Introduction to Monte Carlo methods

Emanuele Coccia

Dipartimento di Scienze Chimiche e Farmaceutiche, Universitá di Trieste

ecoccia@units.it

December 18, 2022

Outline

1. Markov chain
2. Metropolis method
3. Monte Carlo integration

Monte Carlo methods: introduction

- Class of techniques to simulate the behaviour of a physical or mathematical system
- Stochastic methods $\rightarrow$ use of random number sequences
- Used to model random processes
- Evaluation of multidimensional integrals
- Solution of partial differential equations (Schrödinger equation)

Some useful definitions

- Random number: numerical value resulting from a process whose value cannot be predetermined by the initial conditions (attention!)
- State: allowed value of the set of properties of the system
- Sample space: set of all possible states (discrete or continuous)
- Sample point: single point in sample space
- Random variable: variable whose value lies within the sample space with a certain probability distribution

- Probability density function (pdf) p:
  - probability associated to the occurrence $x = x_i$ (discrete case)

$$\sum_i^N p(x_i) = 1$$

  - $p(x)dx$ probability of an event occurring between x and $x + dx$ (continuous case)

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

- Multivariate pdf $p(x_1, x_2, ..., x_n) = p(\mathbf{x})$: probability of the state $x_1 = X_1, ..., x_n = X_n$ (discrete) or $x_1 \leq X_1 \leq x_1 + dx, ..., x_n \leq X_n \leq x_n + dx$ (continuous)

$$P(a_1 \leq x_1 \leq b_1, ..., a_n \leq x_n \leq b_n) = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} p(\mathbf{x}) d\mathbf{x}$$

- Marginal density p_i: probability that the single component X_i lies within $x_i \leq X_i \leq x_i + dx$, regardless of the other components:

$$p_i(x_i) = \int_{-\infty}^{+\infty} p(\mathbf{x}) dx_1 dx_2 ... dx_{i-1} dx_{i+i} ... dx_n$$

- Expected value:

$$\langle f \rangle_p \equiv \int_{-\infty}^{+\infty} f(x) p(x) dx$$

- K-th moment of the distribution:

$$\langle x^k \rangle_p \equiv \int_{-\infty}^{+\infty} x^k p(x) dx$$

- K=1 $\rightarrow \langle x \rangle$ ($\bar{x}$, mean value)
- K-th central moment:

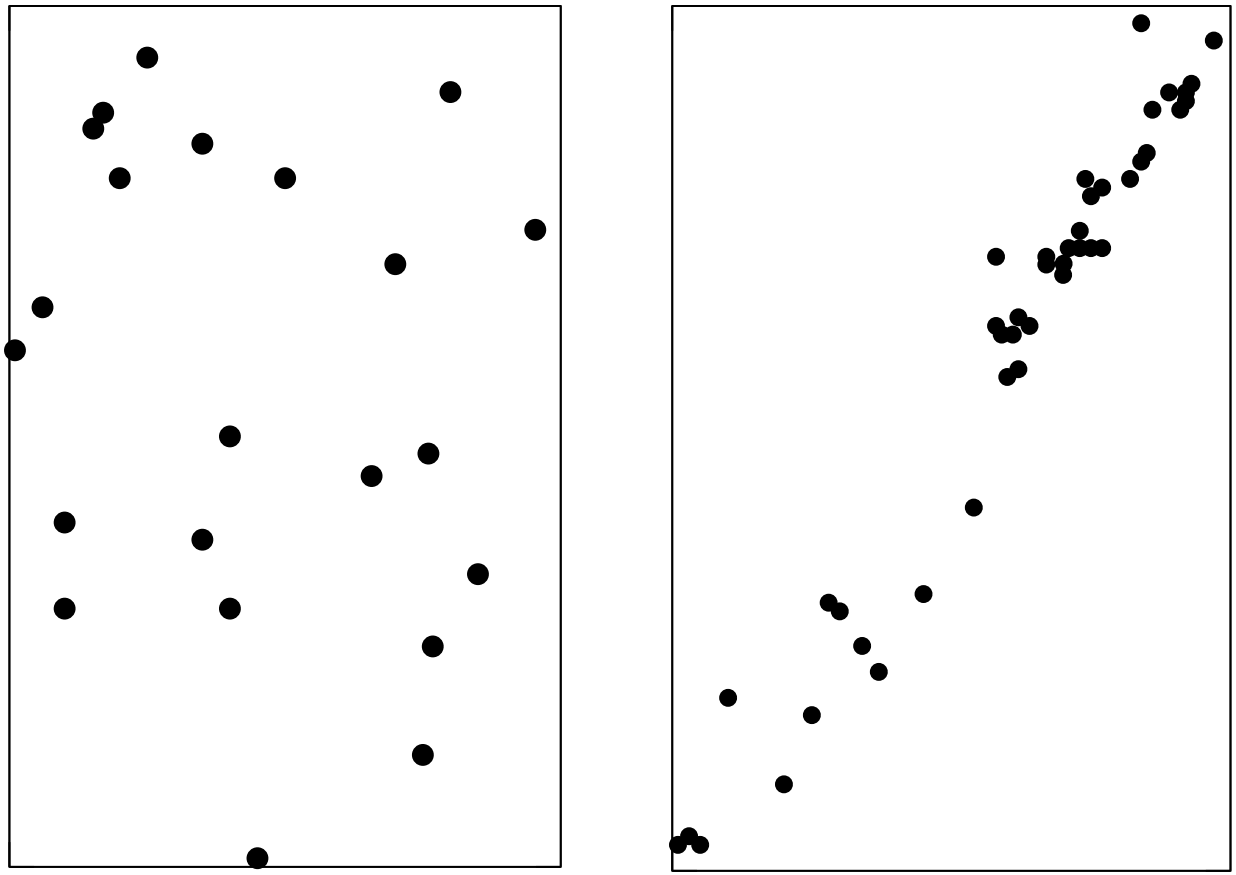
$$\langle (x - \bar{x})^k \rangle_p \equiv \int_{-\infty}^{+\infty} (x - \bar{x})^k p(x) dx$$

- K=2 $\rightarrow \sigma^2$ (variance), σ (standard deviation)

$$\sigma^2 = \langle x^2 \rangle - \bar{x}^2$$

- Straightforward extension to the multivariate case ($\bar{x}_i$, σ_i^2 etc)
- Covariance:

$$\mathrm{Cov}(x_i, x_j) \equiv \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle$$



(a) (b)

Random walks: introduction

- The drunkard's walk: a walker takes random moves on a two-dimensional lattice.
- This walk is readily generalized to higher dimensions and to continuous space.
- The sampling method was originally introduced by Metropolis; this algorithm has been widely applied to the generation of multivariate probability density functions (particularly in statistical mechanics and in Quantum Monte Carlo simulations)

Markov chains

- Walker: mathematical entity whose attributes completely define the state of the system
- The walker moves in a given space by a combination of deterministic and random displacements
- Consider a discrete system of N available states, named S_1 through S_N
- At every discrete point in time i, $x^{(i)} = S_j$ if the system is in the state S_j at time i
- The sequence of events of $x^{(i)}$ from time zero to the end of the walk forms a chain
- Such chain is a Markov chain if given the present state, future states are independent of the past states

- The probability of the system changing from state S_j to S_k in one time step is given by

$$P_{kj} \equiv P(x^{i+1} = S_k \leftarrow x^i = S_j).$$

- Normalization:

$$\sum_{k=1}^{N} P_{kj} = 1.$$

- $p_k^{(i)} \rightarrow$ probability that the system is in state S_k at time i
- Vector representation of $\mathbf{p}$

$$\mathbf{p}^{(i)} = \begin{bmatrix} p_1^{(i)} \\ . \\ . \\ . \\ p_N^{(i)} \end{bmatrix}, \sum_k^N p_k^{(i)} = 1$$

Evolution of p:

$$
\begin{aligned}
p_k^{(i+1)} &= \sum_j P_{kj} p_j^{(i)} \\
\mathbf{p}^{(i+1)} &= \mathbf{P}\mathbf{p}^{(i)} \\
\mathbf{p}^{(1)} &= \mathbf{P}\mathbf{p}^{(0)} \\
\mathbf{p}^{(2)} &= \mathbf{P}\mathbf{p}^{(1)} = \mathbf{P}\mathbf{P}\mathbf{p}^{(0)} \\
\mathbf{p}^{(m)} &= \mathbf{P}^m\mathbf{p}^{(0)}
\end{aligned}
$$

- After a sufficiently long time M, $|\mathbf{p}^{(M+1)} - \mathbf{p}^{(M)}| \to 0$
- Equilibrium probability distribution $\mathbf{p}^*$

$$\mathbf{p}^* = \mathbf{P}\mathbf{p}^*$$

- $\mathbf{p}^*$ is a stationary state of the transition matrix $\mathbf{P}$

Numerical example: three state Markov process with

$$\mathbf{P} = \begin{bmatrix} 1/4 & 1/8 & 2/3 \\ 3/4 & 5/8 & 0 \\ 0 & 1/4 & 1/3 \end{bmatrix}$$

and

$$\mathbf{p}^0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

The first three steps are

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1/4 \\ 3/4 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 5/32 \\ 21/32 \\ 6/32 \end{bmatrix}$$

Iteration	p_1	p_2	p_3
0	1.00000	0.00000	0.00000
1	0.25000	0.75000	0.00000
2	0.15625	0.65625	0.18750
3	0.24609	0.52734	0.22656
4	0.27848	0.51416	0.20736
5	0.27213	0.53021	0.19766
6	0.26608	0.53548	0.19844
7	0.26575	0.53424	0.20002
8	0.26656	0.53321	0.20023
9	0.26678	0.53318	0.20005
10	0.26671	0.53332	0.19998
11	0.26666	0.53335	0.19999
12	0.26666	0.53334	0.20000
13	0.26667	0.53333	0.20000
$\mathbf{p}^*$	0.26667	0.53333	0.20000

- Steady solution of **P**

$$\begin{aligned} p_1^* &= 1/4p_1^* + 1/8p_2^* + 2/3p_3^* \\ p_2^* &= 3/4p_1^* + 5/8p_3^* \\ p_3^* &= 1/4p_2^* + 1/3p_3^* \end{aligned}$$

- with the constraint $p_1^* + p_2^* + p_3^* = 1$
- The final (steady) state is given by

$$\mathbf{p}^* = \begin{bmatrix} 4/15 \\ 8/15 \\ 3/15 \end{bmatrix}$$

- Generalization to continuous variables
- $G(\mathbf{y}, \mathbf{x}; \Delta t)$: probability of moving from $\mathbf{x}$ at time t to $\mathbf{y}$ at time $t + \Delta t$
- $f(\mathbf{x}, t)$: probability density for a particle at $\mathbf{x}$ at time t (continuous analog of $\mathbf{p}^{(i)}$)

$$f(\mathbf{y}, t + \Delta t) = \int f(\mathbf{x}, t) G(\mathbf{y}, \mathbf{x}; \Delta t) d\mathbf{x}$$

$$f(\mathbf{y}, t + m\Delta t) = \int f(\mathbf{x}, t) G(\mathbf{y}, \mathbf{x}; m\Delta t) d\mathbf{x}$$

- Exists an equilibrium distribution function $f^*(\mathbf{y})$, independent of time

$$f^*(\mathbf{y}) = \int f^*(\mathbf{x}) G(\mathbf{y}, \mathbf{x}; \Delta t) d\mathbf{x}$$

Random walks in state space

- Evolution of the distribution in the state space S
- What does it mean for a Markov chain to converge to an equilibrium density?
- For a single walker, its states are sampled during the walk with probability $\mathbf{p}^* \Rightarrow$ the averages will be time independent
- Necessary condition for a random walk to reach equilibrium is ergodicity: spatial averages in the limit of infinite system are equal to time averages
- All the possible states must have a nonzero possibility of being visited (necessary but non sufficient!)
- Given a state S_k, $\sum_j P_{kj} \neq 0$

- A single walker visits the points $X^{(0)}, X^{(1)}, ..., X^{(m)}$ during the walk

$$\langle f \rangle_{\{t\}} = \frac{1}{m} \sum_{i=1}^{m} f(X^{(i)})$$

- Consider an ensemble of walkers $\{X\} = X_1, X_2, ..., X_N$

$$\langle f \rangle_{\{X\}} = \frac{1}{N} \sum_{k=1}^{N} f(X_k)$$

- The two averages are equivalent if drawn from the equilibrium distribution

$$\langle f \rangle_{\mathbf{p}^*} = \frac{1}{mN} \sum_{i=1}^{m} \sum_{k=1}^{N} f(X_k^{(i)})$$

Metropolis method

Metropolis method

- $\mathbf{p}^*$ arises from a Markov process associated to the transition matrix $\mathbf{P}$
- How to invert the procedure and find the appropriate $\mathbf{P}$ for the desired $\mathbf{p}^*$?
- The answer is in the Metropolis method
- Use of an acceptance/rejection step
- S_i: state with the max probability, i.e. $p_i = \max(\mathbf{p}^*)$
- $A_{ki} = p_k^*/p_i^* \Rightarrow$ acceptance probability of moving from S_i to S_k
- Similarly for the second most probable state S_j: $A_{kj} = p_k^*/p_j^*$ $(k \neq i)$ and so on
- $A_{kj} * A_{ji} = A_{ki}$

- Construction of upper (lower) triangular part of the matrix **A**
- The diagonal elements and A_{ik} set to unity

$$\mathbf{A}(3X3) = \begin{bmatrix} 1 & A_{12} & A_{13} \\ 1 & 1 & A_{23} \\ 1 & 1 & 1 \end{bmatrix}$$

- This choice of **A** produces the equilibrium distribution
- Equilibrium: the ratio of populations in two states is p_i^*/p_j^* $(p_i^* > p_j^*)$
- ν_i: current population in S_i
- ν_j: current population in S_j

- All the ν_j walkers at S_j may move to S_i (since $A_{ij} = 1$)
- From S_i to S_j the number of walkers is

$$\nu_i p_j^* / p_i^* = \nu_i A_{ji}$$

- The net change in population for the above transition is

$$\delta\nu_j = \nu_i p_j^* / p_i^* - \nu_j$$

- $\delta\nu_j = 0 \rightarrow \nu_i/\nu_j = p_i^*/p_j^* \rightarrow$ equilibrium
- $\delta\nu_j > 0 \rightarrow \nu_i/\nu_j > p_i^*/p_j^* \rightarrow$ population at S_j increases
- $\delta\nu_j < 0 \rightarrow \nu_i/\nu_j < p_i^*/p_j^* \rightarrow$ population at S_j decreases
- The two inequalities are driven towards equality and equilibrium $\Rightarrow$ right choice of **A**

- For continuous variables (adding a time variable)

$$A(\mathbf{y}, \mathbf{x}; \Delta t) = \min\left(\frac{\mathbf{p}^*(\mathbf{y}, t + \Delta t)}{\mathbf{p}^*(\mathbf{x}, t)}, 1\right)$$

- A simple algorithm

1) Propose a move for each walker k $\mathbf{X}_k^{new} = \mathbf{X}_k^{old} + \eta$;
2) Compute the matrix (or function) A:
3.1) the move is accepted if $A(new, old) \geq 1$;
3.2) the move is also accepted if $A(new, old) < 1$ *but* $> \chi$ *(uniform rnd);*
3.3) the move is otherwise rejected;
4) Update the walkers' position in the given space;
5) Again the point 1) until convergence is achieved.

Monte Carlo integration

Monte Carlo integration

- Monte Carlo integration of definite integrals

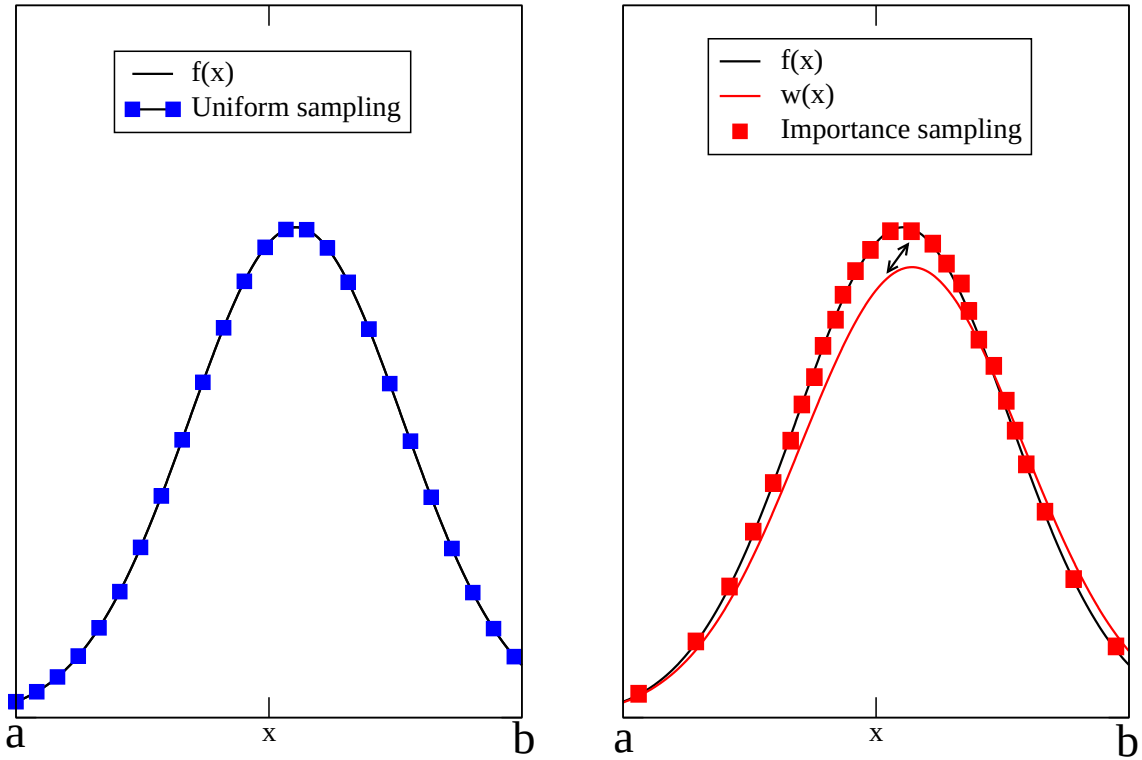
$$
\begin{aligned}
F &= \int_a^b f(x)dx \\
F &= \lim_{N\to\infty} F_N \\
F_N &= \frac{b-a}{N}\sum_{i=1}^{N} f(X_i)
\end{aligned}
$$

- $\{X_i\}$ should fully cover the domain from a to b
- Uniform grid methods $\Rightarrow N^d$ (d dimensionality of the system)
- Gaussian quadrature (for $d \leq 8$)
- Choice of $\{X_i\}$ randomly drawn from a given pdf by Monte Carlo methods

- Uniform sampling: $\{X_i\}$ sampled by an uniform pdf (not efficient!)
- Importance sampling: the majority of sample points "clustered" in the region where the integrand is large
- $w(x) \simeq f(x)$, defined in $[a, b]$
- Generation of (pseudo)random points from

$$
\begin{aligned}
p(x) &= \frac{w(x)}{\int_a^b w(x)dx} \\
F &= \int_a^b g(x)p(x)dx \simeq \frac{1}{M}\sum_{i=1}^{M} g(X_i) \\
g &\equiv f/p
\end{aligned}
$$

- Fluctuations in g (and in F) are greatly reduced (no fluctuation if $w = f$)



- For Monte Carlo methods, if M is the (total) number of sample points, the error $\propto \frac{1}{\sqrt{M}}$
- Independent on the dimensionality!
- Error in grid methods $\propto (\frac{1}{M})^{q/d}$ (q order of the method)
- Monte Carlo more efficient with $d > 2q$

Monte Carlo evaluation of expectation values

- Numerical evaluation of multidimensional integrals (statistical and quantum mechanics)
- $f(\mathbf{x}) \rightarrow$ equilibrium pdf

$$O[f] = \frac{\int d\mathbf{x} O(\mathbf{x}) f(\mathbf{x})}{\int d\mathbf{x} f(\mathbf{x})} \equiv \langle O \rangle_f$$

- Failure of uniform sampling: inefficient
- Generation of sample points based on the integrand
- Monte Carlo sample points $\{\mathbf{X}\}$ drawn from f
- Monte Carlo estimate

$$\langle O \rangle_f = \lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} O(X_i)$$

Simple Metropolis sampling

- The Metropolis algorithm only requires evaluating f for the proposed move
- The unknown normalization $\int d\mathbf{x} f(\mathbf{x})$ is not required
- Convergence to the equilibrium before computing any expectation values
- Step size to optimize the spanning of the system space
- Large step $\rightarrow$ small acceptance ratio $\rightarrow$ actual movement is small
- Small step $\rightarrow$ limited space exploration
- The step size should be optimized empirically according to the behaviour of the sampling algorithm