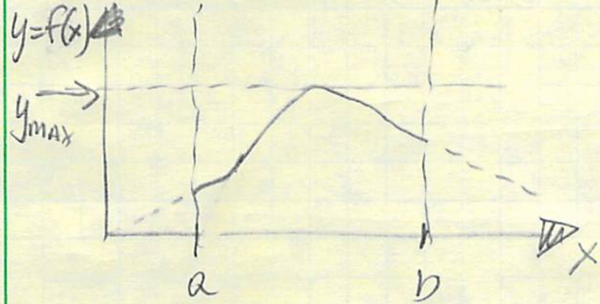
Acceptence/Rejetion and MC integration



area of rectangle $A = y_{MAX}(b-a)$

Pick N values x_i randomly btw b and a

Probability of accepting $x_i = \frac{1}{y_{max}} f(x_i) = p(x_i)$

n = # of successes out of n

Estimate of $I = \int_a^b f(x)dx = A \cdot \frac{n}{N}$

$\langle I \rangle = \frac{A}{N} \langle n \rangle$

$\langle n \rangle$ will vary with random seed

$$n = \sum p(x_i) = \frac{1}{y_{max}} \sum f(x_i)$$

$$I \sim \frac{A}{y_{max} N} \sum_{i=1}^{N} f_i$$ where $f_i = f(x_i)$

$$\boxed{I \sim \frac{b-a}{N} \sum_{i=1}^{N} f_i}$$

Different random seeds will give different $\sum f_i$

How accurate is this.

We can estimate the spread in $\sum_{i=1}^{N} f_i$ from different seeds using a single draw.

Voriance of f $= \frac{1}{N-1} \sum_{i=1}^{N} \left[f_i - \langle f_i \rangle \right]^2 \equiv \sigma_f^2$

N terms here

Then voriance of I in

$$\sigma_I^2 = \frac{(b-a)^2}{N^2} \sum_{n=1}^{N} \sigma_f^2$$ ← like uncertainties in quadrature

$$\boxed{\sigma_i = \frac{b-a}{\sqrt{N}} \sigma_f}$$

Note $b-a = \int_a^b dx$ = length of segment

This generalize to m-dimensional volume for m-voriables. Indeed these MC methods are useful for multivariate functions

eg $f(x_1 x_2 \ldots x_m) \Rightarrow b-a \Rightarrow V = \iiint dx_1 dx_2 dx_3 \ldots dx_m$

So far this does not look much better than splitting into little rectangles and adding up the areas _

Let's rewrite

$$I = \int_a^b f(x)\,dx = \int_a^b g(x)\,p(x)\,dx$$

Where $p(x)$ is the PDF over which we pick our x_i

$$p(x) = \frac{1}{b-a} \qquad g(x) = \frac{f(x)}{p(x)} = (b-a)\,f(x)$$

we had

$$I \sim \frac{b-a}{N}\sum_{i=1}^{N} f_i = \frac{1}{N}\sum_i g_i$$

$$\sigma_i = \frac{\sigma_g}{\sqrt{N}} \quad \text{as before}$$

How does this help us ?

Remember the accuracy depends on

$$\sigma_g = \sum (g_i - \langle g_i \rangle)^2$$

i.e. the spread in g_i

Let me rewrite $\quad I = \int_a^b f(x)\,dx = \int_a^b \left[\frac{f(x)}{p(x)}\right] p(x)\,dx$

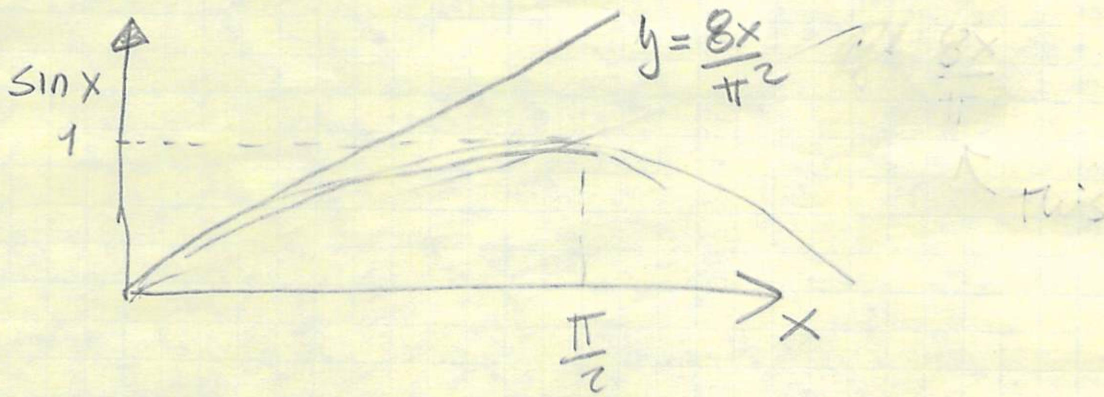
$$g(x) = \frac{f(x)}{p(x)} \qquad g_i = \frac{f_i}{p_i}$$

If I can choose $p(x)$ to be something close to $f(x)$ then the spread in the g_i will be minimized - (or up to proportionality constant)

In other words I will be sampling $F(x)$ where it is most important

Importance Sampling

Example $\int_0^{\pi/2} \sin(x)\,dx = 1$



Code in

/home/pi/physrpi/campagnari/testOfImportanceSampling.py