Lecture 1 Introduction *

Tiejun Li

1 Stochastics: why, what and where

There are still debates on whether the world is deterministic or stochastic. We take a practical point of view on this problem. The reason why we utilize stochastics is as below:

1. The problem itself is stochastic (quantum mechanics).
2. Even the problem is deterministic in nature, the degrees of freedom is too huge to be handled in a deterministic manner (statistical mechanics).
3. The considered problem is in deterministic form, but we utilize its equivalent stochastic form to do computing (Monte Carlo methods).

The course will be composed of three parts:

1. Monte Carlo methods.
2. SDEs and their simulations.
3. Applications.

The main application area of Monte Carlo method:

Statistical Physics, Statistical inference, Mathematical finance, Data Science.

2 Monte Carlo concepts

Example 1. *(Buffon test)*

1. *Parallel lines with distance a in the plane;*
2. *Tossing a needle of length l ($l < a$) randomly;*
3. *Intersection probability?*

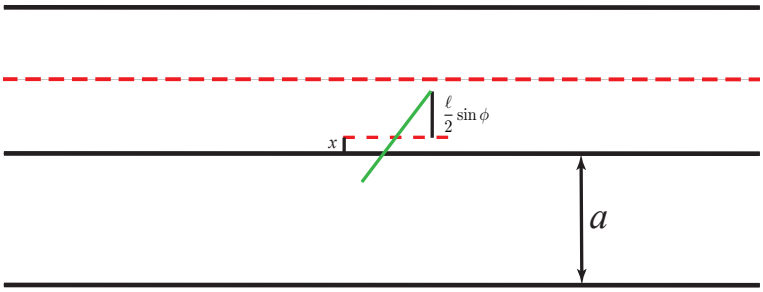

Figure 1: Schematics for Buffon's needle problem.

Solution. This is a geometric probability problem. The admissible set is

$$\Omega := \{0 \le x \le \frac{a}{2}, 0 \le \phi \le \pi\}.$$

The set of intersection is

$$G = \{x \le \frac{l}{2}\sin\phi\},$$

then the probability of intersection

$$P = \frac{meas(G)}{meas(\Omega)} = \Big(\int_0^\pi \frac{l}{2}\sin\phi d\phi\Big)\Big/\Big(\frac{a\pi}{2}\Big) = \frac{2l}{a\pi},$$

thus

$$\pi = \frac{2l}{aP}.$$

Another choice (taking into account more symmetry):

$$\Omega := \{0 \le x \le \frac{a}{2}, 0 \le \phi \le \frac{\pi}{2}\}.$$

and

$$G = \{x \le \frac{l}{2}\sin\phi\},$$

we also have

$$P = \frac{2l}{a\pi},$$

Example 2. *(Monte Carlo integration) Numerically solve*

$$I(f) = \int_0^1 f(x)dx.$$

1. Midpoint rule:

$$I_N^{(1)}(f) = h\sum_{i=1}^N f(x_i), \qquad h = \frac{1}{N},\ x_i = (i+\frac{1}{2})h$$

*School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China

Accuracy: $O(h^2)$.

2. Monte Carlo:

$$I_N^{(2)}(f) = \frac{1}{N}\sum_{i=1}^{N} f(X_i), \qquad X_i \sim i.i.d.\ \mathcal{U}[0,1]$$

one has $\mathbb{E}I_N^{(2)}(f) = I(f)$, and the mean square error

$$\begin{aligned}\mathbb{E}|e_N|^2 = \mathbb{E}(I_N^{(2)}(f) - I(f))^2 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}(f(X_i) - I(f))\right)^2 \\ &= \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbb{E}(f(X_i) - I(f))(f(X_j) - I(f)) \\ &= \frac{1}{N}\mathbb{E}(f(X_i) - I(f))^2 = \frac{1}{N}\mathrm{Var}\,(f),\end{aligned}$$

One obtains $e_N \sim \sqrt{\frac{\mathrm{Var}\,(f)}{N}} \sim O(h^{\frac{1}{2}})$ — half order convergence. (How to generate X_i?)

The above derivations are independent of dimensions.

3. High dimensional case:

Ensemble average in statistical mechanics

$$\langle A\rangle = \frac{1}{Z}\int_{R^{6N}} A(x)e^{-\beta H(x)}dx$$

where $Z = \int_{R^{6N}} e^{-\beta H(x)}dx$ is partition function, $\beta = (k_B T)^{-1}$, k_B is Boltzmann constant, T is the absolute temperature, $dx = dx_1\cdots dx_N dp_1\cdots dp_N$, N is the number of particles.

Deterministic quadrature: 10 segments in each direction, totally 10^{6N} nodes!

Monte Carlo method is the only viable approach!

4. Estimate of computational effort:

Dimension — d, ♯ of quadrature points — N

Midpoint rule $\sim O(N^{-\frac{d}{2}})$, Monte Carlo $\sim O(N^{-\frac{1}{2}})$.

If $d > 4$, Monte Carlo is better.

5. Brief summary:

The advantage of Monte Carlo:

- Half order convergence independent of dimensions;
- Parallel essentially;
- Versatile: If we can find a probabilistic interpretation of a problem, we can apply MC.

The disadvantage of Monte Carlo:

- Half order convergence (slow convergence);
- Noisy result.

3 Further applications

Example 3 (Randomized linear algebra)**.** *Compute the matrix product*

$$C = AB,$$

where $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, *and assume* $n \gg 1$.

When n is huge, which is possible in many applications in big data, the following randomized matrix multiplication was proposed:

Given any probability distribution $\{p_i\}$, *where* $p_i > 0$ *and* $\sum_{i=1}^{n} p_i = 1$, *randomly pick* K *columns with the* i_m*th column from* A, $L^{(m)}$ *and the* i_m*th row from* B, $R^{(m)}$ *according to* $\{p_i\}$. *Correspondingly define*

$$L^{(m)} = \frac{1}{\sqrt{Kp_{i_m}}} A_{\cdot,i_m}, \quad R^{(m)} = \frac{1}{\sqrt{Kp_{i_m}}} B_{i_m,\cdot}, \quad m = 1,\dots,K$$

then compute

$$C \approx \sum_{m=1}^{K} L^{(m)} R^{(m)}. \tag{1}$$

Does it work? Is it possible to generalize and improve it?

Example 4. *(Bayesian methods in statistical learning) Sampling the posterior distribution of the unknown parameters* θ.

In statistics, we have large amount of sampling data, and we want to extract the parameters from some type of probabilistic model. Suppose we have the *likelihood* function

$$L(\boldsymbol{\theta}|\boldsymbol{x}), \quad \boldsymbol{\theta} \in \Theta,$$

and the prior distribution of the parameter θ is $\pi(\boldsymbol{\theta})$, we would like to sample the posteriori distribution of θ

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto L(\boldsymbol{\theta}|\boldsymbol{x})\pi(\boldsymbol{\theta})$$

or compute the expectation of the parameters. Usually $\boldsymbol{\theta}$ is in a high dimensional space, and $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is only known up to a constant. We need the Monte Carlo sampling method here.

Example 5. *(Simulated annealing for optimization)* $\min_x H(x)$, *H(x) is an energy function.*

1. If $H(x)$ is convex, the problem is quite easy by steepest decent method

$$\frac{dx}{dt} = -\nabla H$$

2. If $H(x)$ is non-convex, the problem is complicate. The solution by steepest descent will fall into a local minimum generally.

3. Introduce thermal noise

$$\frac{dx}{dt} = -\nabla H + \epsilon \dot{w}$$

$\epsilon \sim$ temperature. Let $\epsilon \to 0$ with suitable speed, one can achieve the global minimum.

Example 6. *(Harmonic oscillator with random forcing) How to describe the noise mathematically? (Potential* $U(\boldsymbol{x}) = \frac{1}{2}k\boldsymbol{x}^2$*)*

1. Conservative harmonic oscillator

$$\begin{cases} \dot{\boldsymbol{x}} &= \boldsymbol{v} \\ m\dot{\boldsymbol{v}} &= -k\boldsymbol{x} \end{cases}$$

2. Frictional harmonic oscillator (frictional coefficient γ)

$$\begin{cases} \dot{\boldsymbol{x}} &= \boldsymbol{v} \\ m\dot{\boldsymbol{v}} &= -\gamma\boldsymbol{v} - k\boldsymbol{x} \end{cases}$$

3. White noise forcing (mesoscopic particles)

$$\begin{cases} \dot{\boldsymbol{x}} &= \boldsymbol{v} \\ m\dot{\boldsymbol{v}} &= -\gamma\boldsymbol{v} - k\boldsymbol{x} + \sqrt{2k_BT\gamma}\dot{\boldsymbol{w}} \end{cases}$$

$\dot{\boldsymbol{w}}$ is the temporal white noise. How to define $\boldsymbol{w}$?

Example 7. *(First exit time) Connection with PDEs.*

Solving the elliptic PDE

$$\begin{cases} \Delta u &= 0 & D \\ u &= f & \partial D \end{cases}$$

Traditional method: FEM, FD

$$u(\boldsymbol{x}) = \mathbb{E}\big(f(\boldsymbol{X}_{\tau_D})\big)$$

where $\boldsymbol{X}_{\tau_D}$ is the first exit point form ∂D of the Brownian motion starting at $\boldsymbol{x} \in D$.

One can compute the value of u at any point in Ω separately.

Example 8. *(Particle system) Macroscopic behavior from microscopic movements*

1. Deterministic case(without interaction): Liouville equation.

$$\frac{d\boldsymbol{x}_i}{dt} = \boldsymbol{b}(\boldsymbol{x}_i) \;\longrightarrow\; \psi_t + \nabla \cdot (\boldsymbol{b}\psi) = 0$$

2. Stochastic case(without interaction): Fokker-Planck equation.

$$\frac{d\boldsymbol{x}_i}{dt} = \boldsymbol{b}(\boldsymbol{x}_i) + \dot{\boldsymbol{w}}_i \;\longrightarrow\; \psi_t + \nabla \cdot (\boldsymbol{b}\psi) = \frac{1}{2}\Delta\psi$$

3. Stochastic case (with interaction): Mckean-Vlasov equation.

$$\frac{d\boldsymbol{x}_i}{dt} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{b}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \dot{\boldsymbol{w}}_i \;\longrightarrow\; \psi_t + \nabla \cdot (\boldsymbol{U}\psi) = \frac{1}{2}\Delta\psi$$

where $\boldsymbol{U} = \int \boldsymbol{b}(\boldsymbol{x} - \boldsymbol{y})\psi(\boldsymbol{y})d\boldsymbol{y}$.

Example 9. *(Chemical reaction kinetics) Stochastic simulation algorithm.*

Traditional modeling of chemical reaction: reaction rate equation (RRE):

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{a}(\boldsymbol{x}) \tag{2}$$

where $\boldsymbol{x}$ is the concentration of the reactants, $\boldsymbol{a}$ is the reaction rate. In biological reactions, the population of some species are very few. The concept concentration does not make any sense there. The reaction also shows the random character. How to model the chemical reaction kinetics?

Example 10. *(DLA model) Fractal growth of crystallization. (See Fig. 2)*

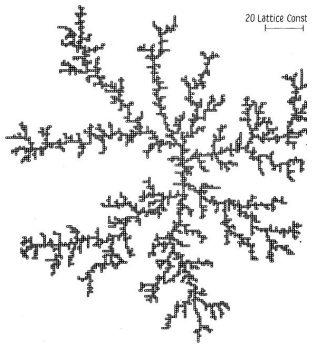

Figure 2: DLA model. Adapted from PRL 47(1981), 1400.

Example 11. *(Complex fluids) Such as the suspensions, colloids and liquid crystals, etc.*

How to describe the behavior of the fluids through describing the polymers?

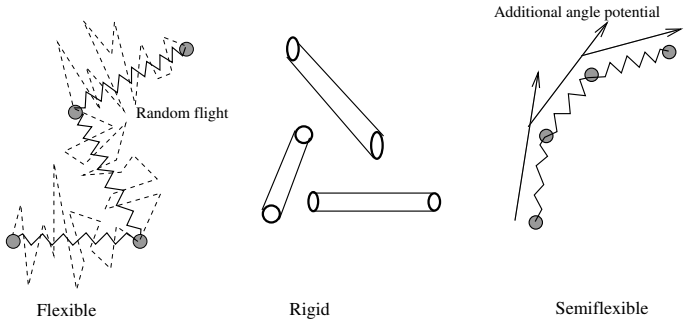

Figure 3: Schematics of flexible, semi-flexible and rigid polymers.

4 Course plan

The following topics will be covered in this course:

- Generation of pseudo random variables,
- Variance reduction methods,
- Simulated annealing and quasi-Monte Carlo,
- Large deviation principle,
- Metropolis algorithm (Markov chain Monte Carlo method),
- Multilevel sampling and kinetic MC,
- Wiener Process and its construction,
- Stochastic differential equations and Ito's formula,
- Fokker-Planck equation and diffusion process,
- Numerical solution of SDEs,
- Path integral methods and Girsanov transformation,
- Applications in material science(rare events),
- Applications in biology,
- Applications in networks,
- Applications in fluids.

We will have 2 numerical projects which will account for 15 pts. The homeworks will account for 15 pts, and the final exam will account for 70 pts.

5 Main references

- W.H. Press et al., Numerical Recipes: the Art of Scientific Computing, Cambridge university press, Cambridge, 1986.

- R.E. Caflish, Monte Carlo and Quasi-Monte Carlo methods, Acta Numerica, Vol. 7, 1-49, 1998.
- P. Glasserman, Monte Carlo methods in financial engineering, Springer-Verlag, New York, 2003.
- C.P. Robert and G. Casella, Monte Carlo Statistical methods, Springer-Verlag, New York, 2004.
- C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences, Springer-Varlag, Berlin, New York, 1983.
- B. Oksendal, Stochastic Differential Equations: an Introduction with Applications, Springer-Verlag, Berlin Heidelberg New York, 2003(6th edition).
- I. Karatzas and S.E. Shreve, Brownian motion and stochastic calculus, Springer-Verlag, New York, 1991.
- P.E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin and Heidelberg, 1992.

6 Homeworks

- Discuss about the method (1). Why is it a valid method? How to characterize its accuracy?
- Show that the midpoint rule has second order convergence if $f \in C^2[0,1]$.
- Numerically testify the half order convergence of Monte Carlo integration for

$$I(f) = \int_0^1 \sin x dx = \mathbb{E} \sin X$$

where X is uniformly distributed in $[0,1]$.