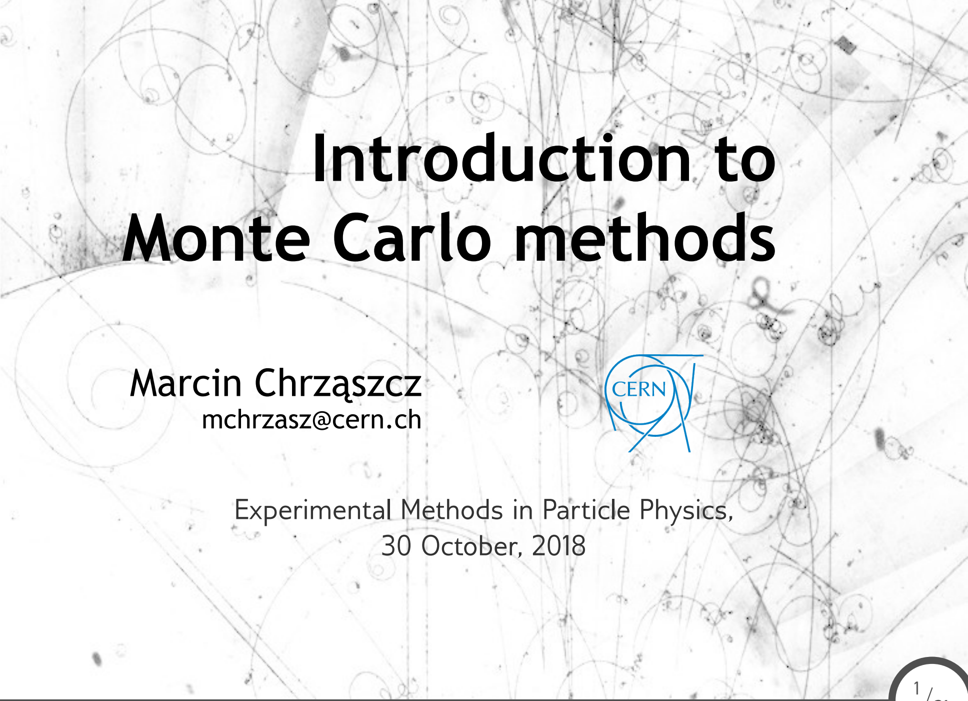


Introduction to Monte Carlo methods

Marcin Chrząszcz
mchrzasz@cern.ch

Experimental Methods in Particle Physics,
30 October, 2018

Literature

1. J. M. Hammersley, D. C. Hamdscomb, “Monte Carlo Methods”, London: Methuen & Co. Ltd., New York: J. Wiley & Sons Inc., 1964
2. I. M. Sobol, “The Monte Carlo Method”, Mir Publishers, Moscow, 1975.
3. M. H. Kalos, P. A. Whitlock, „Monte Carlo Methods”, J. Wiley & Sons Inc., New York, 1986
4. G. S. Fishman, „Monte Carlo: Concepts, Algorithms and Applications”, Springer, 1996.
5. R. Y. Rubinstein, D. P. Kroese, „Simulation and the Monte Carlo Method”, Second Edition, J. Wiley & Sons Inc., 2008.
6. R. Korn, E. Korn, G. Kroisandt, „Monte Carlo methods and models in finance and insurance”, CRC Press, Taylor & Francis Group, 2010.
7. S. Jadach, „Practical Guide to Monte Carlo”, arXiv:physics/9906056, http://cern.ch/jadach/MCguide/.

Course Plan

We will have 6 hours of Monte Carlo (MC) lectures. The lectures will be devoted:

- 1 h: Mathematical introduction to MC methods.
- 1 h: MC integration methods.
- 2 h: Random numbers generators.
- 2 h: Markov Chain MC.
- 2 h: Tutorial and examples.

The hands-on tutorial will consist of program templates in which we will implement couple of algorithms that were explained in the lecture.
⇒ All examples shown in this course are available in the github repository:
`https://github.com/mchrzasz/EMPP_MC`
There will be an indication (in this color) on the adequate slide for each of the macro.

Definitions

⇛ Basic definition:

Monte Carlo method is any technique that uses *random numbers* to solve a given mathematical problem.

↣ Random number: For the purpose of this course we need to assume that we know what it is, although the formal definition is highly non-trivial.

⇛ My favourite definition (Halton 1970): more complicated, but more accurate.

”Representing the solution of a problem as a parameter of a hypothetical population, and using a random sequence of numbers to construct a sample of the population, from which statistical estimates of the parameter can be obtained.”

To put this definition in mathematical language:
Let F be a solution of a given mathematical problem. The estimate of the result $\hat{F}$:

$$\hat{F} = f(\{r_1, r_2, r_3, ..., r_n\}; ...),$$

where $\{r_1, r_2, r_3, ..., r_n\}$ are random numbers.

The problem we are solving doesn’t need to be stochastic!

↠ One could wonder why are we trying to add all the stochastic properties to a deterministic problem. Those are the properties that allow to use all well known statistic theorems.

History of MC methods

- G. Compte de Buffon (1777) - First documented usage of random numbers for integral computation (Buffon thrown niddle on the table with parrarel line; we will do a modern version of this exercise).
- Marquis de Laplace (1886) - Used the Buffon niddle to determine the value of π number.
- Lord Kelvin (1901) - Thanks to drawing randomly numbered cards he managed he managed to calculate some integrals in kinematic gas theorem.
- W. S. Gosse (better knows as Student) (1908) - Used similar way as Lord Kelvin to get random numbers to prove t-Student distribution.
- Enrico Fermi (1930) - First mechanical device (`FERMIAC`) for random number generations. Solved neutron transport equations in the nuclear plants.
- S. Ulam, R. Feynman, J. von Neumann et. al. - First massive usage of random numbers. Most applications were in Manhattan project to calculate neutron scattering and absorption.

  In Los Alamos the name Monte Carlo was created as kryptonim of this kind of calculations.

Euler number determination, Lecture1/Euler_number

⇛ As mentioned before MC methods can be used to solve problems that **do not** have stochastic nature! All the integrals calculated in Los Alamos during the Manhattan project are nowadays solvable without any MC methods.
↣ Let's give a trivial example of solving a non stochastic problem: calculating Euler number e. We know that $e = 2.7182818....$ ⇛ To calculate the $\hat{e}$ we will use the following algorithm:

- We generate a random number in range $(0, 1)$ (in stat. $\mathcal{U}(0, 1)$) until the number we generate is smaller then the previous one, aka we get the following sequence:

$$x_1 < x_2 < ... < x_{n-1} > x_n$$

- We store the number n. We repeat this experiment N times and calculate the arithmetic average of n. The obtained value is an statistical estimator of e:

$$\hat{e} = \frac{1}{N}\sum_{i=1}^{N} n_i \xrightarrow{N\rightarrow\infty} e.$$

⇛ Numerical example:

N	$\hat{e}$	$\hat{e} - e$
100	2.760000	0.041718
10000	2.725000	0.006718
1000000	2.718891	0.000609
100000000	2.718328	0.000046

Is this $\sim \sqrt{N}$?

Let's test the $\sqrt{N}$, $\mathrm{Lecture1/Euler\_number}$

$\Rrightarrow$ In the last example we measured the Euler number using different number of pseudo-experiments.

$\rightarrowtail$ We compared the obtained value to the true and observed roughly a $\sqrt{N}$ dependence on the difference between the true value and the obtained one.

$\rightarrowtail$ Could we test this? YES! Lets put our experimentalist hat on!

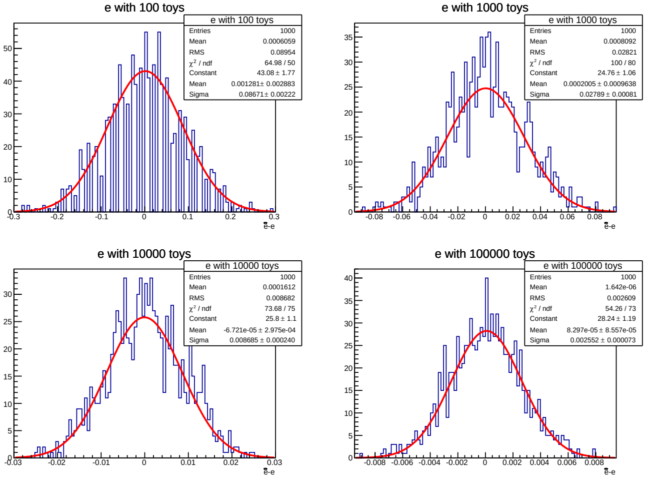
$\rightarrowtail$ From the begging of studies they tooth us to get the error you need to repeat the measurements.

The algorithm:

Previous time we measured Euler number using N events, where $N \in (100, 1000, 10000, 100000)$. Now lets repeat this measurement n_N times (of course each time we use new generated numbers). From the distribution of $\hat{e} - e$ we could say something about the uncertainty of our estimator for given N.

Let's test the $\sqrt{N}$, Lecture1/Euler_number

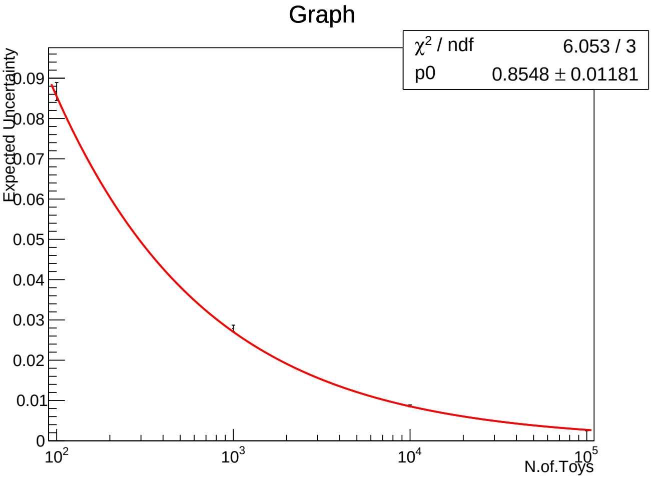
- ⟶ Could we test this? YES! Lets put our experimentalist hat on!
- ⟶ From the begging of studies they tooth us to get the error you need to repeat the measurements.



Let's test the $\sqrt{N}$, Lecture1/Euler_number

⤚⟶ Could we test this? YES! Lets put our experimentalist hat on!

⤚⟶ From the begging of studies they tooth us to get the error you need to repeat the measurements.



Monte Carlo and integration

$\hookrightarrow$ **All MC calculations are equivalent to preforming an integration.**

$\rightrightarrows$ Assumptions: r_i random numbers from $\mathcal{U}(0,1)$. The MC result:

$$F = F(r_1, r_2, ...r_n)$$

is unbias estimator of an integral:

$$I = \int_0^1 ... \int_0^1 F(x_1, x_2, ..., x_n) dx_1, dx_2 ..., dx_n$$

aka the expected value of the I integral is:

$$E(F) = I.$$

$\Rrightarrow$ This mathematical identity is the most useful property of the MC methods. It is a link between mathematical analysis and statistic world. Now we can use the best of the both world!

If we want to calculate the integral in different range then $(0,1)$ we just scale the the previous result:

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \xrightarrow{N\to\infty} E(f) = \frac{1}{b-a}\int_a^b f(x)dx$$

Uncertainty from Monte Carlo methods

$\Rrightarrow$ In practice we do not have $N \rightarrow \infty$ so we will never know the exact result of an integral :(

$\longmapsto$ Let's use the statistical world and estimate the uncertainty of an integral in this case :)

$\rightarrowtail$ A variance of a MC integral:

$$V(\hat{I}) = \frac{1}{n}\left\{E(f^2) - E^2(f)\right\} = \frac{1}{n}\left\{\frac{1}{b-a}\int_a^b f^2(x)dx - I^2\right\}$$

$\looparrowright$ To calculate $V(\hat{I})$ one needs to know the value of I!

$\Rrightarrow$ In practice $V(\hat{I})$ is calculated via estimator:

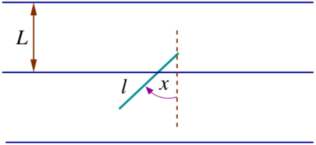
$$\hat{V}(\hat{I}) = \frac{1}{n}\hat{V}(f), \qquad \hat{V}(f) = \frac{1}{n-1}\sum_{i=1}^{n}\left[f(x_i) - \frac{1}{n}\sum_{i=1}^{n} f(x_i)\right]^2.$$

$\Rrightarrow$ MC estimator of standard deviation: $\hat{\sigma} = \sqrt{\hat{V}(\hat{I})}$

Buffon needle - π number calculus

⇛ Buffon needle (Buffon 1777, Laplace 1886): We are throwing a needle (of length l) on to a surface covered with parallel lines width distance L. If a thrown needle touches a line we count a hit, else miss. Knowing the number of hits and misses one can calculate the π number.

Experiment:



n - number of hits
N number of hits and misses, aka number of tries.

Theory:

⇒ x - angle between needle and horizontal line, $x \in \mathcal{U}(0,\pi)$. ⇒ the probability density function (p.d.f.) for x:

$$\rho(x) = \frac{1}{\pi}$$

⇒ $p(x)$ probability to hit a line for a given x value:

$$p(x) = \frac{l}{L}|\cos x|$$

⇒ Total hit probability:

$$P = E[p(x)] = \int_0^\pi p(x)\rho(x)dx = \frac{2l}{\pi L}$$

Now one can calculate $\hat{P}$ from MC : $\hat{P} = \frac{n}{N} \xrightarrow{N\to\infty} P = \frac{2l}{\pi L} \Rightarrow \hat{\pi} = \frac{2Nl}{nL}$

Buffon needle - Simplest Carlo method

Monte Carlo type ”heads or tails”

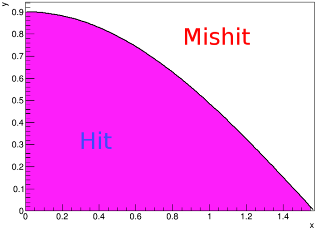
Let’s use the summery of $p(x)$ function nad take $0 < x < \frac{\pi}{2}$.

$\Rightarrow$ Algorithm:

Generate 2 dim. distribution:

$$(x, y) : \mathcal{U}(0, \frac{\pi}{2}) \times \mathcal{U}(0, 1) \text{ and}$$

$$y \begin{cases} \leqslant p(x): & \text{hit}, \\ > p(x): & \text{miss}. \end{cases}$$



Let’s define weight function: $w(x, y) = \Theta(p(x) - y)$, where $\Theta(x)$ is the step function.

$\rightarrowtail$ p.d.f.: $\varrho(x, y) = \rho(x)g(y) = \frac{2}{\pi} \cdot 1$

$\Rightarrow$ Integrated probability:

$$P = E(w) = \int w(x, y)\varrho(x, y)dxdy = \frac{2l}{\pi L} \xleftarrow{N \to \infty} \hat{P} = \frac{1}{N}\sum_{i=1}^{N} w(x_i, y_i) = \frac{n}{N}$$

Standard deviation for $\hat{P}$: $\hat{\sigma} = \frac{1}{\sqrt{N-1}}\sqrt{\frac{n}{N}\left(1 - \frac{n}{N}\right)}$

Buffon needle, Lecture1/Heads_tails

$\Rrightarrow$ Lets make this toy experiment and calculate the π number.
$\hookrightarrow$ We can simulate the central position (y) of an needle between $(-L, L)$ from $\mathcal{U}(-L, L)$.

Symmetry:

Please note the symmetry of the problem, if the position of the needle would be $> L$ then we can shift the needle by any number of L's.

$\hookrightarrow$ New we simulate the angle (ϕ) with a flat distribution from $(0, \pi)$. $\hookrightarrow$ The maximum and minimum y position of the needle are:

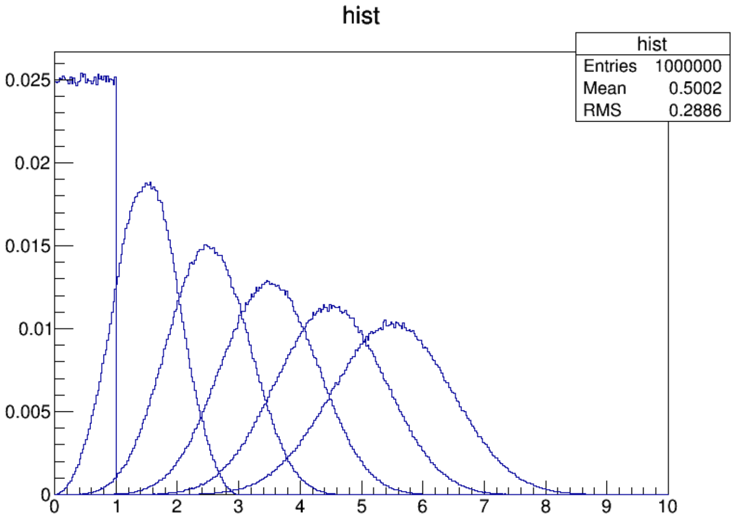
$$y_{\max} = y + |\cos\phi| l$$
$$y_{\min} = y - |\cos\phi| l$$

$\hookrightarrow$ Now we check if the needle touches any of the lines: $y = L$, $y = 0$ or $y = -L$. If yes we count the events.

N	$\hat{\pi}$	$\hat{\pi} - \pi$	$\sigma(\hat{\pi})$
10000	3.12317	−0.01842	0.03047
100000	3.14707	0.00547	0.00979
1000000	3.13682	−0.00477	0.00307
10000000	3.14096	−0.00063	0.00097

Central Limit Theorem, Lecture1/CLT

Large independent random numbers assembly has always Gaussian distribution no matter from what distribution they were generated from as far as they have finite variances and expected values and the assembly is sufficiently large.



Crude Monte Carlo method of integration

⇛ Crude Monte Carlo method of integration is based on Central Limit Theorem (CLT):

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \xrightarrow{N\to\infty} \frac{1}{b-a}\int_a^b f(x)dx = E(f)$$

⇛ The standard deviation can be calculated:

$$\sigma = \frac{1}{\sqrt{N}}\sqrt{\left[E(f^2) - E^2(f)\right]}$$

⇛ From LNT we have:

$$P = \int w(x)\rho(x)dx = \int_0^{\pi/2} (\frac{l}{L}\cos x)\frac{2}{\pi}dx = \frac{2l}{\pi L} \xrightarrow{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} w(x_i)$$

⇛ Important comparison between "Hit and mishit" and Crude MC methods. One can analytically calculate:

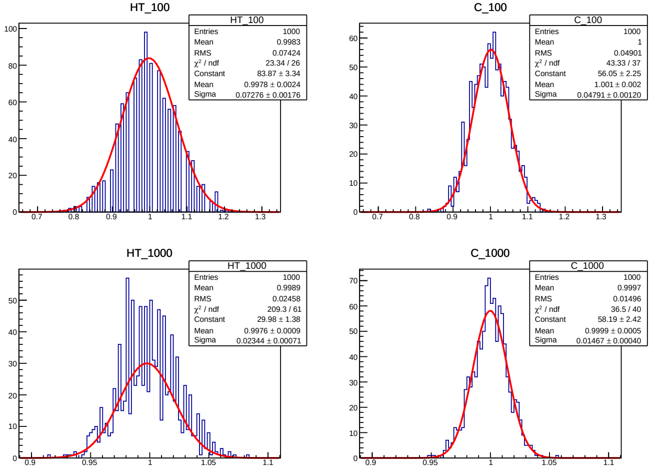
$$\hat{\sigma}^{\text{Crude}} < \hat{\sigma}^{\text{Hit and mishit}}$$

⇛ Crude MC is **always** better then "Hit and mishit" method. We will prove this on an example (can be proven analytically as well).

Crude MC vs "Hit and mishit", Lecture1/Crude_vs_HT

⇒ We can repeat a toy MC studies as we did in the Euler needle case.

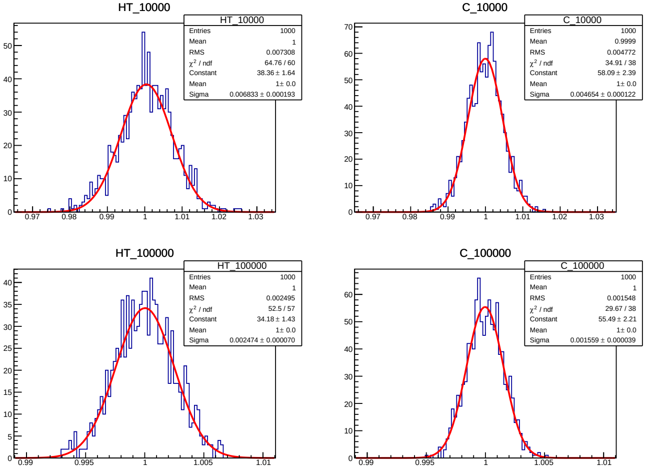
↪ In this example we want to calculate $\int_0^{\pi/2} \cos x dx$



Crude MC vs ”Hit and mishit”, Lecture1/Crude_vs_HT

⇛ We can repeat a toy MC studies as we did in the Euler needle case.

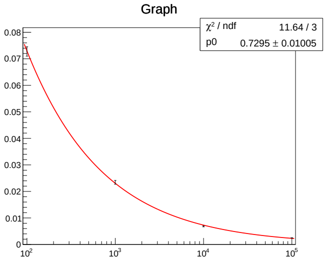
↪ In this example we want to calculate $\int_0^{\pi/2} \cos x dx$

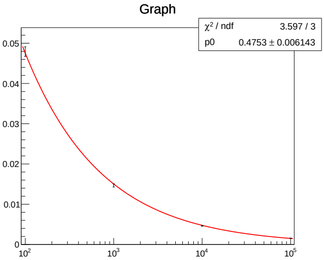


Crude MC vs ”Hit and mishit”, Lecture1/Crude_vs_HT

$\Rrightarrow$ We can repeat a toy MC studies as we did in the Euler needle case.

$\hookrightarrow$ In this example we want to calculate $\int_0^{\pi/2} \cos x dx$





$\Rrightarrow$ One clearly sees that both methods follow $1/\sqrt{N}$ dependence and that the Crude MC is always better then the ”Hit and mishit”.

$\Rrightarrow$ Please note that for the ”Hit and mishit” we are suing 2 times more random numbers than for the Crude method so in terms of timing the Crude MC is also much faster.

Classical methods of variance reduction

⇛ In Monte Carlo methods the statistical uncertainty is defined as:

$$\sigma = \frac{1}{\sqrt{N}}\sqrt{V(f)}$$

⇛ Obvious conclusion:

- To reduce the uncertainty one needs to increase N.
  ⇉ Slow convergence. In order to reduce the error by factor of 10 one needs to simulate factor of 100 more points!

⇛ How ever the other handle ($V(f)$) can be changed! ⟶ Lot's of theoretical effort goes into reducing this factor.

⇛ We will discuss four classical methods of variance reduction:

1. Stratified sampling.
2. Importance sampling.
3. Control variates.
4. Antithetic variates.

Stratified sampling

⇛ The most intuitive method of variance reduction. The idea behind it is to divide the function in different ranges and to use the Riemann integral property:

$$I = \int_0^1 f(u)du = \int_0^a f(u)du + \int_a^1 f(u)du, \ 0 < a < 1.$$

⇛ The reason for this method is that in smaller ranges the integration function is more flat. And it's trivial to see that the more flatter you get the smaller uncertainty.
⇉ A constant function would have zero uncertainty!

General schematic:

Let's take our integration domain and divide it in smaller domains. In the j^{th} domain with the volume w_j we simulate n_j points from uniform distribution. We sum the function values in each of the simulated points for each of the domain. Finally we sum them with weights proportional to w_i and anti-proportional to n_i.

Stratified sampling - mathematical details

Let's define our integrals and domains:

$$I = \int_{\Omega} f(x)dx, \quad \Omega = \bigcup_{i=1}^{k} w_i$$

The integral over j^{th} domain:

$$I_j = \int_{w_j} f(x)dx, \quad \Rightarrow I = \sum_{j=1}^{k} I_i$$

$\rightrightarrows$ p_j uniform distribution in the w_j domain: $dp_j = \frac{dx}{w_j}$.

$\rightrightarrows$ The integral is calculated based on crude MC method. The estimator is equal:

$$\hat{I}_j = \frac{w_j}{n_j} \sum_{i=1}^{n_j} f(x_j^i)$$

Now the total integral is just a sum:

$$\hat{I} = \sum_{j=1}^{k} \hat{I}_j = \sum_{j=1}^{k} \frac{w_j}{n_j} \sum_{i=1}^{n_j} f(x_j^{(i)}),$$

Variance: $V(\hat{I}) = \sum_{j=1}^{k} \frac{w_j^2}{n_j} V_j(f)$, and it's estimator: $\hat{V}(\hat{I}) = \sum_{j=1}^{k} \frac{w_j^2}{n_j} \hat{V}_j(f)$

Importance sampling

⇛ If the function is changing rapidly in its domain one needs to use a more elegant method: make the function more stable.

⇉ The solution is from first course of mathematical analysis: change the integration variable :)

$$f(x)dx \longrightarrow \frac{f(x)}{g(x)}dG(x), \text{ where } g(x) = \frac{dG(x)}{dx}$$

Schematic:

- Generate the distribution from $G(x)$ instead of $\mathcal{U}$.
- For each generate point calculate the weight: $w(x) = \frac{f(x)}{g(x)}$.
- We calculate the expected value $\hat{E}(w)$ and its variance $\hat{V}_G(w)$ for the whole sample.

- If $g(x)$ is choose correctly the resulting variance can be much smaller.
- There are some mathematical requirements:
  - $g(x)$ needs to be non-negative and analytically integrable on its domain.
  - $G(x)$ invertible or there should be a direct generator of g distribution.

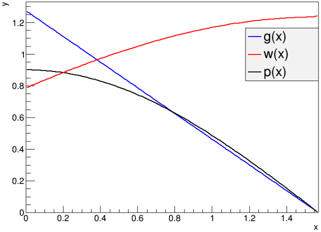
Importance sampling - Example

$\Rrightarrow$ Let's take our good old π determination example.

$\Rrightarrow$ Let's take here for simplicity: $L = l$.

- Let's take a trivial linear weight function: $g(x) = \frac{4}{\pi}(1 - \frac{2}{\pi}x)$
- It's invertible analytically: $G(x) = \frac{4}{\pi}x(1 - \frac{x}{\pi})$
- The weight function:

$$w(x) = \frac{p(x)}{g(x)} = \frac{\pi}{4}\frac{\cos x}{1 - 2x/\pi}$$



- Now the new standard deviation is smaller:

$$\sigma_{\pi}^{\text{IS}} \simeq \frac{0.41}{\sqrt{N}} < \sigma_{\pi} \simeq \frac{1.52}{\sqrt{N}}$$

- Importance sampling has advantages:
  - Big improvements of variance reduction.
  - The only method that can cope with singularities.

Wrap up

⇛ To sum up:

- We discussed basic mathematical properties of MC methods.
- We shown that besides the stochastic nature of MC they can be used to determine totally non stochastic quantities.
- We demonstrated there is a perfect isomorphism between MC method and integration.
- We learned how co calculate integrals and estimate the uncertainties.
- Finally we discussed several classical methods of variance reduction.

Backup

Control variates

⇛ Control variates uses an other nice property of Riemann integral:

$$\int f(x)dx = \int [f(x) - g(x)]dx + \int g(x)dx$$

- $g(x)$ needs to be analytically integrable.
- The uncertainty comes only from the integral: $\int [f(x) - g(x)]dx$.
- Obviously: $V(f \to g) \xrightarrow{f \to g} 0$

⇛ Advantages:

- Quite stable, immune to the singularities.
- $g(x)$ doesn't need to be invertible analytically.

⇛ Disadvantage:

- Useful only if you know $\int g(x)dx$

Antithetic variates

⇛ In MC methods usually one uses the independent random variables. The Antithetic variates method on purpose uses a set of correlated variables (negative correlation is the important property):

- Let f and $f\prime$ will be functions of x on the same domain.
- The variance: $V(f + f\prime) = V(f) + V(f\prime) + 2Cov(f, f\prime)$.
- If $Cov(f, f\prime) < 0$ then you can reduce the variance.

⇛ Advantages:

- If you can pick up f and $f\prime$ so that they have negative correlation one can significantly reduce the variance!

⇛ Disadvantages:

- There are no general methods to produce such a negative correlations.
- Hard to generalize this for multidimensional case.
- You can't generate events from $f(x)$ with this method.