Lec 1 Introduction

Tuesday, February 27, 2024 11:48 AM

0. Angxiu Ni: 双清综合楼 B618

13810493439. 建微信群

TA: 黄崇致

survey: who studied probability (s)?

who studied numerical analysis?

1. Randomness: why, what, where

In practice, why use randomness:

- The problem is random
- We can state symmetry
- $(\Omega, \mathcal{B}, \mathbb{P})$ allows to detect false models
- Randomness is the easiest to understand / generate "universality".

  Universality + observable overcomes curse of dim.

In fact, we might only need universality, not randomness.

1.1. Fun Fact: inadmissible symmetry.

Assume two envelops, one contains twice money as the other.

Once I pick one envelop, should I change? Then change again?

Why absurd consequence?

Assume $\mathbb{P}(X_1=1)=p$, The symmetry we want is $\mathbb{P}(X_2=2X_1 \mid X_1)=\frac{1}{2}$

$\mathbb{P}(X_2=2, X_1=1) = \mathbb{P}(X_2=2 \mid X_1=1)\cdot\mathbb{P}(X_1=1) = p$.

Assume $\mathbb{P}(X_1=1)=p$, The symmetry we want is $\mathbb{P}(X_2=2 \mid X_1)=\frac{1}{2}$

$$\mathbb{P}(X_2=2, X_1=1) = \mathbb{P}(X_2=2 \mid X_1=1)\cdot \mathbb{P}(X_1=1) = \frac{p}{2}.$$

$$\parallel$$

$$\mathbb{P}(X_1=1 \mid X_2=2)\cdot \mathbb{P}(X_2=2) = \frac{1}{2}\cdot \mathbb{P}(X_2=2)$$

$$\Rightarrow \mathbb{P}(X_2=2)=p$$

Similarly, $\mathbb{P}(X_1=4)=\mathbb{P}(X_1=16)=\cdots=\mathbb{P}(4^n)=p \quad \forall n.$

So we can not have a probability measure.

Lesson: Be careful when symmetry $\rightarrow$ procedure

infer symmetry from a well-defined procedure is safer but hard to notice.

1.2. Vague syllabus

(a) Monte Carlo method

(b) SDEs and their simulations

(c). Applications.

(d) Some fun facts on probability

2. Monte Carlo concept

- Typycally MC means integration, sometimes only refers to generating samples.

2.1. Definition of MC.

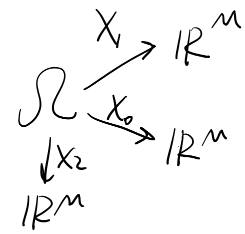
Want $\mathbb{E}[f(X)]$. r.v. $X: \Omega \to \mathbb{R}^d$, $f: \mathbb{R}^d \to \mathbb{R}$ smooth

$$\mathbb{E}[f(X)] \overset{a.s.}{=} \lim \frac{1}{N}\sum_{n=1}^{N} f(X_n)$$

$$\mathbb{E}[f(X)] \overset{a.s.}{=} \lim \frac{1}{N}\sum_{n=1}^{N} f(X_n)$$

MC.

If X_n's are i.i.d copies of X.



a.s. is w.r.t. the overall probability $\mathbb{P}$

on Ω, $\underbrace{\text{or the product of } \mathbb{P}_{X_0} \text{ on } \bigotimes_{n=1}^{\infty} \mathbb{R}^M}_{\text{this is typically better.}}$.

By CLT, the convergence is $\frac{1}{\sqrt{N}}\sum f(X_n) \xrightarrow{d} N(0, \operatorname{var} f(X))$

Why this works in high-dimension $M \gg 1$?

a). $f: \mathbb{R}^M \to \mathbb{R}$ does a dimension reduction

b) $F := f(X)$ has rich details as a r.v. but CLT (universality)

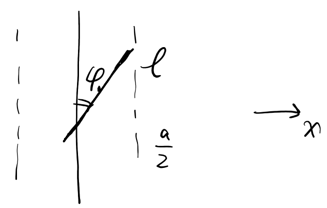
says we can forget details.

2.2. Compute probability of a certain set G.

e.g. use $\mathbb{P}(A)$ to compute π in stone-age.

(Buffon needle)

$\Omega := \{0 \le x \le \frac{a}{2},\ 0 < \varphi \le \pi\}$



$\mathbb{P}(\cdot) \propto Leb(\cdot)$

$G := \{x \le \frac{l}{2}\sin\varphi\}$ interpretation by tossing a needle.

$$\mathbb{P}(G) = \frac{Leb(G)}{Leb(\Omega)}$$

$$Leb(G) = \int_0^{\pi} d\varphi \int_0^{\frac{l}{2}\sin\varphi} dx = \int_0^{\pi} \frac{l}{2}\sin\varphi\, d\varphi = l$$

$$\mathrm{Leb}(G) = \int_0 d\varphi \int_0 dx = \int_0 \frac{l}{2}\sin\varphi \, d\varphi = l$$

$$\mathrm{Leb}(\Omega) = \int_0^{\pi} d\varphi \int_0^{\frac{a}{2}} dx = \frac{a\pi}{2}$$

$$\Rightarrow \mathbb{P}(G) = \frac{2l}{a\pi} \Rightarrow \pi = \frac{2l}{ap}$$

But we can compute

$$p = \mathbb{P}[(X,\Phi) \in G] = \mathbb{E}_{X,\Phi}[\mathbb{1}_G] \underset{\uparrow \atop MC}{\overset{a.s.}{=}} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_G(X_i, \Phi_i)$$

Here a.s. is w.r.t. $\prod_{i=1}^{\infty} \{ [0,\frac{a}{2}] \times [0,\pi], \mathbb{P} \}$

This requires i.i.d sampling of $(X, \Phi)_i$. CLT of iid says

the convergence is fast. But we can have ergodic thm.

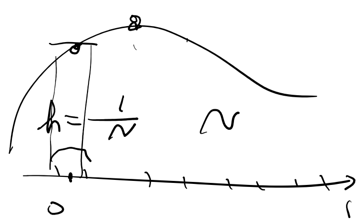
Why this can be done in stone age? $\mathbb{1}_G(x_i, \varphi_i)$ can be

determined by whether needle crosses a line.

2.3. Error analysis.

Want $I[f] := \int_{[0,1]} f(x)\,dx$

(1) Quadrature "finite element"



$$I_N^{(1)}(f) = \sum_{n=0}^{N-1} f\left(\frac{n+\frac{1}{2}}{N}\right) \cdot \frac{1}{N}$$

Error $\quad |I_N^{(1)} - I| \le C \cdot |f''|_{\sup} h^2 = O(h^2)$ prove by Taylor expansion

For dimension $d>1$, $\int_{\Pi^d} f(x)\,dx$, if we still have N samples.

then $N^{\frac{1}{d}}$ cells each direction, so $h = N^{-\frac{1}{d}}$

$\text{Error} = O(N^{-\frac{2}{d}}) \quad \nearrow$ w.r.t. d.

Cost: $N \cdot d$ for d-dim.

$\uparrow$ # samples $\quad \uparrow$ each sample is $\mathbb{R}^d$ even to look at.

(2). MC.

$$I_N^{(2)}(f) = \frac{1}{N}\sum_{n=1}^{N} f(X_n), \quad X_n \overset{iid}{\sim} U[0,1]$$

$$e_N^{(2)} = \left(\frac{1}{N}\sum f(X_n)\right) - \underset{\underset{I[f]}{\parallel}}{\bar{f}} = \frac{1}{N}\sum_{n=1}^{N}(f(X_n) - \bar{f})$$

$$\mathbb{E}\, e_N^{(2)} = 0$$

$$\operatorname{Var} e_N^{(2)} = \mathbb{E}(e_N^{(2)})^2 = \frac{1}{N^2}\mathbb{E}\left[\left(\sum_n f_n - \bar{f}\right)^2\right] \underset{\underset{\mathbb{E}[(f(X_i)-\bar{f})\cdot(f(X_j)-\bar{f})]=0}{\uparrow}}{=} \frac{1}{N^2}\sum \mathbb{E}(f_n - \bar{f})^2 = \frac{1}{N}\mathbb{E}(f(X)-\bar{f})^2$$

$$= \frac{1}{N}\operatorname{Var} f$$

So $e_N^{(2)} \sim \sqrt{\frac{\operatorname{Var} f}{N}} \sim O(N^{-\frac{1}{2}})$

$\uparrow$ mean $e_N^{(2)} \sim N\left(0, \sqrt{\frac{\operatorname{Var} f}{N}}\right)$ by CLT. More generally, means

$\mathbb{P}(e_N^{(2)}$ far from $\sqrt{\frac{\operatorname{Var} f}{N}})$ is small for large N,

e.g. $\mathbb{P}(|e_N^{(2)}| \geq 1) \leq C \cdot \lambda^{-N}$

CLT is much sharper than Chebyshev / Markov $\leq$

In high dimension $d > 1$, the error analysis still works, since we are only working with $F \in \mathbb{R}$. 1-d r.v.!!

Cost: $O(N \cdot d)$.

(3) Cost comparison. for same cost $O(N\cdot d)$, N: # sample.

error of "finite element" $O(N^{-\frac{2}{d}})$

error of MC: $O(N^{-\frac{1}{2}})$

$N^{-\frac{1}{2}} < N^{-\frac{2}{d}} \iff -\frac{1}{2} < -\frac{2}{d} \iff d > 4$ MC is faster.

(4). Many real life examples $d \gg 1$

d is the # of interacting particles $\times$ degree of freedom of each particle.

Ensemble average in statistical mechanics

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

$$y = (x_1^1, x_1^2, x_1^3, p_1^1, p_1^2, p_1^3, x_2^1, \cdots) \in \mathbb{R}^{6M}$$

$M :=$ # interacting particles.

If we want to do quadrature, we need $N = Q^{6M}$

here Q is the # cells each direction in $\mathbb{R}^{6M}$.

(5). Summary.

MC · $\frac{1}{2}$ convergence slow

· # samples ind. of dim, cost $\propto$ dim

· parallel.

3. Other applications. of MC.

3.1. Beysian inference.   data x parameter θ

Know or assume: $\mathbb{P}(X \mid \theta)$, $\mathbb{P}(\Theta)$

Want: $\mathbb{P}(\theta \mid X = x)$ which is a function of θ

meaning: now given x, what is our knowledge on θ?

$$\mathbb{P}(\theta \mid X = x) = \frac{\mathbb{P}(X \mid \theta) \cdot \mathbb{P}(\theta)}{\mathbb{P}(X)}$$

not function of θ, just a constant once X is fixed.

So $\mathbb{P}(X)$ is just a normalizing constant.

$$\mathbb{P}(X) = \int \mathbb{P}(X \mid \theta) \, \mathbb{P}(\theta) \, d\theta$$

Can Run MC ~~to~~ sample $\theta_n \overset{iid}{\sim} \mathbb{P}_\theta$, then.

$$\mathbb{P}(X) \approx \frac{1}{N} \sum_{n=1}^{N} \mathbb{P}(X = x \mid \Theta = \theta_n).$$

3.2. Solving high-dim Elliptic PDE

$$\begin{cases} \Delta u = 0 & \text{in } D \in \mathbb{R}^d. \\ u = f & \text{on } \partial D \end{cases}$$

Finite element expensive when $d \gg 1$

Feyman-Kac:

$$u(x) = \mathbb{E}_B[f(X_{T_D})]$$

$\{X_t\}$ is Brownian motion starting from x.

T_D is the first exist time from D.

MC $\Rightarrow$ $u(x) \approx \frac{1}{N}\sum_{n=1}^{n} f(x_{t_D})$

But we need solve for one $u(x)$ at a time.

Of course we do, more info needs more cost!

if we want $u(x)$ at all x, then still exponential cost.

4. Applications of SDEs. in modelling

4.1. Harmonic oscillator $\quad H := \frac{p^2}{2m} + \frac{1}{2}kx^2$

$$\begin{cases} \dot{x} = p/m = H_p \\ \dot{p} = -kx = -H_x \end{cases}$$

Then $\frac{dH}{dt} = (H_x, H_p)\cdot\frac{dy}{dt} = (H_x, H_p)\cdot(H_p, -H_x) = 0$

conserves energy (Hamiltonian)

Then we add noise $\downarrow H$

white noise forcing $\uparrow H$.

$$\dot{x} = H_p$$

$$\dot{p} = -H_x - \underbrace{\gamma}_{\text{friction coefficient.}} \frac{p}{m} + \sqrt{2k_BT\gamma}\ \underbrace{\dot{w}}_{\text{white noise in time}}$$

4.2. Particle system

$dx = b(x)\,dt + dB_t$. $\quad$ let $X_t \sim$ density φ_t

$\varphi_t + \nabla\cdot(b\varphi) = \frac{1}{2}\Delta\varphi$.

Fokker - Plank.

Coding example.

- compute $\int_0^1 \sin x \, dx$.

Jupyter lite: online

Cocalc: chatgpt support.

Anaconda: local.

Vim + Linux submachine: me.

Hw:

- Show the midpt rule has second order convergence if $f \in C^2[0,1]$

Pen and paper

- Numerically compute $\int_{[-1,1]^5} e^{-x^2} dx$. by MC.

Also plot the convergence wrt # samples

Recruiting for two projects:

1. "time reparametrization for perturbing SDEs"

needs: prob (3). know convergence proof of Ito

2. "optimal response computation"

needs: can solve $\Delta u = 0$ by finite difference method.