Overview

- Sampling / LLN
  - ↳ Computing
    - expected values
    - integrals
- Error in Sampling
- RNG
- Errors in general

Announcements

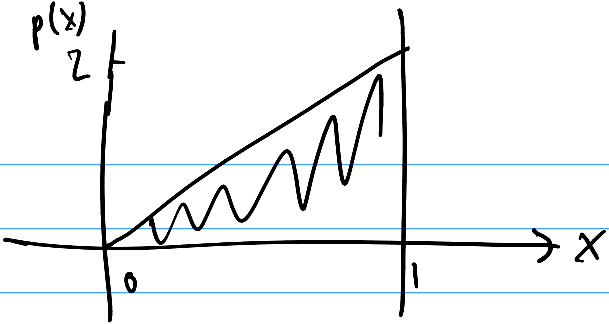
- Quiz 8 : oops
- Exam 1 ed 1 grades: today

LLN:

$$E[X] \approx \frac{1}{N} \sum s_i$$

s_i : samples

from the same distributin as X


p(x)
2
0
1
x

Monte - Carlo algorithm,

- get a random result
- always get a result

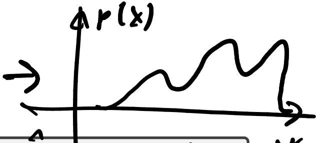
Las Vegas algorithm:

- always computes the correct result (if it computes one)
- but it may only sometimes (randomly) produce a result

Switching Distributions for Sampling

Found: ~~$\frac{1}{N}\sum s_i$~~

$$E[X] = E\left[\tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right] \rightarrow$$



~~draw random numbers from that dist.?~~

Why is this useful for sampling?

$$E[X] = \int x \cdot p(x)\,dx = \int x \cdot \frac{p(x)}{\tilde{p}(x)}\,\tilde{p}(x)\,dx$$

$$= E\left[\tilde{X} \cdot \frac{p(\tilde{X})}{\tilde{p}(\tilde{X})}\right]$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} \tilde{s}_i \cdot \frac{p(\tilde{s}_i)}{\tilde{p}(\tilde{s}_i)}$$

$\tilde{X}$ distributed according to $\tilde{p}$

$\rightarrow \tilde{p}$

Need $\tilde{p} \neq 0$ whenever $p \neq 0$

In-class activity: Monte-Carlo Methods

$$\int g(x)\,dx = \int g(x) \cdot \frac{1}{\tilde{p}(x)}\,\tilde{p}(x)\,dx$$

$$= E\left[\frac{g(\tilde{X})}{\tilde{p}(\tilde{X})}\right]$$

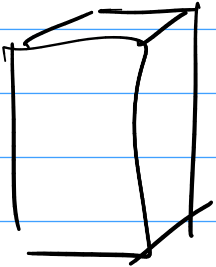
$$= \frac{1}{N}\sum_{i=1}^{N} \frac{g(\tilde{s}_i)}{\tilde{p}(\tilde{s}_i)}$$

$$\int_0^1\int_0^1\int_0^1 f(x,y,z,w)\, dx\, dy\, dz$$

$N \rightsquigarrow 10$

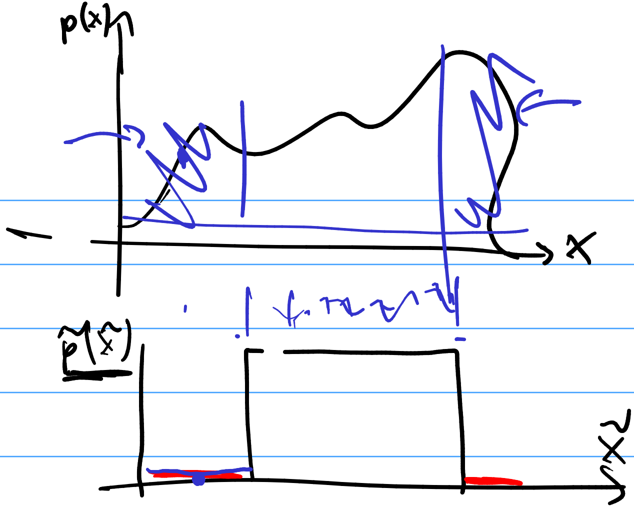
For regular grid: N^3 cost

For MC:

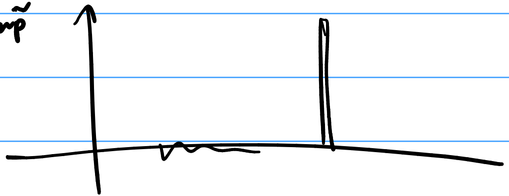


$$\frac{1}{N}\sum_{i=1}^{N} f(\xi_i)$$

cost: $\sim N$ samples



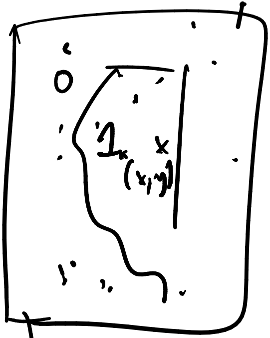
another $\tilde{p}$



Expected Value: Example II

What is the expected snowfall in Illinois?

$$E[\text{Snow}] = \int_0^1 \int_0^1 \text{Snow}(x,y)\, p(x,y)\, dy\, dx$$



$\text{Snow}(x,y)$: Expected snowfall at (x,y)

$p(x,y)$ = prob. with which (x,y) is in IL

$\rightarrow$ $p(x,y) = C \cdot g(x,y)$

$$g(x,y) = \begin{cases} 1 & \text{in IL} \\ 0 & \text{outside of IL} \end{cases}$$

$$C = \frac{1}{\int\int g(x,y)} \leftarrow = \text{Area of IL?}$$

(not quite; need units not (0,1))

Choose $\tilde{p}$ uniform on $[0,1]\times[0,1]$

$$\frac{1}{C} = \int_0^1\int_0^1 g(x,y)\,dx\,dy$$

$$= \int_0^1\int_0^1 \frac{g(x,y)}{\tilde{p}(x,y)}\,\tilde{p}(x,y)\,dx\,dy$$

$$= E\left[\frac{g(\tilde{X},\tilde{Y})}{\tilde{p}(\tilde{X},\tilde{Y})}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{g(\tilde{X}_i,\tilde{Y}_i)}{\tilde{p}(\tilde{X}_i,\tilde{Y}_i)}$$

Dealing with Unknown Scaling

What if a distribution function is only known up to a constant factor, e.g.

$$p(x) = C \cdot \underbrace{\begin{cases} 1 & \text{point } x \text{ is in IL,} \\ 0 & \text{it isn't.} \end{cases}}_{q(x)}$$

Typically $\int_{\mathbb{R}} q \neq 1$. We need to find C so that $\int p = 1$, i.e.

$$C = \frac{1}{\int_{\mathbb{R}} q(x) dx}.$$

Idea: Use sampling.

Demo: Computing π using Sampling
Demo: Errors in Sampling

Sampling: Error

The Central Limit Theorem states that with

$$S_N := X_1 + X_2 + \cdots + X_n$$

for the (X_i) independent and identically distributed according to random variable X with variance σ^2, we have that

$$\frac{S_N - NE[X]}{\sqrt{\sigma^2 N}} \to \mathcal{N}(0,1),$$

i.e. that term approaches the normal distribution. As we increase N, σ^2 stays fixed, so the asymptotic behavior of the error is

$$\left| \frac{1}{N} S_N - E[X] \right| = O\left(\frac{1}{\sqrt{N}} \right).$$