Acceptance/Rejection and MC Integration

\[ y = f(x) \]

\[ y_{max} \]

\[ a \quad b \]

\[ \text{Area of rectangle } A = y_{max} (b-a) \]

Pick \( N \) values \( x_i \) randomly between \( b \) and \( a \).

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

\( n = \# \text{ of successes out of } n \)

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 \( f_i = f(x_i) \)

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

Different random seeds will give different \( \sum f_i \)

How accurate is this?

We can estimate the spread in \( \frac{1}{N} \sum f_i \) from different seeds using a single draw.
Variance of $F = \frac{1}{N-1} \sum_{i=1}^{N} \left[ F_i - \langle F_i \rangle \right]^2 = \sigma_F^2$

Then variance of $I$ is:

$$\sigma_I^2 = \frac{(b-a)^2}{N^2} \sum_{n=1}^{N} \sigma_F^2$$

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

Note: $b-a = \int_a^b dx = \text{length of segment}$

This generalizes to $m$-dimensional volume for $m$-variables. Indeed, these MC methods are useful for multivariate functions.

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

\[ p(x) = \frac{1}{b-a} \quad g(x) = \left( \frac{f(x)}{p(x)} \right) = (b-a) \, f(x) \]

We had

\[ I \sim b-a \frac{1}{N} \sum_{i=1}^{N} f_i = \frac{1}{N} \sum g_i \]

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

How does this help us?

Remember the accuracy depends on

\[ \sigma_g = \sum \left( g_i - \langle g_i \rangle \right)^2 \]

i.e. the spread in \( g_i \)

Let me rewrite

\[ I = \int_a^b f(x) \, dx = \int_a^b \left[ \frac{f(x)}{p(x)} \right] p(x) \, dx \]
\[ g(x) = \frac{f(x)}{p(x)} \quad g_i = \frac{f_i}{p_i} \]

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 \[ \int_{0}^{\pi/2} \sin(x) \, dx = 1 \]

\[ y = \frac{8x}{\pi^2} \]

\[ \sin x \]

\[ 1 \]

Code

```
/home/pi/physrpi/campagnoni/testOfImportanceSampling.py
```