Probabilistic Numerical Methods 2024–2025

Lecture 1: Monte Carlo method and random variable generation

julien.reygner@enpc.fr

1 The Monte Carlo method

The goal of the Monte Carlo method is to numerically approximate an integral which writes under the form

$$\mathcal{I} := \int_{x\in E} f(x)P(\mathrm{d}x), \tag{1}$$

where $(E, \mathcal{E})$ is a measurable space, P is a probability measure on E and $f \in \mathbf{L}^1(P)$.

1.1 Deterministic approach

Assume for simplicity that $E = [0,1]^d$ and that P has a density p with respect to the Lebesgue measure. Then setting $g(x) = f(x)p(x)$, fixing $N \geq 1$ and setting $x_{\vec{k}} = (k_1/N, \dots, k_d/N)$ for $\vec{k} = (k_1, \dots, k_d) \in \{0, \dots, N-1\}^d$, the basic deterministic approximation of $\mathcal{I}$ is given by

$$\mathcal{I}_N := \frac{1}{N^d} \sum_{\vec{k}} g(x_{\vec{k}}),$$

obtained by replacing g with the piecewise constant function which takes the value $g(x_{\vec{k}})$ on the cell $C_{\vec{k}} := [k_1/N, (k_1+1)/N) \times \cdots \times [k_d/N, (k_d+1)/N)$.

The precision of this approximation is given by the fact that, if you assume that g is Lipschitz continuous, then

$$|\mathcal{I} - \mathcal{I}_N| = \left| \sum_{\vec{k}} \int_{C_{\vec{k}}} (g(x) - g(x_{\vec{k}}))\mathrm{d}x \right| \leq \sum_{\vec{k}} \int_{C_{\vec{k}}} \left| g(x) - g(x_{\vec{k}}) \right| \mathrm{d}x \lesssim \frac{1}{N}.$$

As a consequence, to reach a precision $\epsilon \simeq 1/N$, one needs to evaluate f at $N^d \simeq (1/\epsilon)^d$ points. This quantity grows exponentially in d: this is the *curse of dimensionality*.

1.2 Stochastic approach

The formulation (1) of $\mathcal{I}$ allows us to rewrite it under the form

$$\mathcal{I} = \mathbb{E}[f(X)],$$

where X is a random variable in E with law P. Then, if $(X_n)_{n\geq 1}$ is a sequence of iid random variables with common distribution P, the (strong) Law of Large Numbers ensures that

$$\widehat{\mathcal{I}}_n := \frac{1}{n} \sum_{i=1}^n f(X_i)$$

converges almost surely to $\mathcal{I}$. The precision of the approximation of $\mathcal{I}$ by $\widehat{\mathcal{I}}_n$ is measured by the Central Limit Theorem, which ensures that if $\sigma^2 := \mathrm{Var}(f(X)) < +\infty$, then

$$\lim_{n\to+\infty} \frac{\sqrt{n}}{\sigma} \left(\widehat{\mathcal{I}}_n - \mathcal{I} \right) = \mathcal{N}(0,1), \qquad \text{in distribution,}$$

where $\mathcal{N}(0,1)$ denotes the standard Gaussian distribution. This result ensures in particular that, given $\alpha \in (0,1/2)$ and denoting by $\phi_{1-\alpha/2}$ the quantile of order $1-\alpha/2$ of $\mathcal{N}(0,1)$, the interval $[\widehat{\mathcal{I}}_n \pm \phi_{1-\alpha/2}\sigma/\sqrt{n}]$ contains $\mathcal{I}$ with probability converging to $1-\alpha$ when $n \to +\infty$. Therefore, to reach a precision $\epsilon \simeq \sigma/\sqrt{n}$, one needs to evaluate f at $n \simeq \sigma^2/\epsilon^2$ points, which only depends on the underlying dimension of E through the prefactor σ^2. So this method avoids the curse of dimensionality.

The remainder of this lecture serves two objectives:

1. describe *how*, in practice, one may generate iid samples from a given distribution P;
2. present methods allowing to improve the precision of the Monte Carlo method when the factor σ^2 is too large.

2 Random variable generation

2.1 Uniform distribution

When you ask a random number to a computer without being more specific, it generally returns a number uniformly distributed in $[0,1]$. Historically, computers used *Linear Congruential Generators* for this task. But one of the limits of these generators is that they have a relatively small period: after (at most) 2^{32} or 2^{64} iterations, the sequence becomes periodic. Nowadays computers mostly use the Mersenne Twister which has a much larger period. So now the game is: given a 'black box' generator which returns independent numbers uniformly distributed on $[0,1]$, and a target probability measure P, how to use the former in order to generate samples from the latter?

Basic application: Bernoulli, Binomial, Geometric; link with `if`, `for` and `while` commands.

2.2 Inverse CDF method

Explanation on a random variable $X \in \{1,\dots,K\}$ such that $\mathbb{P}(X=k)=p_k$: generalisation of the `if` condition for Bernoulli. Can also be seen as a particular case of the following approach: given an arbitrary random variable $X \in \mathbb{R}$ with CDF $F(x) := \mathbb{P}(X \le x)$, define its pseudo-inverse by $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \ge u\}$, $u \in (0,1)$. Then $F^{-1}(u) \le x$ if and only if $u \le F(x)$. This ensures that if $U \sim \mathcal{U}[0,1]$, then $F^{-1}(U)$ has the same law as X.

Application: Exponential random variables.

2.3 Box–Muller method for Gaussian random variables

Since $X \sim \mathcal{N}(\mu,\sigma^2)$ if and only if $X = \mu + \sigma G$ with $G \sim \mathcal{N}(0,1)$, it suffices to focus on $\mathcal{N}(0,1)$.

Box–Muller method: if $\Theta \sim \mathcal{U}[0,2\pi]$ and $R \sim \mathcal{E}(1/2)$ are independent, then $X := \sqrt{R}\cos\Theta$ and $Y := \sqrt{R}\sin\Theta$ are independent and $\mathcal{N}(0,1)$-distributed.

2.4 Rejection sampling

Principle of the method for sampling from the uniform distribution on some bounded subset of $\mathbb{R}^d$ with positive Lebesgue measure. General method when considering $D \subset \mathbb{R}^d \times [0,+\infty)$ as the graph of the target density p.

Application: the Ziggurat algorithm.

3 Variance reduction

3.1 The rare event problem

Come back to the estimation of $\mathcal{I} = \mathbb{E}[f(X)]$. Assume that $f(x) = \mathbb{1}_{\{x \in A\}}$ for some subset A such that $\mathcal{I} = \mathbb{P}(X \in A) \ll 1$. Then basically $\widehat{\mathcal{I}}_n$ is always 0 for reasonable values of n. More quantitatively, to reach a relative precision δ, that is to say that the length of the Monte Carlo confidence interval $\simeq \sigma/\sqrt{n} \simeq \sqrt{\mathcal{I}/n}$ be of order $\delta\mathcal{I}$, one needs $n \simeq 1/(\delta\mathcal{I})^2$. Therefore there is an interest in reducing the variance σ^2.

3.2 Importance sampling

Principle of the method and optimal choice of the density. In practice it cannot be employed, but the goal is to find tractable densities which 'look like' the optimal one.

Application: we want to estimate $\mathbb{P}(X \geq 20)$ for $X \sim \mathcal{N}(0, 1)$. Which density q should we use? Two propositions:

1. q is the law of $20 + E$, where $E \sim \mathcal{E}(\lambda)$ for some $\lambda > 0$.
2. q is the density of $\mathcal{N}(20, 1)$.

Homework: which one is the best?

Importance sampling can be implemented in an adaptative way. Other variance reduction techniques include: control variate, stratified sampling, splitting...