Monte Carlo Methods in High Energy Physics

WIESŁAW PŁACZEK

Institute of Physics, Jagiellonian University,

Cracow, Poland.

- **Lecture 1: Basics of Monte Carlo methods.**
- **Lecture 2: Random number generators.**
- **Lecture 3: Monte Carlo simulations of HEP processes.**
- **Lecture 4: Markovian Monte Carlo.**

$\Rightarrow$ http://cern.ch/placzek

Outline of Lecture 1:

- **Indroduction.**
  - ▷ Definition.
  - ▷ History.
- **Mathematical foundations of Monte Carlo (MC) methods.**
  - ▷ Random variables and distributions.
  - ▷ Expectation, variance, covariance.
  - ▷ The Law of Large Numbers.
  - ▷ The Central Limit Theorem.
- **Buffon's needle and basic Monte Carlo methods.**
- **Classical variance-reducing techniques.**
- **Adaptive variance-reducing techniques.**
- **Comparisons of MC integration with numerical quadrature.**

Introduction

- **Definition.**

  $\Rightarrow$ General definition:

  > A **Monte Carlo technique** is any technique making use of random numbers to solve a problem.

  $\Rightarrow$ Halton (1970) – a narrower but more enlightening definition:

  > The **Monte Carlo method** is defined as 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 the statistical estimates of the parameter can be obtained.

  Let F – the solution of the problem (a real number, a set of numbers, a yes/no decision, etc.) $\Rightarrow$ MC estimate of F:

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

  where $\{r_1, r_2, \ldots, r_n\}$ – the random numbers used in the calculation.

  The problem does not have to be of the stochastic nature!

History

- G. Comte de Buffon (1777) – perhaps the earliest documented use of random sampling to find the solution to the integral (by throwing a needle onto horizontal plane ruled with straight lines).
- Marquis Pierre-Simon de Laplace (1886) – use of Buffon's method to evaluate π.
- Lord Kelvin (1901) – use random sampling (drawing numbered pieces of paper from a bowl) to aid in evaluating some integrals in the kinetic theory of gases.
- W.S. Gossett (as "Student") (1908) – similar random sampling helped in his discovery of the distribution of the correlation coefficient.
- Enrico Fermi (1930s) – numerical sampling experiments on neutron diffusion and transport in nuclear reactors (deviced FERMIAC – a mechanical sampling device).
- J. von Neumann, S. Ulam, N. Metropolis, R. Feynman (1940s) – first large-scale random-numbers based calculations of neutron scattering and absorption during the "Manhattan" project (work on a nuclear bomb). The name **"Monte Carlo"** was invented for this type of calculations and respective mathematical methods.

History – cont.

- H. Kahn (1950) – invention of "importance sampling".
- N. Metropolis, A.W. Rosenbluth, N.M. Rosenbluth, A.H. Teller, E. Teller (1953) – invention of "Metropolis algorithm" (or "M(RT)^2 algorithm") for solving some random-walk (Markov chain) problems in statistical mechanics.
- I. Kopylov, JETP **35** (1958) 1426 – use of Monte Carlo methods in particle physics to solve the Fermi phase-space problem.
- ...
- G. Peter Lepage, J. Comput. Phys. **27** (1978) 195 – "VEGAS algorithm" for adaptive integration.
- S. Jadach, Comput. Phys. Commun. **130** (2000) 244; e-Print: physics/9910004 – FOAM: a general-purpose cellular Monte Carlo sampler.
- ...

Mathematical foundations of Monte Carlo methods

- **Random variables and distributions**

A **random variable** is a variable that can take on more than one value (generally a continuous range of values), and for which any value that will be taken cannot be predicted in advance (however, its distribution may well be known).

The **distribution of a random variable** gives the probability of a given value (or infinitesimal range of values).

▷ For continuous variables we define

$$\rho(u)du = \mathcal{P}[u < u' < u + du],$$

$\rho(u)$ – **the probability density function (pdf)** of u (gives the probability of finding the random variable u' within du of a given value u).

▷ The **cumulative (integrated) distribution function (cdf)**:

$$R(u) = \int_{-\infty}^{u} \rho(x)dx, \qquad \rho(u) = \frac{dR(u)}{du}\,.$$

Note: $R(u)$ – monotonically non-decreasing function and $0 \leq R(u) \leq 1$.

Expectation, variance, covariance

- **Expectation value of a function $f(u')$:**

$$E(f) = \int f(u) dR(u) = \int f(u)\rho(u)du\,.$$

If $u' \in \mathcal{U}(0,1)$, i.e. uniformly distributed between 0 and 1, then $E(f) = \int_0^1 f(u)du$

- **Variance of a function $f(u')$:**

$$V(f) = E[f - E(f)]^2 = \int [f - E(f)]^2 dR = E(f^2) - E^2(f)\,.$$

$\Rightarrow$ **Standard deviation:** $\sigma(f) = \sqrt{V(f)}\,.$

- Let x and y be random variables and c be a constant, then:

$$E(cx + y) = cE(x) + E(y),$$

$$V(cx + y) = c^2 V(x) + V(y) + 2c\,\mathrm{Cov}(x, y),$$

where $\mathrm{Cov}(x, y) = E([x - E(x)][y - E(y)])$ **– covariance between x and y.**

$\triangleright$ If $\mathrm{Cov}(x, y) = 0$: x and y are **uncorrelated**, but not necessarily independent!

The Law of Large Numbers (LLN)

Let's choose n numbers u_i randomly with a probability density uniform on the interval (a, b), and for each u_i evaluate the function $f(u_i)$. Then, as n becomes large:

$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) \underset{n\to\infty}{\longrightarrow} E(f) = \frac{1}{b-a}\int_a^b f(u)du\,.$$

That is, in statistical language, the left-hand side is a **consistent estimator** of the integral on the right-hand side, since (under certain conditions) it converges to the exact value of the integral as n approaches infinity.

▷ **'Certain conditions'** concern the behaviour of the function f – it must be:

- integrable,
- piecewise continuous (it may have a finite number of discontinuities),
- everywhere finite.

The law of large numbers can be interpreted as a statement that the Monte Carlo estimator of an integral converges to the correct answer as the random sample size becomes very large.

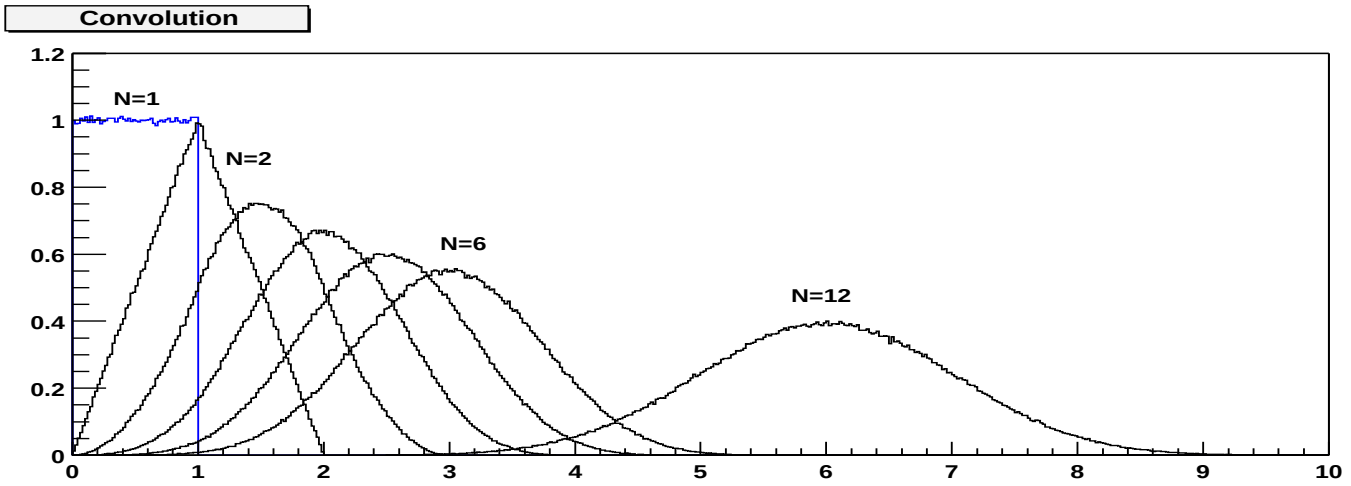
The Central Limit Theorem (CLT)

The sum of a large number of independent random variables is always normally distributed (i.e. a Gaussian distribution), no matter how the individual random variables are distributed, provided they have finite expectations and variances and provided n is 'large enough' (in practice the convergence is pretty fast)**.**

▷ Gaussian distribution $N(\mu, \sigma^2)$ with the expectation value μ and the variance σ^2:

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] .$$

▷ The illustration of CLT for $x_i \in \mathcal{U}(0,1), i = 1, \ldots, 12$:



▷ Gaussian random number generator based on the CLT:

Let $x_i \in \mathcal{U}(0,1), i = 1, \ldots, n,$

take $R_n = \sum_{i=1}^{n} x_i,$ then:

$$\left.\begin{array}{l} E(x_i) = \frac{1}{2} \\ V(x_i) = \frac{1}{12} \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} E(R_n) = \frac{n}{2} \\ V(R_n) = \frac{n}{12} \end{array}\right.$$

$\longrightarrow$ From the above we have:

$$\frac{R_n - n/2}{\sqrt{n/12}} \underset{n\to\infty}{\longrightarrow} N(0,1),$$

i.e. we get the standardized Gaussian random number generator.

A convenient choice for practical purposes is:

$$n = 12 \quad \longrightarrow \quad R_{12} - 6\,.$$

Warning: The tails of the Gaussian distribution are not well reproduced by this kind of a generator!

Résumé: mathematical properties of the Monte Carlo method

Let $u_i \in \mathcal{U}(a,b)$, then from the Law of Large Numbers:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} f(u_i)}_{\text{MC estimator of the integral}} \quad \underset{n\to\infty}{\longrightarrow} \quad \frac{1}{b-a}\int_a^b f(u)du$$

Mathematical properties of the Monte Carlo estimator:

(i) If $V(f) < \infty$, the MC estimator is **consistent**, i.e. it converges to the true value of the integral for very large n.

(ii) The MC estimator is **unbiased** for all n, i.e. the expectation value of the MC estimator is the true value of the integral (easy to check – from the linearity of the operator E).

(iii) The MC estimator is asymptotically **normally distributed** (Gaussian density).

(iv) The **standard deviation** of the MC estimator is given by: $\sigma = \frac{1}{\sqrt{n}}\sqrt{V(f)}\,.$

▷ The MC estimator of the standard deviation: $\hat{\sigma} = \frac{1}{\sqrt{n}}\sqrt{\hat{V}(f)}\,,$

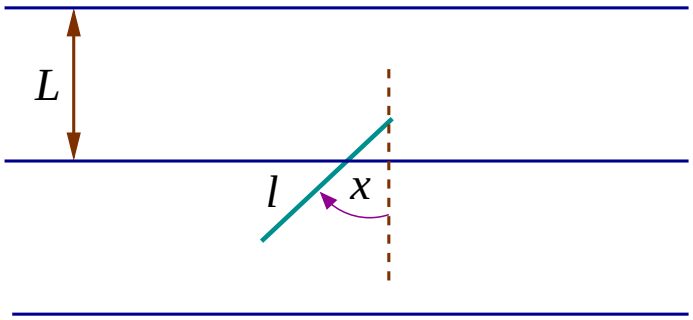
where: $\hat{V}(f) = \frac{1}{n-1}\sum_{i=1}^{n}\left[f(u_i) - \frac{1}{n}\sum_{i=1}^{n} f(u_i)\right]^2.$

Buffons's needle and basic Monte Carlo methods

▷ **Buffon's needle** (Buffon 1777, Laplace 1886)**:**

A needle of length l is thrown at random onto a horizontal plane ruled with straight lines a distance L, $(L \geq l)$ apart. A **'hit'** is counted when the needle crosses the line, and a **'miss'** otherwise. By counting the hits and misses calculate the value of π.

Experiment:



n – number of 'hits'

N – number of tries ('hits' and 'misses')

Theory:

x – the angle between the needle and the perpendicular to the lines, $x \in \mathcal{U}(0, \pi)$,

$\Rightarrow$ probability density function of x: $\rho(x) = \frac{1}{\pi}$.

$p(x)$ – probability of a 'hit' for a given angle x:

$$p(x) = (l/L)|\cos x|$$

$\Rightarrow$ Total probability of the 'hit':

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

▷ **Law of Large Numbers:** $\frac{n}{N} \underset{N\to\infty}{\longrightarrow} P = \frac{2l}{\pi L} \quad \Longrightarrow \quad \boxed{\frac{2Nl}{nL} \underset{N\to\infty}{\longrightarrow} \pi}$.

Buffons's needle – Monte Carlo modeling

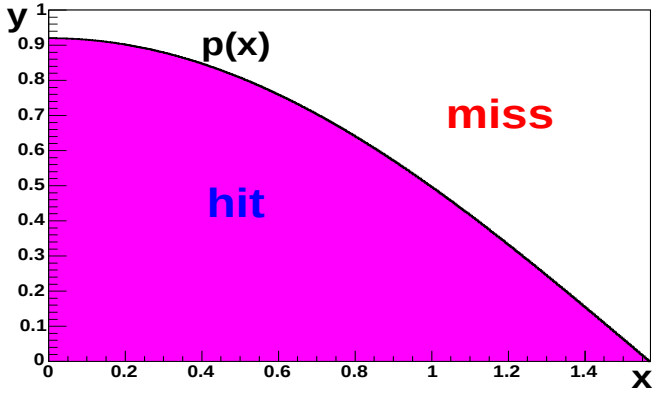
- **Hit-or-miss Monte Carlo**

  Due to symmetry of $p(x)$ we can consider: $0 < x < \frac{\pi}{2}$.

  ▷ **Algorithm:**

  Generate uniformly two random variables (x, y) : $0 < x < \frac{\pi}{2}$ and $0 < y < 1$.

$$y \begin{cases} \leq p(x) : \textbf{hit}\,, \\ > p(x) : \textbf{miss}\,. \end{cases}$$



Let's define the weight function: $w(x, y) = \Theta(p(x) - y)$,

where $\Theta(z) = 1$ for $z \geq 0$, and $= 0$ otherwise (a step function).

▷ Probability density function: $\varrho(x, y) = \rho(x)g(y) = \frac{2}{\pi} \cdot 1\,.$

$\Rightarrow$ Total probability:

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

$\Rightarrow$ Standard deviation of MC estimate of P:

$$\sigma = \frac{1}{\sqrt{N}}\sqrt{P(1-P)} \underset{N\to\infty}{=} \frac{1}{\sqrt{N}}\sqrt{\frac{n}{N}\left(1-\frac{n}{N}\right)}\,.$$

- **Hit-or-miss Monte Carlo – cont.**

  ▷ Standard deviation of MC estimate of π for $l = L$: $\boxed{\sigma_\pi^{\text{hit-or-miss}} \simeq \frac{2.374}{\sqrt{N}}}$

  This means that the uncertainty on the value of π is :

after	100	tries:	0.2374
after	10 000	tries:	0.0237
after	1 000 000	tries:	0.0024

  → **These uncertainties are very high!**

  Can we improve it?

- **Crude Monte Carlo**

  Let's define the weight function: $w(x) = p(x) = \frac{l}{L}\cos x$

  and generate x uniformly in the range $(0, \frac{\pi}{2})$.

  ▷ From the Law of Large Numbers we have:

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

  ▷ Standard deviation of MC estimate of π for $l = L$: $\boxed{\sigma_\pi^{\text{crude}} \simeq \frac{1.52}{\sqrt{N}}}$

→ **Crude Monte Carlo is more efficient than hit-or-miss Monte Carlo (always true)!**

▶ But hit-or-miss MC provides **unweighted** events while crude MC only **weighted** events!

Classical variance-reducing techniques

The uncertainty of a Monte Carlo integral is: $\sigma = \sqrt{V(f)}/\sqrt{n}\,.$

▷ It can be decreased by increasing n – very slow convergence!

▷ Another way is to try do decrease the effective variance $V(f)$.

Intuitively, large uncertainties of MC integration are due to the fact that the integrand differs considerably from the flat distribution from which points are generated → large fluctuations of the functions values in the MC estimate of the integral!

► A possible solution: To make the distribution of points to be closer to the function f.

- **Stratified sampling**

  ▷ Based on the fundamental property of the integral:

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

  ► General scheme:

  The full integration domain is divided into sub-domains. In the jth sub-domain whose volume is ω_j one chooses uniformly n_j random points. Then, partial sums are formed over each sub-domain, and these partial sums are added, weighted proportionally to ω_j and inversely to n_j.

Stratified sampling – cont.

Let $I = \int_\Omega f(x)dx$, $\Omega = \sum_{i=1}^{k} \omega_j$, and $I_j = \int_{\omega_j} f(x)dx \;\Rightarrow\; I = \sum_{j=1}^{k} I_j$.

p_j – uniform distribution over jth sub-domain, i.e. $dp_j = \frac{dx}{\omega_j} \;\Rightarrow\; I_j = \omega_j \int_{\omega_j} f(x)dp_j$.

▷ Integral I_j is evaluated using the crude MC method, i.e. its MC estimator is:

$$\hat{I}_j = \frac{\omega_j}{n_j} \sum_{i=1}^{n_j} f(x_{j,i}) \,.$$

Points over each sub-domains are chosen independently, therefore the total estimator is:

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

and the variance:

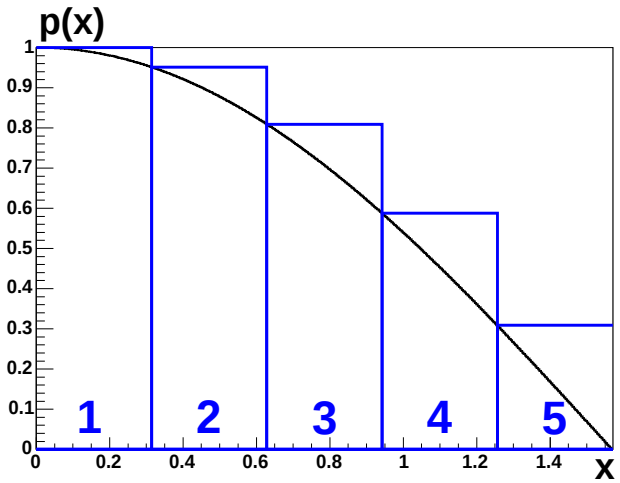
$$V(\hat{I}) = \sum_{j=1}^{k} \frac{\omega_j^2}{n_j} V_j(f) \quad \rightarrow \text{ its MC estimator : } \hat{V}(\hat{I}) = \sum_{j=1}^{k} \frac{\omega_j^2}{n_j - 1} \hat{V}_j(f) \,,$$

where $V_j(f)$ is the variance of the function f in the jth sub-domain.

It can be easily shown that uniform stratification ($\omega_j = \omega_l$ and $n_j = n_l$ for all j, l) cannot increase the variance and will, in general, decrease it, if the expectation of the function is different in the different sub-domains. (→ Try for the case $k = 2$.)

Stratified sampling for Buffon's needle

Let's, for simplicity, take $l = L$ and perform uniform stratification with $k = 5$ sub-domains:



We have: $\omega_j = \frac{\Omega}{5} = \frac{\pi}{10}$ and $n_j = \frac{N}{5}$.

▷ MC estimator of the total probability of the hit:

$$\hat{P} = \frac{1}{\Omega} \sum_{j=1}^{5} \frac{\omega_j}{n_j} \sum_{i=1}^{N/5} p(x_{j,i}) = \frac{1}{N} \sum_{i=1}^{N} p(x_i) \, .$$

▷ Standard deviation of MC estimate of π:

$$\boxed{\sigma_\pi^{\text{strat}} \simeq \frac{0.345}{\sqrt{N}}} \quad < \quad \sigma_\pi^{\text{crude}} \simeq \frac{1.52}{\sqrt{N}} \, .$$

- In the above uniform stratification one generates the same number of points in each sub-interval, independently of their contribution to the total integral.
- This can be improved either by generating the numbers of point proportionally to the area of the blue rectangles or by dividing the integration interval into the sub-intervals for which the corresponding rectangles have equal area. In such cases, one can also obtain unweighted events.
- Uniform stratification is the safest choice if we know nothing about an integrand, but it is suited only for integration – one cannot obtain unweighed events.

Importance sampling

▷ A large variation in the value of the funtion f leads to a large uncertainty in the MC estimate of its integral.

▶ MC calculations will be more efficient when each point has nearly the same weight.

⟶ A possible solution: **a change of integration variable(s) (mapping)**

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

● Importance sampling – the basic scheme:

(i) Points are generated according to $G(x)$ instead of uniformly.

(ii) The weight $w(x) = f(x)/g(x)$ is calculated for each point.

(iii) The expectation $\hat{E}_G(w)$ and the variance $\hat{V}_G(w)$ are evaluated for the whole sample.

▶ If $g(x)$ is appropriately chosen, the variance of the weight $w = f/g$ can be much smaller than the variance of the function f! (one may even have $w = \text{const}$).

▶ One can easily obtain unweighed events by applying hit-or-miss MC to the weight $w(x)$.

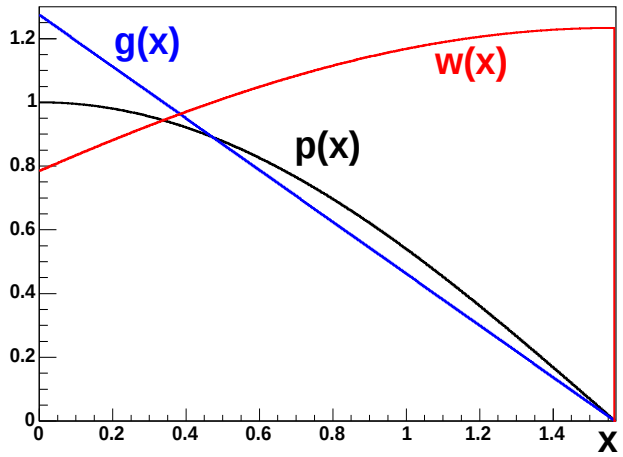
● Requirements:

▷ The function $g(x)$ should be non-negative and integrable analytically.

▷ The cumulative function $G(x)$ can be inverted analytically (only handful of functions). or g-distributed random number generator is available.

Importance sampling for Buffon's needle

▷ Let $l = L$, for simplicity.



Let's take: $g(x) = \frac{4}{\pi}\left(1 - \frac{2}{\pi}x\right)$.

$\Rightarrow G(x) = \frac{4}{\pi}x\left(1 - \frac{x}{\pi}\right)$ – invertible analytically.

▶ The weight function:

$$w(x) = \frac{p(x)}{g(x)} = \frac{\pi}{4}\frac{\cos x}{1-2x/\pi} \quad \text{and} \quad \frac{\pi}{4} \leq w \leq \frac{\pi^2}{8}\,.$$

▷ Standard deviation of MC estimate of π:

$$\boxed{\sigma_\pi^{\text{imp}} \simeq \frac{0.406}{\sqrt{N}}} \quad < \quad \sigma_\pi^{\text{crude}} \simeq \frac{1.52}{\sqrt{N}}\,.$$

- We can do even better by chosing: $g(x) = \cos x$.

▷ Then: $G(x) = \sin x$ – invertible analytically.

▶ The weight function: $w(x) \equiv 1$!!!

▷ This means that we can calculate our integral exactly, i.e. the variance: $V(w) = 0$!!!

- **Importance sampling is one of the most basic and useful Monte Carlo techniques, but it has to be used with care!** (e.g. zeros of the function $g(x)$ may be dangerous!)

We skip other classical variance-reduction techniques, such as **control variates** and **antithetic variates**, because they cannot be used for event generation.

Adaptive variance-reducing techniques

- All presented so far variance-reduction methoths, with the possible exception of uniform stratification, require some advance knowledge of behaviour of the function.
- A natural extension is toward **adaptive** techniques which learn about the function as they proceed, preferably requiring no *a priori* knoledge about the function.
- A typical scheme of adaptive algorithm $\longrightarrow$ two phases:
  1. **Exploration phase**: Usually recursive division (stratification) of the integration region into sub-regions (hyper-rectangles, simplices, etc.).
  2. **Calculation phase**: Based on the results of the exploration phase, the integral is evaluated. Sometimes there is also a possibility to generate points (events) according to a distribution given by the integrand.
- Examples of adaptive algorithms (in terms of computer programs):
  - ▷ RIWIAD of Sheppey & Lautrup: one of the earliest multidimensional MC integrators.
  - ▷ DIVONNE2 of J. Friedman, SLAC CGTM No. 188 (1977): MC integrator.
  - ▷ VEGAS of G. P. Lepage, J. Comp. Phys. **27** (1978) 192: MC integrator.
  - ▷ FOAM of S. Jadach, CPC **130** (2000) 244: both MC integrator and event generator.

Comparisons of MC integration with numerical quadrature

- All quadrature formulae approximate the value of the integral by a linear combination of function values:

$$I_Q = \sum_{i=1}^{m} w_i f(x_i) \, .$$

Different methods correspond to different choices of the points x and the weights w.

Note: The crude Monte Carlo method can be considered as a quadrature formula with unit weights and point chosen uniformly but randomly.

- Efficiencies of integration methods:

Uncertainty as a function of number of points n	In 1 dimension	In d dimensions
Monte Carlo	$n^{-1/2}$	$n^{-1/2}$
Trapezoidal rule	n^{-2}	$n^{-2/d}$
Simpson's rule	n^{-4}	$n^{-4/d}$
m-point Gauss rule	n^{-2m}	$n^{-2m/d}$ $(m < n)$

- However, quadrature methods are difficult to apply in many dimensions (> 2), for complicated integration regions and integration errors are not easy to estimate!

Summary

- A Monte Carlo method is, in brief, a technique which uses random numbers to solve a problem.
- Formally, the Monte Carlo method is based on two basic theorems of the mathematical statistics: the **Law of Large Numbers** and the **Central Limit Theorem**.
- Convergence of the MC calculation is slow – inversely proportional to the square root of the number of generated random points, however, there exist efficient variance-reducing techniques.
- MC methods can be used to solve various kinds of problems, not only the ones of stochastic nature.
- Most commonly, the MC methods are used for complicated multi-dimensional integration and event generation according to multi-dimensional distributions.
- MC techniques were first applied in physics, but now they are also used in many other areas, such as: biology, chemistry, biomedical technology, engineering, telecommunication, economics, sociology, etc.