=1}^{d} g(x_j) \cdot f_{X_j}(x_j) \, dx_1 \cdots dx_d \\
&\overset{(\star)}{=} \prod_{j=1}^{d} \int_{\mathbb{R}} g(x_j) \cdot f_{X_j}(x_j) \, dx_j,
\end{aligned} \tag{19}$$

where we used Fubini's theorem to switch the product and the integral (Durrett, 2019).

Observe that equations (18) and (19) only provide an approximation technique in special cases. Generally, for complicated multi-dimensional integrals, it is not possible to use the approximation technique described in equation (16) because it is often infeasible to directly draw from a multi-dimensional PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$. When the PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$ in equation (16) is a 'common' multi-dimensional PDF from which we can directly draw samples, then equation (16) can be used. An example of a common multi-dimensional PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$, that does not satisfy equation (17) but from which we can directly draw samples, is a multivariate normal PDF with a covariance matrix that has non-zero off-diagonal elements.

The main appeal of Monte Carlo integration compared to deterministic quadrature methods has to do with the convergence rate $\mathcal{O}(n^{-\frac{1}{2}})$ from Theorem 2. Generally speaking, this means that to halve the standard deviation of the estimator, we need four times as many samples. In contrast, approximating the d-dimensional integral in equation (13) by deterministic methods, such as Riemann summation, yields a standard deviation of order $\mathcal{O}(n^{-\frac{1}{d}})$. Note that the order depends on the dimension d, this is referred to as the curse of dimensionality. This is however not the case with Monte Carlo integration, because $\mathcal{O}(n^{-\frac{1}{2}})$ does not depend on d. This makes Monte Carlo integration a useful technique for high-dimensional integrals, which are relevant for applications such as MSL. A comparison of different convergence rates are given in the exercise titled "Convergence speed through Big O notation".

3. Monte Carlo integration algorithm

To approximate the integral using Monte Carlo simulation, follow this algorithm:

Algorithm: Monte Carlo integration

Input: Sample size n
for $i = 1, \ldots, n$
    Sample $X_i \sim f_X$
    Compute $g(X_i)$
end
Output: Estimator $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$

4. Final notes

This file is prepared and copyrighted by Jelmer Wieringa and Tunga Kantarcı. This file and the accompanying MATLAB file are available on GitHub and can be accessed using this link. We have used four references. First is Robbins, H., 1948. On the asymptotic distribution of the sum of a random number of random variables, Proc. Natl. Acad. Sci. U.S.A. 34 (4) 162-163. Second is Durrett, R. T. (Ed.), 2019. Probability: Theory and Examples. Cambridge University Press: Cambridge, New York. Third is Robert, C. P., Casella, G. (Eds.), 2010. Introducing Monte Carlo Methods with R. Springer: New York, London. Fourth is Metropolis, N. C., Ulam, S. M., 1949, The Monte Carlo Method. Journal of the American Statistical Association , Vol. 44, No. 247, p. 335-341.