importance sampling

Suppose we want to integrate a normal distribution $N(0,1)$,

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad F = \int_{x_1}^{x_2} g(x) dx$$

using monte-carlo techniques. We are going to sample 1000 times.

Integration can be made using a uniform pdf or in better way, by using importance sampling.

Reminder:

- MC integral using uniform distribution for x:
  $I = \int_{x_1}^{x_2} g(x)\, dx = (x_2 - x_1) \, \langle g \rangle_N$
- Error associated to I calculation:
  $\sigma_I = (x_2 - x_1) \, \sigma_{<g>_N}$
  - Variance of $g(x)$ distribution (calculated for every x),
    $Var(g) = \mathcal{E}\left[(g- < g >)^2\right] = \langle g^2 \rangle - \langle g \rangle^2$
  - $\sigma^2_{<g>} = \frac{Var(g)}{N}$
- Error associated to I calculation:
  $\sigma_I = (x_2 - x_1) \sqrt{Var(g)/N}$

The integration is going to be performed on interval $[-50, 50]$.

1. using a uniform $pdf(x)$

> generate random variable x according to $pdf(x)$ from $[-50, 50]$ let's compute the integral 5000 times and make its distribution (histogram),
>
> $$I = (x_2 - x_1) \frac{1}{N} \sum_{i=1}^{N} g(x) = (x_2 - x_1) \, \langle g(x) \rangle$$

2. using importance sampling

> find an auxiliary function $p(x)$ to "smooth" the integrand,

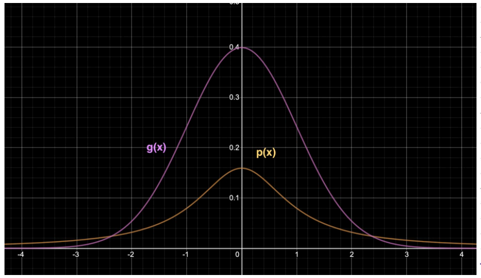
$$F = \int_{x_1}^{x_2} g(x) \, \frac{p(x)}{p(x)} \, dx = \int_{x_1}^{x_2} \frac{g(x)}{p(x)} \, p(x) \, dx$$

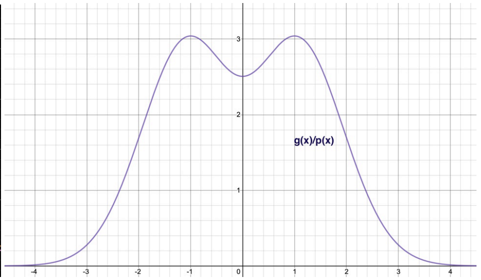
The monte-carlo integral is obtained

- making a variable transformation, using a uniform $p(y)$ distribution normalized to one on $[0,1]$ interval: $p(y)dy = p(x)dx \Rightarrow y = \int p(x)dx$
- $I = \left\langle \frac{g(x(y))}{p(x(y))} \right\rangle_y$

> For instance using a Cauchy function[a],
>
> [a] https://www.itl.nist.gov/div898/handbook/eda/section3/eda3663.htm

$$p(x) = \frac{1}{1 + x^2}$$





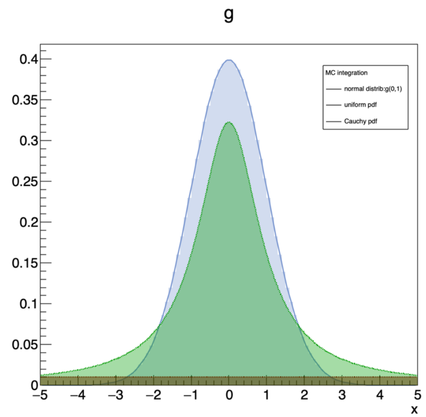
> Normalize Cauchy pdf on interval $[x_1, x_2]$,

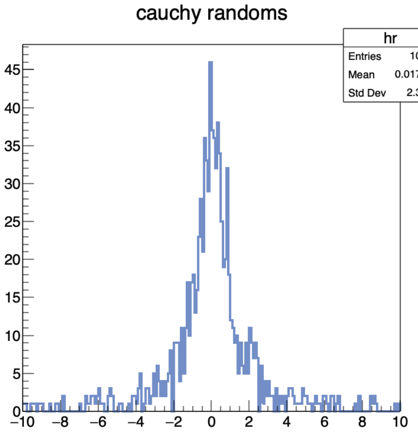
$$k \int_{x_1}^{x_2} p(x)\, dx = 1 \quad \Rightarrow \quad k = \frac{1}{atan(x_2) - atan(x_1)}$$

> find cumulative and invert it,

$$y = \int_{x_1}^{x} \frac{k}{1 + x^2} \, dx = k \left[atan(x) - atan(x_1) \right] \quad \Rightarrow \quad x = tan\left[\frac{y}{k} + atan(x_1) \right]$$

> Now, generating $y \sim [0,1]$ we get x correctly distributed according to Cauchy distribution (see next figure on the right)





3. Results

> The integral results are computed for every method (uniform pdf and Cauchy pdf) 500 times. The distribution of the integral values are shown on next figure (left=uniform pdf, right=Cauchy pdf). It is evident the huge difference on the precision of every method: while uniform pdf provides an error of ~ 0.17 the Cauchy pdf has a precision ten times higher