TAM 598 Lecture 8 :

THE MONTE CARLO METHOD FOR ESTIMATING EXPECTATIONS

Announcements:

- HW 2 covers lectures 4-8 ; due on Feb 26

Our problem today: high dimensional integrals

say we have a random vector $\underset{\sim}{X} = \{x_1, x_2, \ldots, x_d\} \in \mathbb{R}^d$
and some function $g(\underset{\sim}{X})$.

things we want to compute:

expectation $\mathbb{E}[g(\underset{\sim}{X})] = \int \cdots \int g(x_1, \ldots, x_d)\, p(x_1, \ldots, x_d)\, dx_1 \ldots dx_d$

$$= \underline{\int \cdots \int g(\underset{\sim}{x})\, p(\underset{\sim}{x})\, d\underset{\sim}{x}}$$

variance $\mathbb{V}[g(\underset{\sim}{x})] = \mathbb{E}\left[\left(g(\underset{\sim}{x}) - \mathbb{E}[g(\underset{\sim}{x})]\right)^2\right]$

$$= \int \cdots \int \left(g(\underset{\sim}{x}) - \mathbb{E}[g(\underset{\sim}{x})]\right)^2 p(\underset{\sim}{x})\, d\underset{\sim}{x}$$

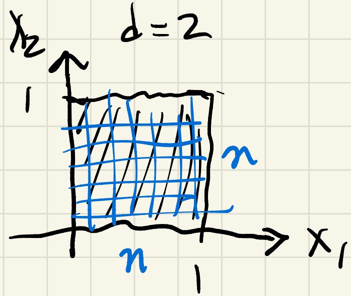
$\mathbb{1}_A(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \underset{\sim}{x} \in A \\ 0 & \text{else} \end{cases}$

CDF $F(\underset{=}{x}) = p[g(\underset{\sim}{Y}) \le \underset{\sim}{x}] = \int \cdots \int \mathbb{1}_{[-\infty, \underset{\sim}{x})}\, g(\underset{\sim}{Y})\, p(\underset{\sim}{Y})\, d\underset{\sim}{Y}$

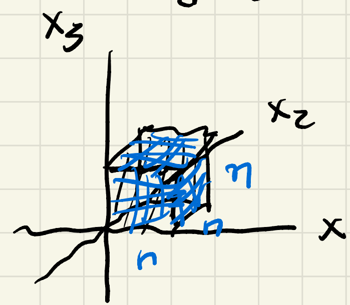
CURSE OF DIMENSIONALITY - why high dim. integrals are hard

let $\underset{\sim}{X} = \{ X_1 \dots X_d \}$ be uniform in $[0,1]^d$ $\quad$ d = dimension

d = 2



d = 3



want to evaluate

$$\mathbb{E}[g(\underset{\sim}{x})] = \underbrace{\int \cdots \int}_{\text{inside cube}} g(x_1 \dots x_d)\, dx_1 \dots dx_d$$

use quadrature

$$\int \cdots \int g(\underset{\sim}{x})\, d\underset{\sim}{x} \approx \frac{1}{n^d} \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} g\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right)$$

terms in sum $= n^d$

grows exponentially w/ d

say $n=10$	$d=2$	$d=3$	$d=5$	$d=10$	$d=20$
	$10^2 = 100$ terms	1000	100,000	10^{10}	10^{20}
if each term requires 0.1 ms	0.1 s	1 s	100 s	115 days	3.1 billion years

LAW OF LARGE NUMBERS

Say we have an infinite series of independent random variables $X_1, X_2, \ldots$ all with the same distribution.

We call this an independent, identically distributed (i.i.d) sequence

Let $\mu = \mathbb{E}[X_i]$ be the common mean of the random variables

$$\boxed{\text{Then } \tilde{X}_N = \frac{X_1 + X_2 + \ldots X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i \text{ converges to } \mu \text{ as } N \to \infty}$$

MONTE CARLO Ulam, von Neumann (1940s)

KI 4ú44

want to compute expectation

$$\int g(\underset{\sim}{x})\, p(\underset{\sim}{x})\, d\underset{\sim}{x} = \mathbb{E}_{p(\underset{\sim}{x})}\left[g(\underset{\sim}{x})\right]$$

$p(\underset{\sim}{x})$ = "target distribution"

$$\text{OR} \quad \sum_{\underset{\sim}{x}} g(\underset{\sim}{x})\, p(\underset{\sim}{x}) = \mathbb{E}_{p(\underset{\sim}{x})}\left[g(\underset{\sim}{x})\right] \quad \text{for discrete}$$

Say we can draw independent samples $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \dots \underset{\sim}{x}_N \sim p(\underset{\sim}{x})$

Then

$$\boxed{\int g(\underset{\sim}{x})\, p(\underset{\sim}{x})\, d\underset{\sim}{x} \approx \boxed{\frac{1}{N}\sum_{n=1}^{N} g(\underset{\sim}{x}_n)} = I_N}$$

Monte Carlo estimator of $\mathbb{E}$

from the law of large numbers, the sum converges to $\mathbb{E}_{p(\underset{\sim}{x})}\left[g(\underset{\sim}{x})\right]$ for large N

say we want to estimate not only the expectation of a distribution by sampling, but also the variance. We can do this by deriving an expression for a variance estimator.

$$\mathbb{V}_{p(\underset{\sim}{x})}[g(\underset{\sim}{x})] = \underbrace{\mathbb{E}_{p(\underset{\sim}{x})}[g^2(\underset{\sim}{x})]} - \underbrace{\mathbb{E}_{p(\underset{\sim}{x})}[g(\underset{\sim}{x})]^2}$$

consider the random variables

$$\underset{\sim}{Y}_1 = g^2(\underset{\sim}{x}_1),\ \underset{\sim}{Y}_2 = g^2(\underset{\sim}{x}_2) \dots\ Y_N = g^2(\underset{\sim}{x}_N)$$

these are iid so

$$\frac{1}{N}\sum_{n=1}^{N} g^2(\underset{\sim}{x}_n)$$

$$\left(\frac{1}{N}\sum_{n=1}^{N} g(\underset{\sim}{x}_n)\right)^2$$

$$\boxed{\mathbb{V}_{p(\underset{\sim}{x})}[g(\underset{\sim}{x})] = \frac{1}{N}\sum_{n=1}^{N} g^2(\underset{\sim}{x}_n) - \tilde{I}_N^2}$$

variance estimator

6

Show that the average of the estimator converges to the expectation.

$$\mathbb{E}_{\{\underset{\sim}{x}_n\}}\Bigg[\underbrace{\frac{1}{N}\sum_{n=1}^{N} g(\underset{\sim}{x}_n)}_{\text{estimator}}\Bigg] = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{\{x_n\}}\left[g(\underset{\sim}{x}_n)\right]$$

$$= \frac{1}{N}\sum_{n=1}^{N} \int\cdots\int \underbrace{p(\underset{\sim}{x}_1')\cdots p(\underset{\sim}{x}_n')\cdots p(\underset{\sim}{x}_N')}\, g(\underset{\sim}{x}_n')\, d\underset{\sim}{x}_1' \cdots d\underset{\sim}{x}_N'$$

because samples are independent

$$= \frac{1}{N}\sum_{n=1}^{N} \underbrace{\int p(\underset{\sim}{x}_n')\, g(\underset{\sim}{x}_n')\, d\underset{\sim}{x}_n'}_{\mathbb{E}_{p(\underset{\sim}{x})}[g(\underset{\sim}{x})]}$$

$$= \mathbb{E}_{p(\underset{\sim}{x})}\left[g(\underset{\sim}{x})\right]$$

3/4
1/4
1/2
TTHT
THHH
n=1 THHT
n=2 HTHT 1/2
n=3
⋮
n=5
N

VARIANCE OF A MONTE CARLO ESTIMATOR

$$\mathbb{V}_{\{\mathbf{x}_n\}}\left[\frac{1}{N}\sum_{n=1}^{N} g(\mathbf{x}_n)\right]$$

$$= \frac{1}{N^2}\sum_{n=1}^{N} \mathbb{V}_{\{\mathbf{x}_n\}}\left[g(\mathbf{x}_n)\right]$$

$$= \frac{1}{N^2}\sum_{n=1}^{N} \mathbb{V}_{p(\mathbf{x})}\left[g(\mathbf{x})\right]$$

$$= \frac{1}{N}\,\mathbb{V}_{p(\mathbf{x})}\left[g(\mathbf{x})\right]$$

our error $\sim \dfrac{1}{\sqrt{N}}$