Monte Carlo Integration I. Exercises. [RC] Chapter 3.1

Anatoli Iambartsev
IME-USP

Classical Monte Carlo integration, [S].

Calculate integral of a function $h(x, y)$ on an area $G \subset \mathbb{R}^2$

$$m = \int_G h(x, y) f(x, y) dx dy$$

where $f(x, y)$ is a density on G, i.e.

$$\int_G f(x, y) dx dy = 1.$$

Classical Monte Carlo integration, [S].

Observe that any integral on a finite area G with $S_G = |G|$

$$\int_G h(x,y)dxdy$$

can be represented as an integral of type

$$m = \int_G h_1(x,y)f(x,y)dxdy$$

Indeed, considering

$$h_1(x,y) = S_G h(x,y), \quad f(x,y) = 1/S_G,$$

we obtain

$$\int_G h(x,y)dxdy = \int_G h_1(x,y)f(x,y)dxdy.$$

Classical Monte Carlo integration, [S].

Remember that in order to estimate the integral $m = \int_G h(x,y)f(x,y)dxdy$ we construct (simulate) a random point $p = (x,y)$ with density $f(x,y)$ and random variable $Z = h(p) = h(x,y)$ which expectation is equal to m :

$$\mathbb{E}(Z) = \int_G h(p)f(p)dp.$$

Thus, if there exists $\mathbb{E}|Z|$ we have a convergence in probability and also almost sure convergence

$$m_n = \frac{1}{n}\sum_{i=1}^{n} h(p_i) \to m,$$

as $n \to \infty$.

Classical Monte Carlo integration. [S]

If the function h is bounded, suppose $0 \leq h(p) \leq c$ for any $p \in G$, we can suggest a *geometric method* to estimate the integral. Consider a volume $\tilde{G} = G \times [0, c]$. Consider a random point $q = (x, y, z)$ on $\tilde{G}$ with density $f(x, y, z) = f(x, y)/c$. Note that the marginal distribution of q em area G has a density $f(x, y)$. Choosing n independent realizations $q_1, \ldots, q_n$ of q we construct the following estimator

$$\tilde{m}_n = \frac{c\nu}{n},$$

where

$$\nu = \#\{\text{times when } q_i \text{ stays under the surface } h(\cdot)\}$$

Prove, $\nu \sim B(n, p_\nu)$, where $p_\nu = m/c$.

Classical Monte Carlo integration.[S]

$$\nu = \#\{\text{times when } q_i \text{ stayed under the surface } h(\cdot)\}$$

Prove, $\nu \sim B(n, p_\nu)$, where $p_\nu = m/c$.

Indeed,

$$\begin{aligned} \nu &= \sum_{i=1}^{n} \mathbb{1}(q_i \text{ stays under the surface } h(\cdot)) \\ &= \sum_{i=1}^{n} \mathbb{1}(z_i < h(x_i, y_i)) \end{aligned}$$

Classical Monte Carlo integration. [S]

$$\nu = \#\{\text{times when } q_i \text{ stays under the surface } h(\cdot)\}$$

Prove, $\nu \sim B(n, p_\nu)$, where $p_\nu = m/c$.

And,

$$\begin{aligned}
p_\nu &= \mathbb{E}(\mathbb{1}(Z_i < h(X_i, Y_i))) \\
&= \iiint \mathbb{1}(z < h(x,y)) f(x,y,z)\,dx\,dy\,dz \\
&= \iint \Big(\int_0^1 \frac{1}{c}\mathbb{1}(z < h(x,y))\,dz\Big) f(x,y)\,dx\,dy \\
&= \frac{1}{c}\iint \Big(\int_0^{h(x,y)} dz\Big) f(x,y)\,dx\,dy \\
&= \frac{1}{c}\iint h(x,y) f(x,y)\,dx\,dy \;=\; \frac{m}{c}.
\end{aligned}$$

Classical Monte Carlo integration. [S]

We see that $\mathbb{E}\tilde{m}_n = m$, and also we have almost sure convergence $\tilde{m}_n \to m$. Note that we represented $\tilde{m}_n$ in alternative form

$$\tilde{m}_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i, \text{ where } \tilde{Z}_i = \begin{cases} c, & \text{if } z_i < h(x_i, y_i), \\ 0, & \text{if } z_i \geq h(x_i, y_i). \end{cases}$$

Compare this estimator with

$$m_n = \frac{1}{n}\sum_{i=1}^{n} h(p_i) \;\; (= Z_i).$$

Classical Monte Carlo integration. [S]

In order to compare two estimators

$$m_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \text{ and } \tilde{m}_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i,$$

we require the existence of the second moment for calculation of the variances

$$\mathbb{V}ar Z = \int_G h^2(p) f(p) dp - m^2 \text{ and } \mathbb{V}ar \tilde{Z} = cm - m^2,$$

because $\mathbb{E}(\tilde{Z}^2) = c^2 \mathbb{P}(Z < h(X,Y)) = c \cdot m$.

If $0 \leq h(p) \leq c$, then

$$\int_G h^2(p) f(p) dp \leq c \int_G h(p) f(p) dp = c \cdot m \Rightarrow \mathbb{V}ar Z \leq \mathbb{V}ar \tilde{Z}.$$

Classical Monte Carlo integration, [S].

Example. Calculate the integral

$$I = \int_0^1 e^x dx$$

the estimators corresponding to the estimators considered before are

$$\theta_N = \frac{1}{N}\sum_{i=1}^{N} e^{U_i}, \qquad \tilde{\theta}_N = e\frac{\nu}{N},$$

where ν is the number of pairs $(U_i, U_i'), \ldots, (U_N, U_N')$ such that $U_i' < e^{U_i}$ (as usial $U_i, U_i' \sim U[0,1]$ are i.i.d.).

$$\mathbb{V}ar(Z) = \int_0^1 e^{2x} dx - m^2 = \frac{1}{2}(e^2 - 1) - (e-1)^2 \cong 0.2420$$

$$\mathbb{V}ar(\tilde{Z}) = e \cdot m - m^2 = e - 1 \cong 1.7183$$

Classical MC integration. Efficiency.

Consider two estimators

$$m_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \text{ and } m_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(2)}$$

where

$$X_i^{(1)} = X_i^{(1)}(U_1, \ldots, U_{n_1}),\ \ X_i^{(2)} = X_i^{(2)}(U_1, \ldots, U_{n_2}).$$

Let $T^{(1)}, T^{(2)}$ be time spent in calculation a value of $X^{(1)}$ and $X^{(2)}$ correspondingly. It is natural to suppose that an algorithm is more efficient if it spends less time to achieve, say, the probable error accuracy $r_n = \varepsilon$.

Classical MC integration. Efficiency.

Let $T^{(1)}, T^{(2)}$ be time spent in calculation a value of $X^{(1)}$ and $X^{(2)}$ correspondingly. It is natural to suppose that an algorithm is more efficient if it spends less time to achieve, say, the probable error accuracy $r_n = \varepsilon$.

Thus the efficiencies of two estimators can be defined as

$$T^{(1)}n^{(1)} \quad \text{and} \quad T^{(2)}n^{(2)}$$

where $n^{(1)}, n^{(2)}$ are sample sizes needed to achieve the accuracy r_n. Thus, using estimation for sample sizes we obtain the efficiency of algorithms

$$T^{(1)}\mathbb{V}ar(X^{(1)})\Big(\frac{0.6745}{\varepsilon}\Big)^2 \text{ and } T^{(2)}\mathbb{V}ar(X^{(2)})\Big(\frac{0.6745}{\varepsilon}\Big)^2.$$

Classical Monte Carlo integration.

Example. Estimate the integral

$$m = \int_0^1 \sqrt[5]{x} dx = \frac{5}{6}.$$

We consider two methods: direct method and “geometric” method.

Classical Monte Carlo integration.

Example. Estimate the integral

$$m = \int_0^1 \sqrt[5]{x} dx = \frac{5}{6}.$$

Direct method: $h(x) = \sqrt[5]{x}$ and $f(x) = 1$ if $x \in (0,1)$, $f(x) = 0$ if $x \notin (0,1)$, thus $X^{(1)} = h(U_1)$ and

$$m_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \sqrt[5]{U_i}.$$

$$\mathbb{V}ar(X^{(1)}) = \int_0^1 x^{2/5} dx - \left(\frac{5}{6}\right)^2 = \frac{5}{252}.$$

Classical Monte Carlo integration.

Example. Estimate the integral

$$m = \int_0^1 \sqrt[5]{x}dx = \frac{5}{6}.$$

"Geometric" method: since $0 \le \sqrt[5]{x} \le 1$, then

$$m_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(U_i' < \sqrt[5]{U_i}).$$

and $X^{(2)} = X^{(2)}(U_1, U_1') = \mathbb{1}(U_i' < \sqrt[5]{U_i})$ variance

$$\mathbb{V}ar\big(\mathbb{1}(U_i' < \sqrt[5]{U_i})\big) = \frac{5}{6} - \left(\frac{5}{6}\right)^2 = \frac{5}{36}.$$

Classical Monte Carlo integration.

Example. Estimate the integral

$$m = \int_0^1 \sqrt[5]{x} dx = \frac{5}{6}.$$

Direct and "geometric" methods:

$$m_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \sqrt[5]{U_i}, \quad m_n^{(2)} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(U_i' < \sqrt[5]{U_i}).$$

$$\mathbb{V}ar(X^{(1)}) = \frac{5}{252} < \mathbb{V}ar\big(X^{(2)}\big) = \frac{5}{36}.$$

Classical Monte Carlo integration. Infinite variance.

Consider two integrals

$$I_1 = \int_0^1 \frac{dx}{\sqrt{x}} = 2 \quad \text{and} \quad I_2 = \int_0^1 \frac{dx}{\sqrt[3]{x}} = \frac{3}{2}.$$

Consider their estimators

$$\begin{aligned} \widehat{I}_N^{(1)} &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sqrt{U_i}} =: \frac{1}{N}\sum_{i=1}^{N} X_i, \\ \widehat{I}_N^{(2)} &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sqrt[3]{V_i}} =: \frac{1}{N}\sum_{i=1}^{N} Y_i, \end{aligned}$$

where $U_i, V_i \sim U[0,1]$ i.i.d.

Classical Monte Carlo integration. Infinite variance.

$$f_X(x) = \begin{cases} \frac{2}{x^3} & x \in [1, \infty) \\ 0, & \text{otherwise} \end{cases}, f_Y(y) = \begin{cases} \frac{3}{y^4} & x \in [1, \infty) \\ 0, & \text{otherwise} \end{cases}$$

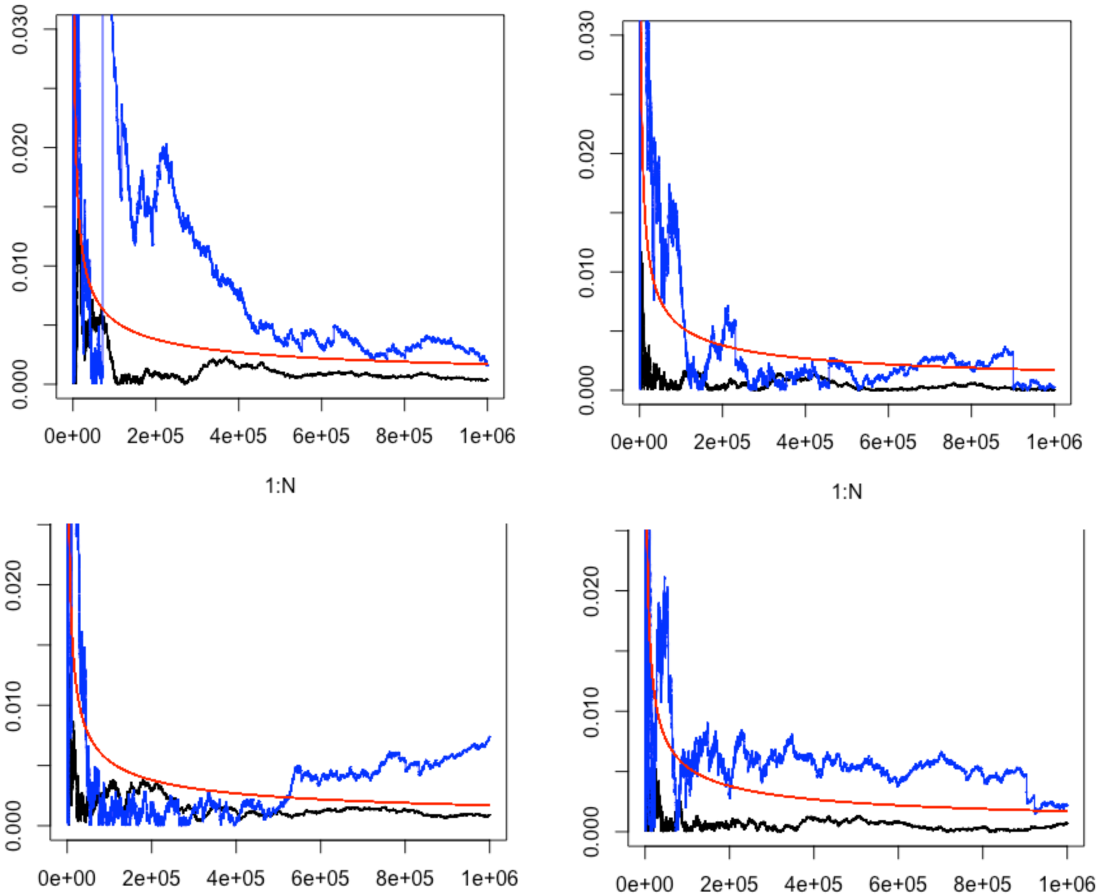
$$\mathbb{E}(X_i) = 2,\ \mathbb{V}ar(X_i) = \infty,\ \mathbb{E}(Y_i) = 2,\ \mathbb{V}ar(Y_i) = \frac{3}{4}.$$

We will plot: for $n \in \{1, 2, \ldots, N\}$

black line: $|\widehat{I}_n^{(2)} - 3/2|$;

blue line: $|\widehat{I}_n^{(1)} - 2|$;

red line: $1.96\sqrt{(3/4)/n}$.


0.000
0.010
0.020
0.030
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06
1:N
0.000
0.010
0.020
0.030
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06
1:N
0.000
0.010
0.020
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06
0.000
0.010
0.020
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06

References:

[RC] Cristian P. Robert and George Casella. *Introducing Monte Carlo Methods with R*. Series "Use R!". Springer

[S] Sobol, I.M. *Monte-Carlo numerical methods*. Nauka, Moscow, 1973. (In Russian)

[As] Asmussen, S. and Glynn, P.W. *Stochastic Simulation. Algorithms and Analysis.* Springer. 2010