Monte Carlo Integration I

[RC] Chapter 3

Anatoli Iambartsev
IME-USP

There is no exact definition of the Monte Carlo methods. In the course Monte Carlo Method means *numeric method of a solution of a mathematical problem using a modeling of random variables and statistical estimation of their parameters.* Thus, we are talking

- numerical method, and not analytical;
- can resolve not only probabilistic/statistical problems.

The "official" year of birth is 1949:

Metropolis N., Ulam S.M. *The Monte Caro method*. J. Amer. Assoc., **44**, 247, 335 341, 1949

[RC, p62] "The major classes of numerical problems that arise in statistical inference are *optimization* problems and *integration* problems. Numerous examples show that it is not always possible to analytically compute the estimators associated with a given paradigm (maximum likelihood, Bayes, method of moment, ext.)."

[RC, p62] "The possibility of producing an almost infinite number of random variables distributed according to a given distribution gives us access to the use of frequentist and asymptotic results much more easily than in the usual inferential settings, where the sample size is most often fixed. One can therefore apply probabilistic results such as the LLN or CLT, since they allow assessment of the convergence of simulation methods (which is equivalent to the deterministic bounds used by numerical approaches)."

Classical Monte Carlo integration.

The generic problem is about the representation of an integral

$$\int F(x)dx = \int h(x)f(x)dx =: \mathbb{E}\Big(h(X)\Big)$$

as the expectation of some function $h(\cdot)$ of a r.v. X with density $f(\cdot)$. The principle of the Monte Carlo method for approximating

$$\int h(x)f(x)dx =: \mathbb{E}\Big(h(X)\Big)$$

is to generate a sample $(X_1, \ldots, X_n)$ from the density f and propose as an approximation

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

which converges almost surely to $\mathbb{E}\Big(h(X)\Big)$ by the Strong LLN.

Classical Monte Carlo integration.

The following questions arise.

- how to choose a good r.v. X;
- how to simulate values of r.v. X;
- convergence, when to stop.

Let X be a real-valued random variable, and let $X_1, X_2, \ldots$ be an infinite sequence of independent and identically distributed copies of X. Let $\overline{X}_n := \frac{1}{n}(X_1 + \ldots X_n)$ be the empirical average of this sequence. *Law of Large Numbers* (LLN) is known in two form, weak and strong:

Weak LLN.

If the first moment of X is finite, i.e. $\mathbb{E}(|X|) < \infty$, then $\overline{X}_n$ converges *in probability* to $\mathbb{E}(X)$:

$$\lim_{n\to\infty} \mathbb{P}\Big(|\overline{X}_n - \mathbb{E}(X)| \geq \varepsilon\Big) = 0, \text{ for every } \varepsilon > 0.$$

Strong LLN.

If the first moment of X is finite, i.e. $\mathbb{E}(|X|) < \infty$, then $\overline{X}_n$ converges *almost surely* to $\mathbb{E}(X)$:

$$\mathbb{P}\Big(\lim_{n\to\infty} \overline{X}_n = \mathbb{E}(X)\Big) = 1.$$

Classical Monte Carlo integration. By the Strong LLN the integral approximation

$$m_n = \frac{1}{n}\sum_{i=1}^{n} h(x_i) \to \mathbb{E}\big(h(X)\big), \text{ a.s.}$$

If there exists the second moment, by CLT, for large n

$$\frac{m_n - \mathbb{E}\big(h(X)\big)}{\sqrt{\frac{\mathbb{V}ar\big(h(X)\big)}{n}}} \approx N(0,1),$$

This leads to the confidence bounds on the approximation of $\mathbb{E}\big(h(X)\big)$.

Classical Monte Carlo integration.

Indeed, let $m := \mathbb{E}(X)$, and $v := \mathbb{V}ar(X)$ (further we omit h)

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - m)\right| < x\sqrt{\frac{v}{n}}\right) = 2\Phi(x) - 1.$$

thus for n large

$$\mathbb{P}\left(|m_n - m| < x\sqrt{\frac{v}{n}}\right) \approx 2\Phi(x) - 1.$$

Given $1 - \alpha/2 = \Phi(x)$ we find quantile $z_{1-\alpha/2}$ and

$$\mathbb{P}\left(|m_n - m| < z_{1-\alpha/2}\sqrt{\frac{v}{n}}\right) \approx 1 - \alpha.$$

Classical Monte Carlo integration.

[As. p70] "Being the same dimension as the expectation m, the standard deviation $\sqrt{v}$ is the natural measure of precision rather then the variance v. The rate $n^{-1/2}$ is intrinsic for virtually any Monte Carlo experiment as the optimal rate which can be achieved. In fact, there are quite a few situations in which one get slower rates such as $O(n^{-(1/2-\varepsilon)})$ with $\varepsilon > 0$, e.g., in gradient estimation via finite differences (VII.1), or variance estimation via time series methods (IV.3). In view of these remarks, one often refers to $O(n^{-1/2})$ as the *canonical Monte Carlo convergence rate*.

There are, however, a few isolated examples (but just a few!) in which $O(n^{-1/2})$ can in fact be improved to supercanonical (faster) rates; see V.7.4 for examples."

Classical Monte Carlo integration.

Example ([As,p.70], see also [RC] Example 3.4). As an illustration of the slow convergence rate of the Monte Carlo method, consider the estimation of $\pi = 3.1415...$ via the estimator $Z := 4 \cdot \mathbb{1}(U_1^2 + U_2^2 < 1)$, where $U_1, U_2 \sim U[0,1]$ are independent r.v.s. Here $\mathbb{1}(U_1^2 + U_2^2 < 1) \sim B(\pi/4)$, and $m = \mathbb{E}(Z) = \pi, v = \mathbb{V}ar(Z) = 4^2\frac{\pi}{4}(1-\frac{\pi}{4}) \cong 2.70$. Thus, to get the leading 3 in π correct w.p. 95% we need

$$|m_n - m| \leq 0.5 = z_{0.975}\sqrt{\frac{v}{n}} = 1.96\sqrt{\frac{2.70}{n}}$$

thus $n \cong \frac{1.96^2 2.7}{0.5^2} \cong 41$, whereas for the next 1 we need $n = \frac{1.96^2 2.7}{0.05^2} \cong 4,150$, for the 4, $n = \frac{1.96^2 2.7}{0.005^2} \cong 415,000$, and so on.

Classical Monte Carlo integration.

If there exists a third central moment $\beta_3 = \mathbb{E}(X - m)^3 < \infty$, the rate convergence in CLT can be improved:

$$\left|\mathbb{P}(S_n - mn > \sigma\sqrt{nx}) - \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-u^2/2}du\right| \leq \frac{c_0\beta_3}{v^{3/2}\sqrt{n}}$$

where c_0 is an absolute constant such that $1/\sqrt{\pi} \leq c_0 < 0.9051$.

Classical Monte Carlo integration.

Sometimes the *probable error of a method*, r_n is used:

$$r_n = z_{0.75}\sqrt{\frac{v}{n}} = 0.6745\sqrt{\frac{v}{n}}.$$

The name "probable" is because

$$\mathbb{P}\{|m_n - m| < r_n\} \cong 1/2 \cong \mathbb{P}\{|m_n - m| > r_n\},$$

i.e. the fluctuations greater then r_n and fluctuation less then r_n are equiprobable. In practice r_n is used to estimate an order of an error.

Classical Monte Carlo integration. Errors.

[As.p71-72] In many computational settings, one wishes to calculate the solution to a desired level of accuracy. Two accuracy criteria are commonly used:

•absolute accuracy: compute the quantity m to an accuracy ε

$$|m_n - m| < \varepsilon;$$

•relative accuracy: compute the quantity m to an accuracy $\varepsilon|m|$

$$|m_n - m| < \varepsilon|m|.$$

The confidence interval suggests, in principle, what one should do to achieve such accuracies.

Classical MC integration. Production runs.

[As.p71-72] The confidence interval suggests, in principle, what one should do to achieve such accuracies.

$$n \approx \frac{z^2_{1-\alpha/2} v}{\varepsilon^2} \qquad n \approx \frac{z^2_{1-\alpha/2} v}{\varepsilon^2 |m|^2}$$

As in the setting of confidence intervals, these run-length determination rules (i.e., rules for selecting n) cannot be implemented directly, since they rely on problem-dependent parameters such as v and m that are unknown.

Classical MC integration. Production runs.

[As.p.72] As a consequence, the standard means of implementing this idea involves first generating a small number of trial runs (say 50) to estimate v and m^2 by $\widehat{v}_{trial}$ and $\widehat{m}^2_{trial}$. One then determines the number n of so-called *production runs* from either

$$n \approx \frac{z^2_{1-\alpha/2}\widehat{v}_{trial}}{\varepsilon^2} \quad \text{or} \quad n \approx \frac{z^2_{1-\alpha/2}\widehat{v}_{trial}}{\varepsilon^2 \widehat{m}^2_{trial}},$$

depending on whether absolute or relative precision is desired.

Classical MC integration. Error without "$\widehat{v}$".

Suppose $v < \infty$. There is a way to estimate an error without calculation of $\widehat{v}$. Let $n = kn_1$, where $k \geq 3$ is integer and not large. Suppose that n_1 is large enough to apply the CLT. Calculate the averages

$$\zeta_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \zeta_2 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{i+n_1}, \ldots, \zeta_2 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{i+n_1(k-1)}.$$

If $\zeta_1, \ldots, \zeta_k$ are i.i.d. with normal distribution with mean m then

$$t = \sqrt{k-1}\frac{\bar{x} - m}{s}, \text{where } \bar{x} = \frac{1}{k}\sum_{i=1}^{k}\zeta_i,\ s^2 = \frac{1}{k}\sum_{i=1}^{k}(\zeta_i - \bar{x})^2,$$

has t-Student distribution with $k-1$ degree of freedom.

Classical MC integration. Error without "$\widehat{v}$".

$$t = \sqrt{k-1}\frac{\bar{x}-m}{s}, \text{where } \bar{x} = \frac{1}{k}\sum_{i=1}^{k}\zeta_i,\ s^2 = \frac{1}{k}\sum_{i=1}^{k}(\zeta_i - \bar{x})^2,$$

in our case $\mathbb{E}(\zeta_i) = m, \bar{x} = m_n$ and $s^2 = \frac{1}{k}\sum_{i=1}^{k}(\zeta_i - m_n)^2$. Thus

$$\mathbb{P}\left(|m_n - m| < x\sqrt{s^2/(k-1)}\right) \cong 2\int_0^x p_{k-1}^{Student}(y)dy$$

and

$$\mathbb{P}\left(|m_n - m| < t_{k-1,1-\alpha}\sqrt{s^2/(k-1)}\right) \cong 1-\alpha.$$

The probable error of this method is

$$r_n = t_{k-1,0.5}\sqrt{s^2/(k-1)}$$

Classical MC integration. Infinite variance.

Recommendation: do not use variables with infinite variance!

Classical MC integration. Partial integration. [S]

Monte Carlo error is proportional $\sqrt{\mathbb{V}ar(h(X))/n}$. We cannot to improve $\sqrt{n}$, but we can try to decrease $\mathbb{V}ar(h(X))$. Thus the aim is to find estimator with lower variance.

The aim is to calculate $m = \int_G h(x)f(x)dx$, where $h \in L_2(G; f)$. Suppose that there exists the function $g(x)$ which integral can be calculated explicitly

$$I = \int_G g(x)f(x)dx.$$

Classical MC integration. Partial integration. [S]

Aim: $m = \int_G h(x)f(x)dx$, known: $I = \int_G g(x)f(x)dx$.

Again, $X_i \sim f(\cdot)$ and consider two estimators

$$m_n = \frac{1}{n}\sum_i h(X_i), \quad m'_n = \frac{1}{n}\sum_i \Big(I + h(X_i) - g(X_i)\Big),$$

with $\mathbb{E}\Big(h(X)\Big) = \mathbb{E}\Big(I + h(X) - g(X)\Big) = m$,

$$\mathbb{V}ar\Big(I + h(X) - g(X)\Big) = \int_G \Big(h(x) - g(x)\Big)^2 f(x)dx - (m - I)^2.$$

If g is close enough to h such that

$$\int_G \Big(h(x) - g(x)\Big)^2 f(x)dx \leq \varepsilon,$$

then

$$\mathbb{V}ar\Big(I + h(X) - g(X)\Big) \leq \int_G \Big(h(x) - g(x)\Big)^2 f(x)dx \leq \varepsilon.$$

Classical MC integration. Partial integration. [S]

Example. Let $h(x) = e^x, g(x) = 1 + x$. Estimate the integral $m = \int_0^1 e^x dx$. Compare variance of two estimators:

$$m_n = \frac{1}{n}\sum_i e^{U_i}, \text{ and } m'_n = \frac{1}{n}\sum_i (\frac{3}{2} + e^{U_i} - 1 - U_i).$$

Essentially we compare

$$\mathbb{V}ar(e^U) \text{ and } \mathbb{V}ar(e^U - U).$$

Explicit calculations gives us:

$$\mathbb{V}ar(e^U) = \frac{(3-e)(e-1)}{2} \cong 0.2420$$

$$\mathbb{V}ar(e^U - U) = \frac{-6e^2 + 36e - 53}{12} \cong 0.0436$$

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