# Lecture 1.

Chp 1.1, 1.2. Liu's Book.

Note Title 4/2/2006

Goal: Teach Monte Carlo methods with some related material on optimization.

Text Book: "Monte Carlo Strategies in Scientific Computing". J.S. Liu.

Grading: 4 Homework Assignments.
+ Final.

Motivation: Many problems in Statistics can be formulated as probabilistic inference, or as optimization.

Find $\hat{\underline{x}} = \underset{\underline{x}}{\text{ARG-MIN}}\, E(\underline{x})$ $\qquad \underset{\underline{x}}{\text{MIN}}\, E(\underline{x}) = E(\hat{\underline{x}})$ —

$$\hat{\underline{x}} = \underset{\underline{x}}{\text{ARG-MAX}}\, P(\underline{x} \mid \underline{d})$$

Or as evaluating an integral

$$I = \int_D g(\underline{x})\, d\underline{x}$$

Monte Carlo gives a way to perform this

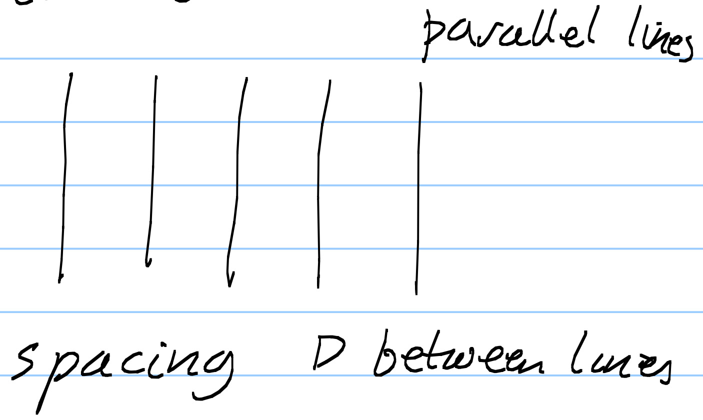
<u>History</u>: Developed after the 2nd world war. The Manhattan project. Computers available.

# Count de Buffon's Needle to Estimate π.

## Early Sampling Experiment.

Needle length $l < D$

Drop needle at random.



probability that needle will intersect a line is $\frac{2l}{\pi D}$. (Check $\frac{2l}{\pi D} < 1$, because $l < D$).

Let $p_N$ be the proportion of "intersects" in $N$ samples (i.e. drop the needle $N$ times, count no. times it intersects line - say $M$, set $p_N = M/N$).

$$\lim_{n \to \infty} p_N = \frac{2l}{\pi D}.$$

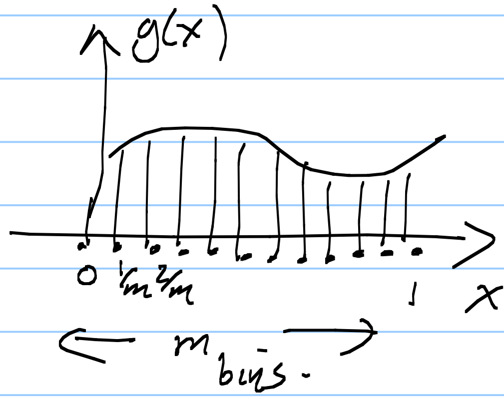
Hence $\pi = \lim_{n \to \infty} \frac{2l}{p_N D}$. //

Note: Sampling has been around for 1,000's years. For example, Banks / Kings would do sampling to estimate their wealth.

# Integration

$$I = \int_0^1 g(x)\,dx$$

$D = [0,1]$



## Riemannian Approximation

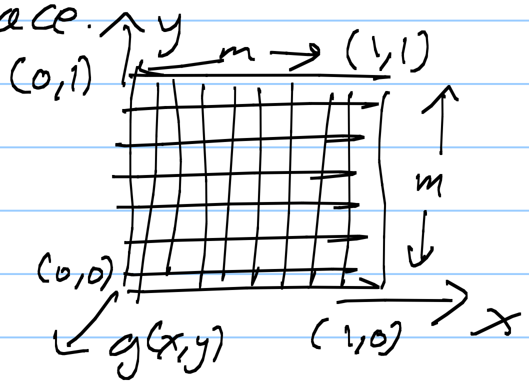
Approximate $I$ by $\tilde{I}_m = \frac{1}{m}\{g(1/m)+g(2/m)+\dots+g(1)\}$

The typical error is $O(m^{-1})$

Not so bad.

As $m \to \infty$, $\tilde{I}_m \to I$.

But no. of bins increases exponentially with the dimensionality of the space.



In 2-dimensions

$D = [0,1]^2$

$$\tilde{I}_m = \frac{1}{m^2}\sum_{\substack{i=1\\j=1}}^{m} g(i/m, j/m)$$

$\tilde{I}_m \to I$, as $m \to \infty$

But requires evaluating $m^2$ points to get error $O(m^{-1})$

In n-dimensions, to get $O(m^{-1})$ error requires evaluating $m^n$ points — too many!

Page 4.

# Monte Carlo approximation.

$I = \int_{D} g(\underline{x})\,d\underline{x}$ , $\underline{x}$ in n-dimensional space

Define a uniform distribution on $D$

$\mu(\underline{x}) = 1/|D|$ , $\underline{x} \in D$. (E.G. if $D = [0,1]^2$

$\int_{D} \mu(\underline{x})\,d\underline{x} = 1.$ $\mu(x,y) = 1/m^2$

$0 \le x \le 1,\ 0 \le y \le 1.$)

Use MC methods to draw

m independent & identically distributed (i.i.d.)

<u>samples</u> from $\mu(\underline{x})$

$\underline{x}^{(1)}, \ldots\ldots\ldots \underline{x}^{(m)}$

Approximate $I$ by $\hat{I}_m = \frac{1}{m}\{g(\underline{x}^{(1)}) + \cdots + g(\underline{x}^{(m)})\}$

By the law of large numbers

$\lim_{m\to\infty} \hat{I}_m = I$ , with probability 1.

$(\forall \epsilon > 0,\ \lim_{m\to\infty} P(|\hat{I}_m - I| > \epsilon) = 0)$

By central limit theorem

$\sqrt{m}\,(\hat{I}_m - I) \to N(0, \sigma^2)$ – error is $O(m^{-1/2})$

a miracle? $\longrightarrow$ $\boxed{\text{independent of space dimension}}$

Where $N(0,\sigma^2)$ is zero mean Gaussian, with variance. $\sigma^2 = \int_{D} d\underline{x}\, \pi(\underline{x}) (g(\underline{x}) - \bar{g})^2$

with $\bar{g} = \int_{D} d\underline{x}\, \pi(\underline{x})\, g(\underline{x})$.

Page 5.

The Monte Carlo approach gives better error rates with far fewer samples 1600

<u>But</u>

(1.) We need to be able to draw samples from $\pi(\underline{x})$ — this is not easy (the main purpose of the course is to show how to sample from probability distributions).

(2.) The variance $\sigma^2$ may be very large. Recall Error $\sim \sigma/\sqrt{m}$.

There are ways to reduce the variance E.G. Importance Sampling.

Get i.i.d samples $\underline{x}^{(1)}, \dots, \underline{x}^{(m)}$ from distribution $\pi(\underline{x})$ that puts more probability on important parts of $\mathcal{D}$.

Estimate $\hat{I}_m = \frac{1}{m} \sum_{j=1}^{m} \frac{g(\underline{x}^j)}{\pi(\underline{x}^j)}$ $\Big|$ $E_\pi\left(\frac{g(\underline{x})}{\pi(\underline{x})}\right) = \int_D g(\underline{x})\, d\underline{x}$

As before $\sqrt{m}\,(\hat{I}_m - I) \to N(0, \sigma_\pi^2)$

with $\sigma_\pi^2 = \mathrm{var}_\pi\left(\frac{g(\underline{x})}{\pi(\underline{x})}\right)$. If $\pi(\underline{x}) = \overset{\swarrow\ \text{constant}}{K}\, g(\underline{x})$

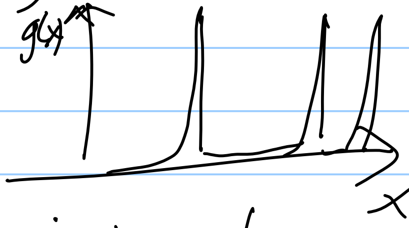
then $\sigma_\pi^2 = 0$, <u>ideal!</u>

(6) Best distribution to sample from is $g(\underline{x})$ — but this may not be possible.

|| Some distributions are easy to sample from but others are very difficult. ||

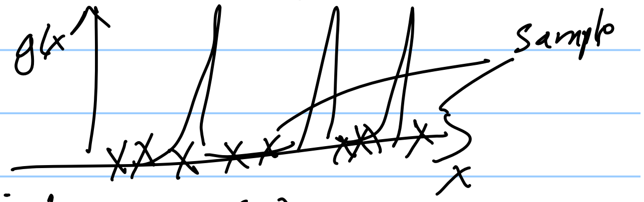
Notice that if $\pi(\underline{x})$ and $g(\underline{x})$ are very different, then the variance $\sigma_\pi^2 = \mathrm{var}_\pi\left(\frac{g(\underline{x})}{\pi(\underline{x})}\right)$ may be huge — so sampling might be even worse than using the Riemmann approximation.

For example, in 1-D suppose $g(x)$ is very 'spikey'



and you sample from a uniform distribution $\pi(x) = 1/|D|$

There is very low probability that the samples from $\pi(x)$ will lie on the spikes — so you estimate of $\int g(x)dx$ will often be bad:



Moral → you need to pick a sampling distribution $\pi(x)$ which is close to $g(x)$.

So the "miracle" of MC — the ability to estimate an integral with error independent of dimension — occurs only if you know a lot about the function $g(x)$ that you want to integrate.