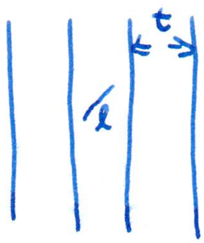
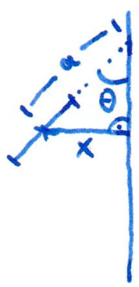
AM207 - Lecture 02 Monte Carlo Integration
Sampling

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



t: distance between two parallel lines

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



θ: acute angle of the needle and the closest line

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

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:

$$P = \int_{\theta=0}^{\frac{\pi}{2}} \int_{x=0}^{\frac{\ell}{2}\cdot\sin(\theta)} \frac{4}{t\pi}\, dx\, d\theta = \frac{2\ell}{t\pi}$$

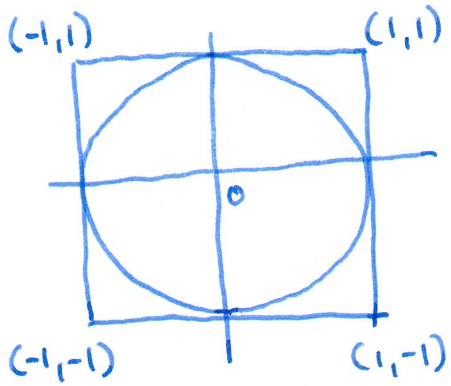
Now we can solve for π: $\pi = \frac{2\ell}{tP}$

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

Frequentist estimate: $P = \frac{\#\text{ needles crossing the line}}{\#\text{ needles in total}}$

Another way of estimating π:

Hit-and-miss method:



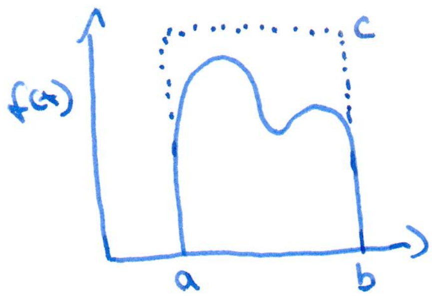
Area of a circle is computed as $A = \pi \cdot r^2$, where r is the radius.

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

$\Rightarrow$ We can sample points from the square area and count how many land inside or outside the circle.

$\Rightarrow$ What happens if we change the size of the square?

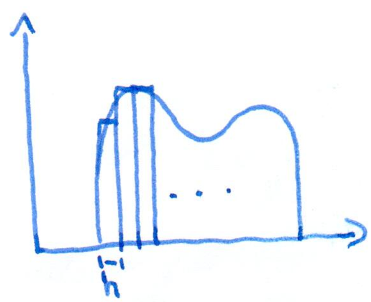
Consider a simpler example:



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

- all points above the curve are kind of wasted
  - ⇒ c too large ⇒ most points are outside
  - ⇒ c too small ⇒ we miss part of the relevant area

Basic numerical integration:



- take regular intervals h
- calculate area of boxes as $h \cdot y$
- make h smaller, y better and the overall estimate more precise.

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

<u>Mean-value-theorem for integrals</u>

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

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

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

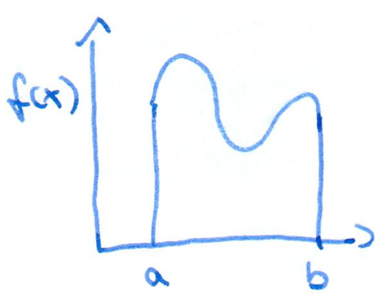
⇒ The area under the curve is the base $(b-a)$ times the "average height" $f(c)$. Instead of $f(c)$ we also write $\langle f \rangle$ or $E[f(x)]$.

Simple Monte Carlo integration:

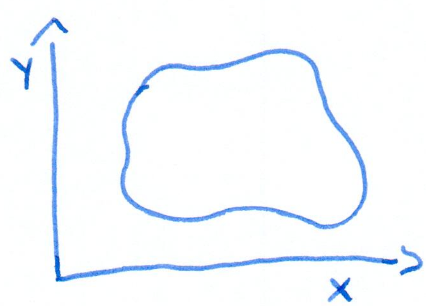
The idea is to estimate $\langle f \rangle$ to compute $\int_a^b f(x)\,dx$ using the mean-value-theorem for integrals.

1. Choose random samples $x_i \sim U[a,b]$
2. Compute $f(x)$ 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:



vs.



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

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

For the multi-dimensional case we need to make sure that the point (x_i, y_i) is inside the area we want to integrate over.

What if we don't know V? You can either estimate it or define a rectangular region instead and define $f(x_i, y_i)$ to be zero outside of the valid area.

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 basic Monte Carlo integration one can show that:

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

$\Rightarrow$ The error depends on the variance of the function f!

$\Rightarrow$ The error is independent of the dimensionality of f!

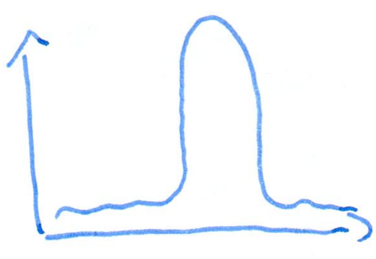
This is why MC methods are very popular for high-dimensional problems.

We will look closer into how to reduce the variance of f to come up with better (more precise) estimates.

Importance sampling (Intuition only):

Idea: Choose random points such that more points are sampled where $f(x)$ is large.

$\Rightarrow$ Draw points where the action is



We need to learn how to sample from non-uniform distributions.

Inverse transform:

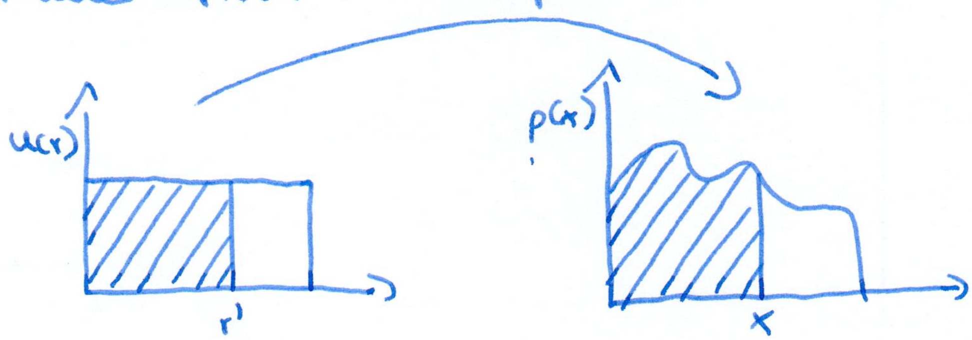
Idea: Transform uniform samples r directly into samples of a different distribution.

$$\begin{array}{ccc} r & \rightarrow & x \\ u(r) & & p(x) \end{array}$$

To find the right mapping we need to look at the CDF:

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

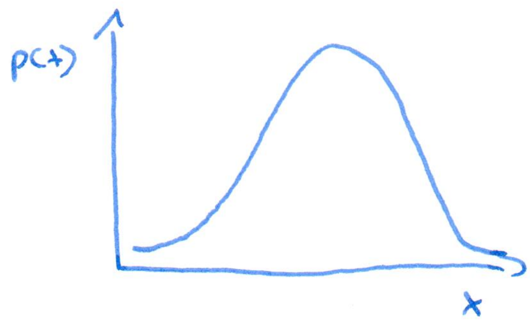
$\Rightarrow$ We want to preserve the probability of being smaller than r or respective x.

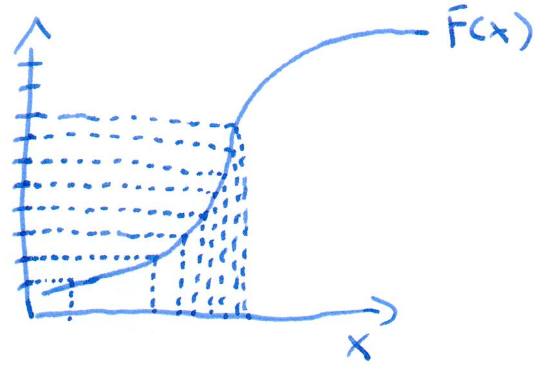


$$r = F(x) \quad \Leftrightarrow \quad x = F^{-1}(r)$$

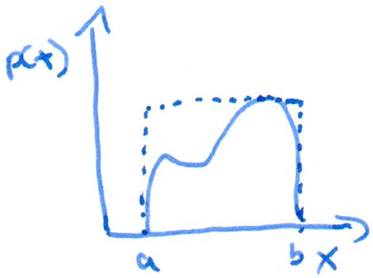
$\Rightarrow$ To use the inverse transform we need to know the antiderivative $F(x)$ and its inverse $F^{-1}(r)$

Different view:





Basic rejection sampling:

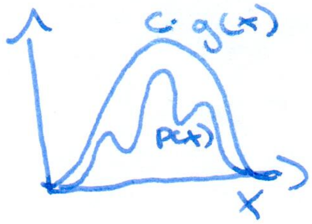


- draw $x \sim U[a, b]$
- draw $y \sim U[0, \max(p(x))]$

if $y < p(x)$: accept the sample otherwise reject.

Rejection sampling with an envelope function:

$\Rightarrow$ needs $g(x)$ such that we can easily sample from $g(x)$ and $c \cdot g(x) > p(x)$ for entire domain of interest.



- draw $x \sim g(x)$
- draw $y \sim U[0, 1]$
- if $y < \frac{p(x)}{c \cdot g(x)}$: accept otherwise reject.

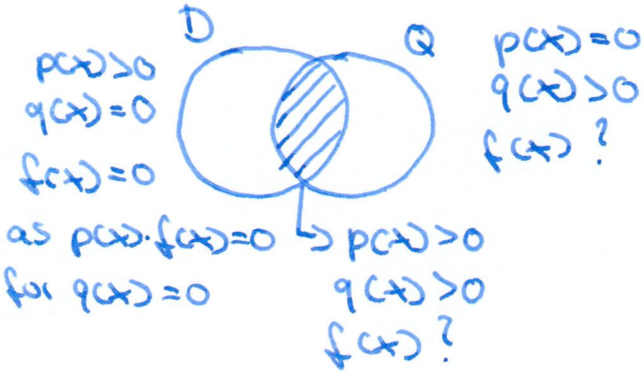
Importance sampling (now the math): (see Art Owen)

$$E_p[f(x)] = \int_D f(x)\cdot p(x)\,dx = \int_D \frac{f(x)\cdot p(x)}{q(x)}\cdot q(x)\,dx$$

$$\underset{\text{claim}}{\overset{\uparrow}{=}} E_q\left[\frac{f(x)\cdot p(x)}{q(x)}\right] = \int_Q \frac{f(x)\,p(x)}{q(x)}\cdot q(x)\,dx$$

$q(x)$ does not have to be positive everywhere, but we need $q(x)>0$ whenever $f(x)\cdot p(x)\neq 0$

Let's define $D=\{x \mid p(x)>0\}$ and $Q=\{x \mid q(x)>0\}$



$$E_q\left[\frac{f(x)\cdot p(x)}{q(x)}\right] = \int_Q \frac{f(x)\cdot p(x)}{q(x)}\cdot q(x)\,dx$$

$$= \int_Q f(x)\cdot p(x)\,dx$$

$$= \int_D f(x)\cdot p(x)\,dx + \int_{Q\cap D^c} f(x)\cdot \overset{\Rightarrow p(x)=0}{p(x)}\,dx$$

$$- \int_{D\cap Q^c} \overset{\Rightarrow f(x)=0}{f(x)}\cdot p(x)\,dx$$

$$\Rightarrow \int_Q f(x)\cdot p(x)\,dx = \int_D f(x)\cdot p(x)\,dx$$

and thus $E_p[f(x)] = E_q\left[\frac{f(x)\cdot p(x)}{q(x)}\right]$

For us $p(x)$ is typically uniform (a,b) and $q(x)$ is some distribution which looks similar to $f(x)$ but is easy to sample from.

$$\int_a^b f(x)\,dx = (b-a)\cdot E_p[f(x)]$$

$$= (b-a)\cdot E_q\left[\frac{f(x)\cdot p(x)}{q(x)}\right] \qquad p(x) = \frac{1}{b-a}$$

$$= \frac{(b-a)}{(b-a)}\cdot E_q\left[\frac{f(x)}{q(x)}\right] = E_q\left[\frac{f(x)}{q(x)}\right]$$

$$E_q\left[\frac{f(x)}{q(x)}\right] \approx \frac{1}{N}\sum_{x\sim q(x)} \frac{f(x)}{q(x)}$$

$\Rightarrow$
- sample $x_i \sim q(x)$
- compute $\frac{1}{N}\sum_{x\sim q(x)} \frac{f(x)}{q(x)}$
- this is the estimate for $\int_a^b f(x)\,dx$.

Intuition: as $q(x)$ and $f(x)$ are similar $\frac{f(x)}{q(x)}$ has lower variance than $f(x)$

$\Rightarrow$ we reduce the error in our estimate.

Art B. Owen, "Monte Carlo theory, methods and examples", 2013