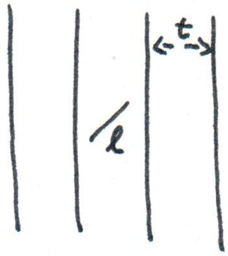
Lecture 3 of AM 207

Verena Kaynig-Fittkau
vkaynig@seas.harvard.edu

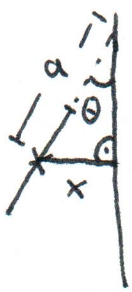
Buffon's needle: Estimating π Monte-Carlo style in 1777



t: distance between two lines

l: length of the needle ($l < t$)

Θ: acute angle of the needle and closest line



x: distance of the needle center to nearest line $\left(0 \leq x \leq \frac{t}{2}\right)$

we know from high school:

$$\sin(\Theta) = \frac{x}{a} \iff a = \frac{x}{\sin(\Theta)}$$

the needle intersects the line if $\frac{l}{2} \geq \frac{x}{\sin(\Theta)}$

pdf of x being anywhere between $[0, \frac{t}{2}]$ is $\frac{2}{t}$.

— " — Θ ——— " ——— $[0, \frac{\pi}{2}]$ is $\frac{2}{\pi}$

x and Θ are independent, so the joint pdf is: $\frac{4}{t\pi}$

The probability of the needle crossing the line is given by the integral of the joint pdf:

$$P = \int_{\Theta=0}^{\frac{\pi}{2}} \int_{x=0}^{\frac{l}{2}\cdot\sin\Theta} \frac{4}{t\pi}\, d\Theta\, dx = \frac{2l}{t\pi}$$

Now we can solve for π: $\pi = \frac{2l}{tP}$

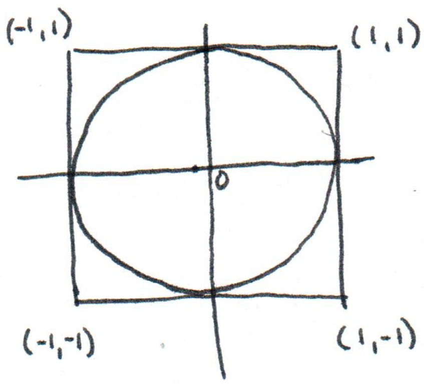
$\Rightarrow$ If we can estimate P, we can estimate π:

$$P = \frac{\text{\# needles crossing line}}{\text{\# needles in total}}$$

$\Rightarrow$ show interactive applet with error preview

$\Rightarrow$ Fun fact: ants seem to estimate area of nesting places using Buffon's needle.

Hit-and-miss method:

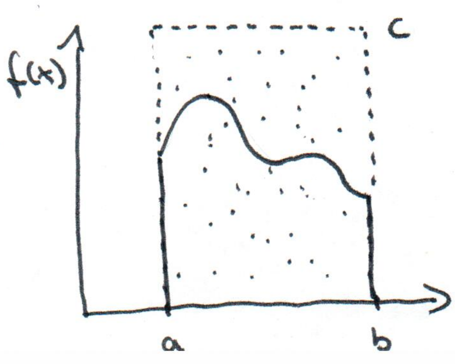


Area of a circle is computed as $A = \pi \cdot r^2$, where r is the radius.

$\Rightarrow$ Area of a circle with radius one is π.

$\Rightarrow$ Estimate π by estimating the area of a unit circle.

Show example or do as in-class exercise.



$$\int_a^b f(x)\,dx = c \cdot (b-a) \cdot \frac{N_{below}}{N_{total}}$$

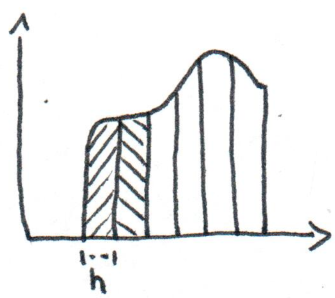
Problems: • all points above the curve are kind of wasted

• If c is too large most points are outside

• If c is too small we might miss part of the relevant area (by cutting it off)

"Crude" Monte Carlo

How we have learned integration in school:



- take regular intervals h
- calculate area of boxes as $h \cdot y$

We then go on to make h smaller, Y better and the overal estimate more precise.

But we can also make h large, if we have the perfect y!

Mean-value-theorem for integrals

If $f(x)$ is continuous over $[a,b]$ then there exists a real number c with $a<c<b$ such that

$$\frac{1}{b-a} \cdot \int_a^b f(x)\,dx = f(c)$$

This means that

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

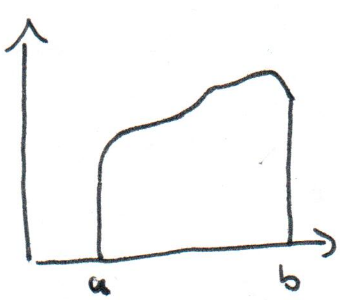
$\Rightarrow$ The area under the curve is the base $(b-a)$ times the "average height" $f(c)$. Instead of $f(c)$ we also write $\langle f \rangle$ or $E(f(x))$.

$\Rightarrow$ If we know $\langle f \rangle$ it is easy to compute $\int_a^b f(x)\,dx$:

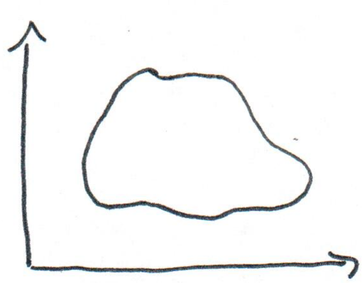
1. Choose random samples x_i from $[a,b]$
2. Compute $f(x_i)$ and $\langle \hat{f} \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$
3. Compute the approximate value of the integral $\int_a^b f(x)\,dx \approx (b-a)\cdot\langle \hat{f} \rangle$

Multi-dimensional case:

compare:



$$\frac{V}{N}\cdot\sum_{i=1}^{N} f(x_i)$$



$$\frac{V}{N}\cdot\sum_{i=1}^{N} f(x_i, y_i)$$

need to make sure that x_i and y_i are inside the area we want to integrate over.

Errors in MC:

From previous experiments we have seen that the error depends on the number of samples. In fact the error gets lower by following $O(\frac{1}{\sqrt{N}})$

By the central limit theorem our error should follow a normal distribution with mean zero. We can estimate the variance by doing multiple experiments and computing a histogram.

For crude MC one can also show that

$$\sigma_{MC}^2 = \frac{\sigma^2}{N} = \frac{\langle f^2 \rangle - \langle f \rangle^2}{N}$$

⇒ This means the error depends on the variance of our function f.

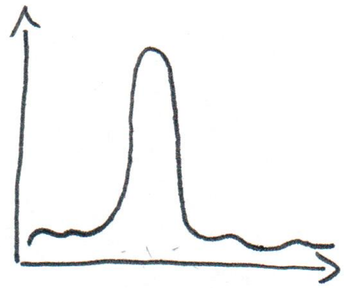
⇒ The error is independent of the dimensionality of f!

This is why MC methods are very populer for high dimensional problems.

We will learn later more about how to reduce the variance of f to come up with better estimates.

Importance Sampling

Idea: Choose the random points such that more points are sampled around the peak of $f(x)$, less where the integrand is small.



$\Rightarrow$ want to draw points where the action is.

$$I = \int_V f(x)\, dx = \int_V p(x) \frac{f(x)}{p(x)}\, dx$$

$p(x)$ is a distribution which is close to $f(x)$, but simple enough that we can sample from it.

remember: $\frac{1}{N} \sum_x p(x) \cdot h(x) = \frac{1}{N} \sum_{x \sim p(x)} h(x)$

$$\Rightarrow I = \lim_{N \to \infty} \frac{1}{N} \sum_{x_i \sim p(x)} \frac{f(x_i)}{p(x_i)}$$

Where is our volume factor? We can omit it because $p(x)$ is a probability density function and therefore normalized.

Intuition: $\frac{f(x_i)}{p(x_i)}$ is flatter than $f(x_i)$, so the variance of $\frac{f(x_i)}{p(x_i)}$ is smaller and so also our error.

Important: • $p(x)$ needs to be normalized

• We need to know how to sample from $p(x)$.

Inverse transform

Idea: transform uniform samples r directly into samples of a different distribution

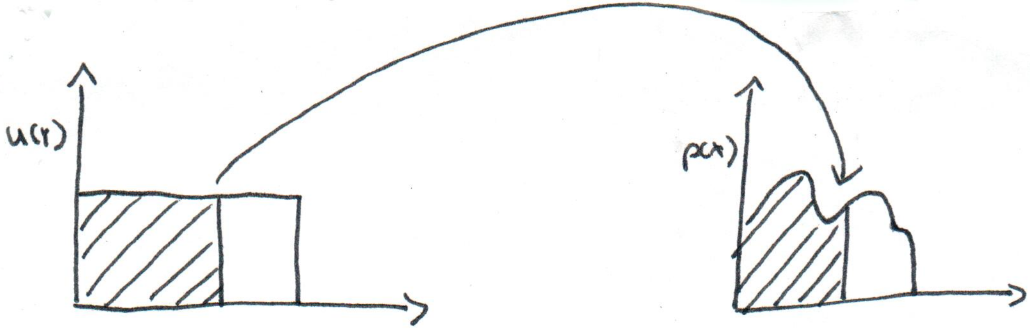
$$r \rightarrow x$$

$$u(r) \qquad p(x)$$

To find the right mapping we look at the cdf:

$$r = \int_0^r u(r')\,dr' = \int_{-\infty}^{x} p(x)\,dx = F(x)$$

The basic idea is that we want to preserve the probability of being smaller than r or respective x.

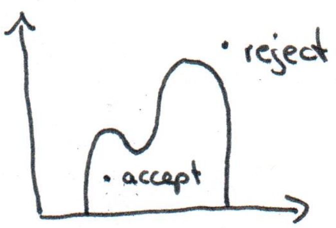


$$r = F(x) \iff x = F^{-1}(r)$$

$\Rightarrow$ To use the inverse transformation we need to be able to get the anti derivative $F(x)$ and to get the inverse $F^{-1}(r)$.

Basic rejection sampling:

- Same as hit-and-miss:



- draw $x \sim U[x_{min}, x_{max}]$
- draw $y \sim U[0, \max(p(x))]$
- if $y < p(x)$: accept the sample
- otherwise reject.

Rejection sampling on steroids:

- Use a proposal density, just like in importance sampling
    - draw $x \sim g(x)$
    - draw $y \sim U[0,1]$
    - if $y < \frac{p(x)}{c \cdot g(x)}$: accept
    - otherwise reject

c: constant factor such that $c \cdot g(x) > p(x)$ for entire domain of interest.