Lecture 3 -Crude Monte Carlo, importance sampling

Monte Carlo method

Example of a multidimensional integral

$$\int_a^b dx_1 \int_a^b dx_2 \cdots \int_a^b dx_N \; f(x_1, \ldots, x_N)$$

Has to be evaluated.

A determonistic method:

bin each variable $\{x_i\}$ into 100 points

Question? Largest N which can be directly used?

A1 1 Answer memory limit 1GB ~ 10^9 values

Total number of points

$$\begin{array}{ccccc} x_1 & x_2 & \cdots & x_N & \\ 100 \cdot 100 & \cdot & \cdots & \cdot\, 100 & = 100^N = 10^{2N} \end{array}$$

Memory $10^9 \simeq 10^{2N}$; $N \sim 4-5$

A2 whole capacity of internet ~ 10^{24} bytes

$10^{24} \sim 10^{2N}$; $N \simeq 12$

Size of the digital universe

- in 2020 44 zettabytes
- is doubled ever two years
- $1\ ZB = 10^{21}\ B$

→ instead stochastic methods can be used

Crude Monte Carlo method (hit-or-miss meth.)

Crude Monte Carlo method (hit-or-miss meth.)

$$\int_a^b f(x)\,dx = (b-a)\int_a^b f(x)\,p(x)\,dx$$

where $p(x) = \frac{1}{b-a}$ is the uniform distribution

CLT can be applied:

- expectation value of $f(x)$
- with respect to the uniform distribution on $[a; b]$

$$\langle f \rangle = \int_a^b f(x)\,dx = \lim_{N\to\infty} (b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

where x_i are taken from a uniform distribution

- expectation → approximation for the integral

Statistical error:

$$\frac{\sigma^2}{N} = \lim_{N\to\infty} \frac{(b-a)^2}{N}\left[\frac{1}{N}\sum_{i=1}^{N} f^2(x_i) - \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2\right]$$

Monte Carlo method with importance sampling

- sample the important region
- another application of CLT in a form

$$\int f(x)\,p(x)\,dx$$

where $p(x)$ is a PDF

- any integral can be written in the form

$$\int f(x)\,dx = \int \left[\frac{f(x)}{p(x)}\right] p(x)\,dx$$

from MC interpretation:

from MC interpretation:

x in uniform distribution

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i)\Big|_{x_i \text{ uniform}}$$

x in distribution $p(x)$

$$\frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}\Big|_{x_i \text{ accordin to } p(x)}$$

Variance:

$$\int f^2(x)\,dx - \langle f\rangle^2$$

$$\underbrace{\int \left[\frac{f(x)}{p(x)}\right]^2 p(x)\,dx - \langle f\rangle^2}_{\text{can be reduced by a proper choice of } p(x)}$$

Choose $p(x)$ in such a way that is samples more important points, the ones which provide large contribution

Generating a non-uniform PDF

- random values u_i from uniform random distribution can be used to generate a PDF $p(x)$ for which its cumulant function $P(x)$ is known and can be inverted
- $x_i = P^{-1}(u_i)$ are distributed according to $p(x)$

Ex: generate distribution $p(x) = 2xe^{-x^2}, x \geq 0$

$$\int_0^\infty p(x) = 1$$

Its cumulant function

$$P(x) = \int_0^x p(x')\,dx' = \int_0^x 2x' e^{-x'^2} dx = e^{-x'^2}\Big|_0^x$$

$$= -e^{-x^2} + 1$$

Inverse function: $x(u)$

$$u = 1 - e^{-x^2}$$

$$e^{-x^2} = 1 - u$$

$$-x^2 = \ln(1-u)$$

$$x^2 = -\ln(1-u)$$

$$x = \sqrt{-\ln(1-u)}$$

← desired random variable where $0 < u < 1$ is a uniformly distributed random variable

Actually $1-u$ is also a uniform distribution

$\Rightarrow$ $x = \sqrt{-\ln u}$ can be as well used

Normal Random generator

Goal: generate random variables corresponding to Normal PDF:

$$p(x) = \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Box-Muller method, polar coordonates

- use uniform random generator to calculate two random values u_1, u_2: $\quad 0 < u_1 < 1$, $\quad 0 < u_2 < 1$
- calculate $r = \sqrt{-2 \ln u_1}$ - radius
  $\varphi = 2\pi u_2$ - angle
- calculate two independent random variables
  $$z_0 = r \cdot \cos\varphi$$
  $$z_1 = r \cdot \sin\varphi$$
  distributed according PDF: $p(z) = \exp\left\{-\frac{z^2}{2}\right\}$
- adjust the mean value and the variance
  $$x_0 = \mu + z_0 \cdot \sigma$$
  $$x_1 = \mu + z_1 \cdot \sigma$$