Monte Carlos — Lecture I

Stefan Gieseke

Institut für Theoretische Physik
Universität Karlsruhe

2nd Terascale MC School, 20-23 April 2009

Outline

- Lecture I — Basics:
  - Introduction
  - Monte Carlo techniques
- Lecture II — Perturbative physics
  - Hard scattering
  - Parton showers
- Lecture III — Non–perturbative physics
  - Hadronization
  - Hadronic decays
  - Underlying event
  - MC programs

Outline Lecture I

Basics

- Introduction, motivation
- Monte Carlo event generators
- Monte Carlo methods
  - Hit and miss
  - Simple MC integration
  - Variance reduction
  - Multichannel MC

Thanks

Thanks to my colleagues

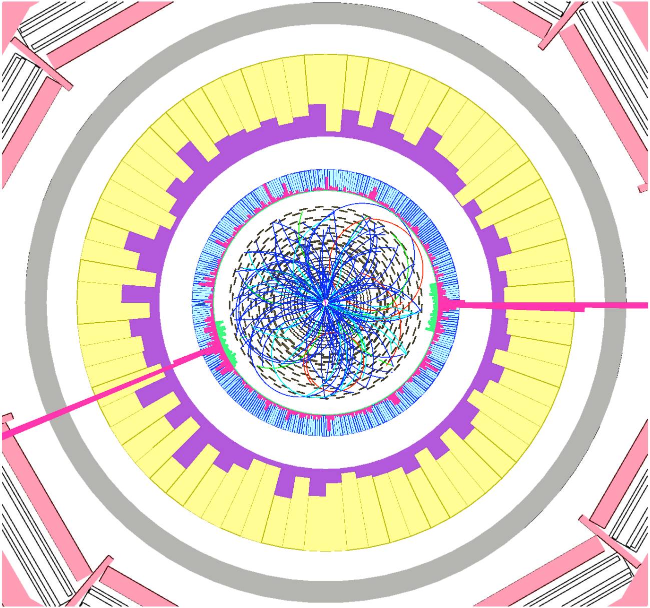
Frank Krauss, Leif Lönnblad, Steve Mrenna, Peter Richardson, Mike Seymour, Torbjörn Sjöstrand.

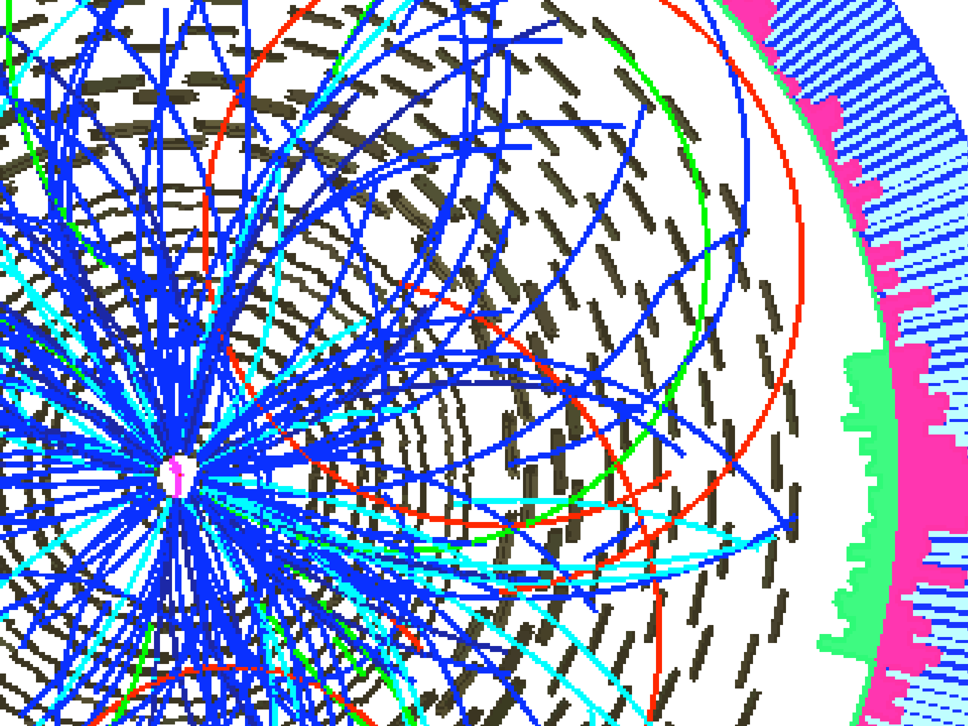
Why Monte Carlos?

We want to understand

$$\mathscr{L}_{\text{int}} \longleftrightarrow \text{Final states}\ .$$

Can you spot the Higgs?





Why Monte Carlos?

LHC experiments require
sound understanding of signals and *backgrounds*.

$\uparrow$

Full detector simulation.

$\uparrow$

Fully exclusive hadronic final state.

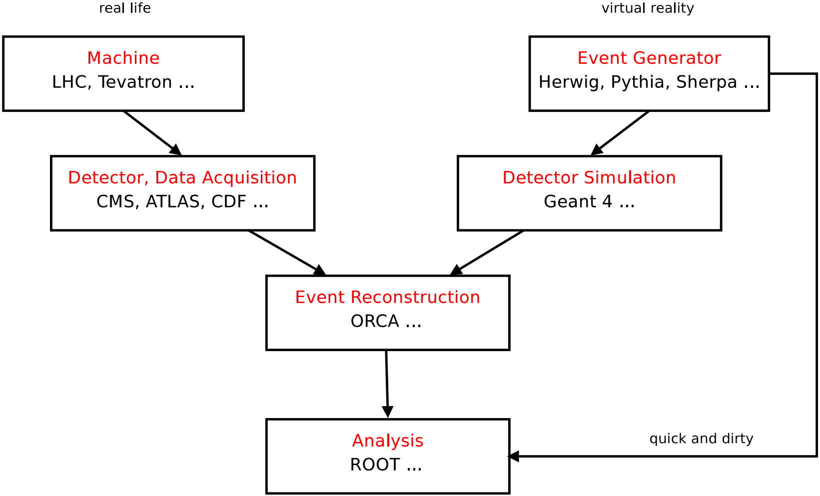
$\uparrow$

Monte Carlo event generator with
parton shower, hadronization model, decays of unstable particles.

$\uparrow$

Parton level computations.

Experiment and Simulation



Monte Carlo Event Generators

- Complex final states in full detail (jets).
- Arbitrary observables and cuts from final states.
- Studies of new physics models.

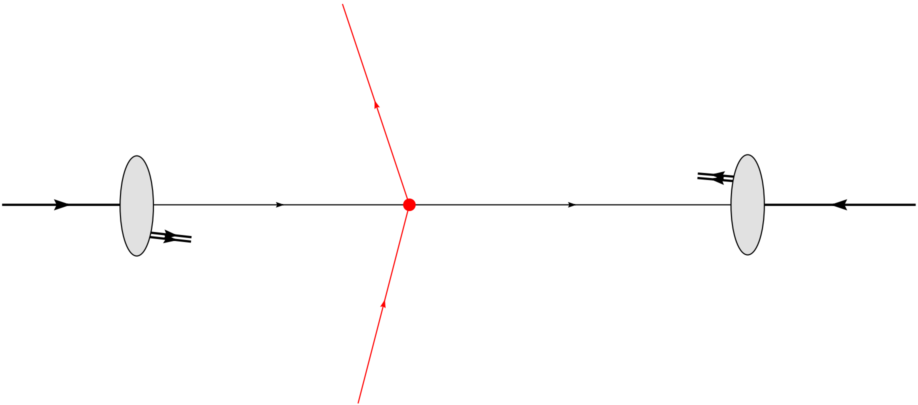
- Rates and topologies of final states.
- Background studies.
- Detector Design.
- Detector Performance Studies (Acceptance).

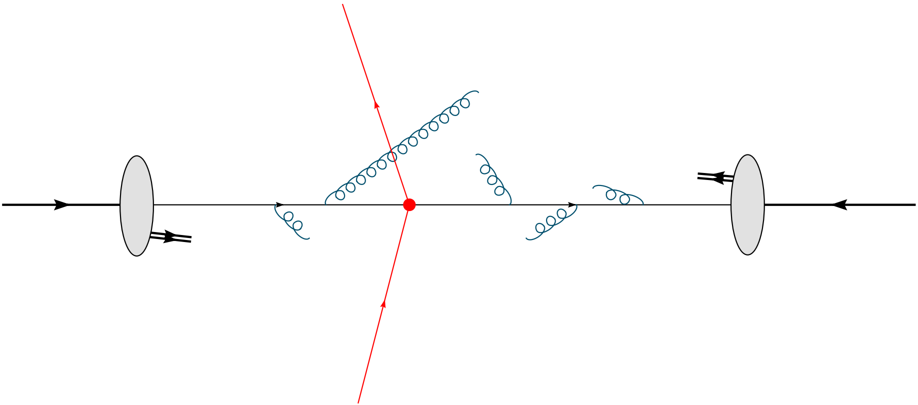
- *Obvious* for calculation of observables on the quantum level

$$|A|^2 \longrightarrow \text{Probability}.$$

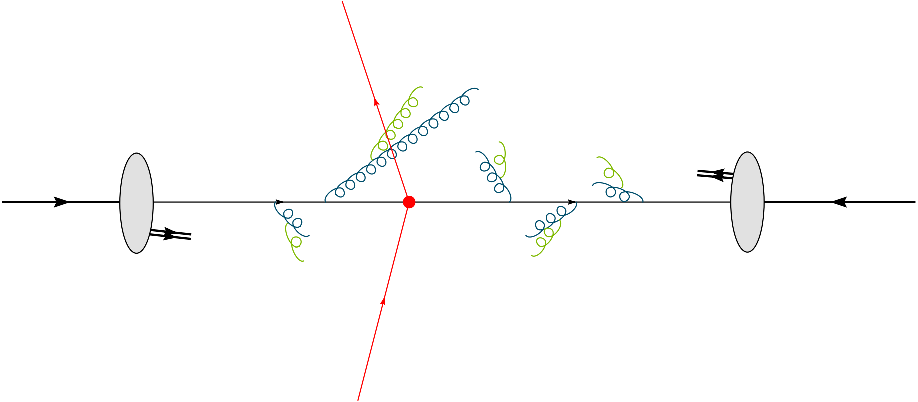
pp Event Generator

pp Event Generator

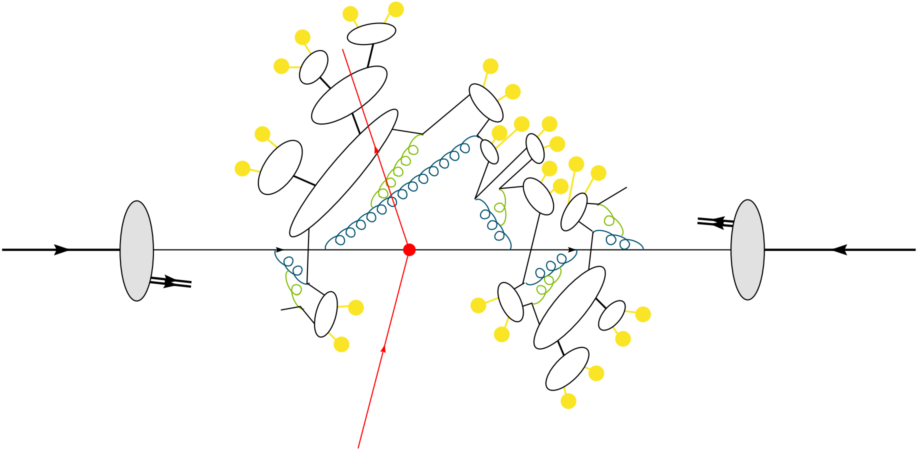


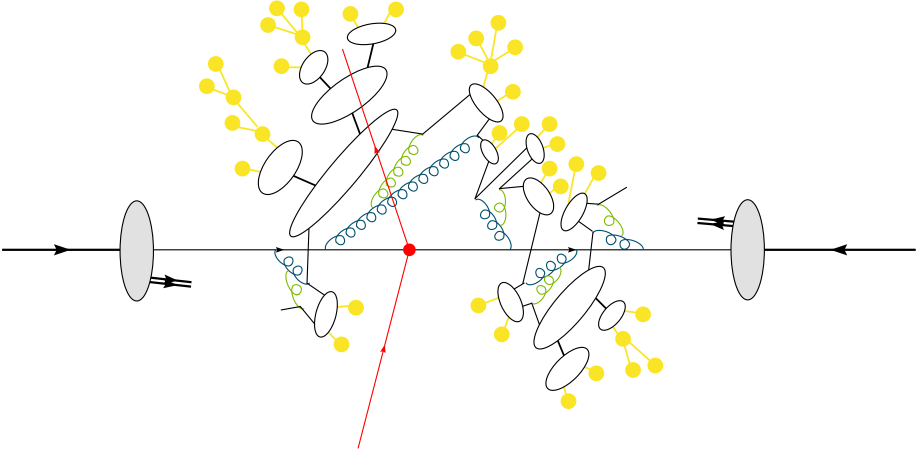


pp Event Generator

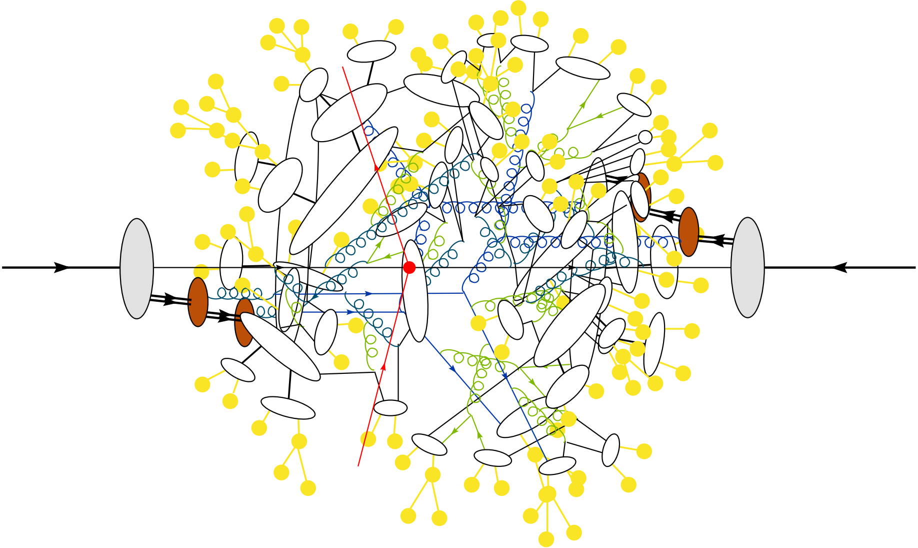


pp Event Generator





pp Event Generator



Divide and conquer

Partonic cross section from Feynman diagrams

$$d\sigma = d\sigma_{\rm hard} dP(\text{partons} \to \text{hadrons})$$

Note, that

$$\int dP(\text{partons} \to \text{hadrons}) = 1 \,,$$

- σ remains unchanged
- introduce realistic fluctuations into distributions.

Divide and conquer

Partonic cross section from Feynman diagrams

$$\mathrm{d}\sigma = \mathrm{d}\sigma_{\mathrm{hard}} \mathrm{d}P(\text{partons} \to \text{hadrons})$$

Note, that

$$\int \mathrm{d}P(\text{partons} \to \text{hadrons}) = 1\ ,$$

- σ remains unchanged
- introduce realistic fluctuations into distributions.

Simulation steps governed by different scales
$\longrightarrow$ separation into ($Q_0 \approx 1\,\mathrm{GeV} > \Lambda_{\mathrm{QCD}}$)

$$\begin{aligned}
\mathrm{d}P(\text{partons} \to \text{hadrons}) = \ & \mathrm{d}P(\text{resonance decays}) && [\Gamma > Q_0] \\
& \times \mathrm{d}P(\text{parton shower}) && [\mathrm{TeV} \to Q_0] \\
& \times \mathrm{d}P(\text{hadronisation}) && [\sim Q_0] \\
& \times \mathrm{d}P(\text{hadronic decays}) && [O(\mathrm{MeV})]
\end{aligned}$$

Divide and conquer

$$
\begin{aligned}
\mathrm{d}P(\text{partons} \to \text{hadrons}) = \; & \mathrm{d}P(\text{resonance decays}) && [\Gamma > Q_0] \\
& \times \mathrm{d}P(\text{parton shower}) && [\text{TeV} \to Q_0] \\
& \times \mathrm{d}P(\text{hadronisation}) && [\sim Q_0] \\
& \times \mathrm{d}P(\text{hadronic decays}) && [O(\text{MeV})]
\end{aligned}
$$

Quite complicated integration.

Divide and conquer

$$\begin{aligned}
\mathrm{d}P(\text{partons} \to \text{hadrons}) = \ & \mathrm{d}P(\text{resonance decays}) & [\Gamma > Q_0] \\
& \times \mathrm{d}P(\text{parton shower}) & [\text{TeV} \to Q_0] \\
& \times \mathrm{d}P(\text{hadronisation}) & [\sim Q_0] \\
& \times \mathrm{d}P(\text{hadronic decays}) & [O(\text{MeV})]
\end{aligned}$$

Quite complicated integration.

Monte Carlo is the only choice.

Monte Carlo Methods

Introduction to the most important MC sampling
(= integration) techniques.

1. Hit and miss.
2. Simple MC integration.
3. (Some) methods of variance reduction.
4. Multichannel.

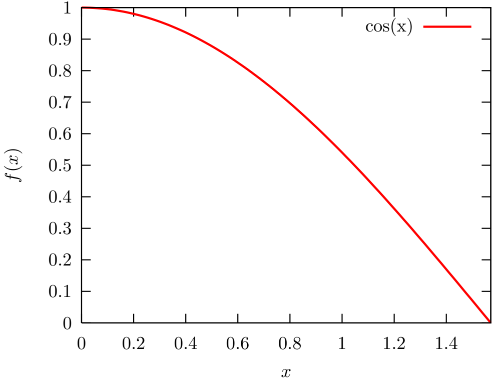
Probability

Probability density:

$$dP = f(x)\,dx$$

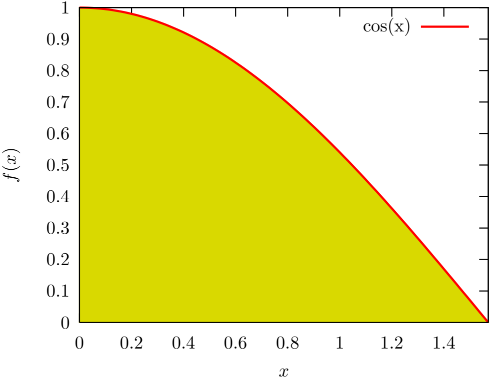
is probability to find value x.

Example: $f(x) = \cos(x)$.



Probability

Example: $f(x) = \cos(x)$.



Probability density:

$$dP = f(x)\,dx$$

is probability to find value x.

$$F(x) = \int_{x_0}^{x} f(x)\,dx$$

is called *probability distribution*.

Probability

Probability density:

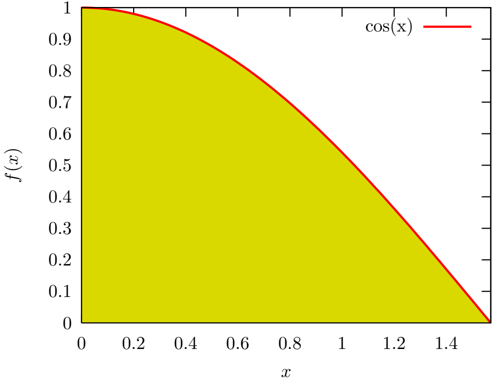
$$dP = f(x)\,dx$$

is probability to find value x.

$$F(x) = \int_{x_0}^{x} f(x)\,dx$$

is called *probability distribution*.

Example: $f(x) = \cos(x)$.



Probability ∼ Area

Hit and Miss

Hit and miss method:

- throw N random points (x, y) into region.
- Count hits N_{hit}, i.e. whenever $y < f(x)$.

Then

$$I \approx V \frac{N_{\text{hit}}}{N}.$$

approaches 1 again in our example.

Hit and Miss

Hit and miss method:

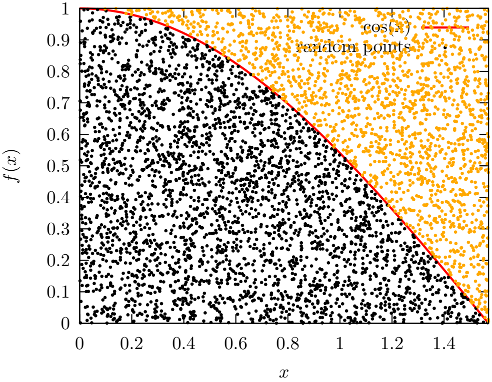
- throw N random points (x, y) into region.
- Count hits N_{hit}, i.e. whenever $y < f(x)$.

Then

$$I \approx V \frac{N_{\text{hit}}}{N}.$$

approaches 1 again in our example.

Example: $f(x) = \cos(x)$.



Hit and Miss

Hit and miss method:

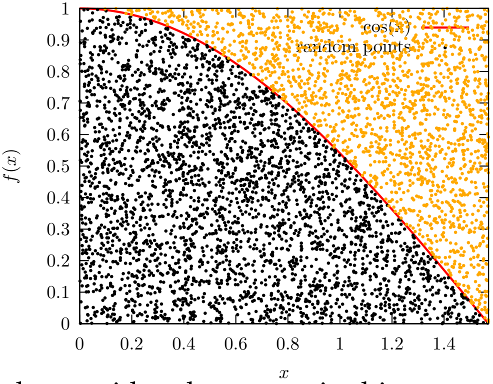
- throw N random points (x, y) into region.
- Count hits N_{hit}, i.e. whenever $y < f(x)$.

Then

$$I \approx V \frac{N_{\text{hit}}}{N}.$$

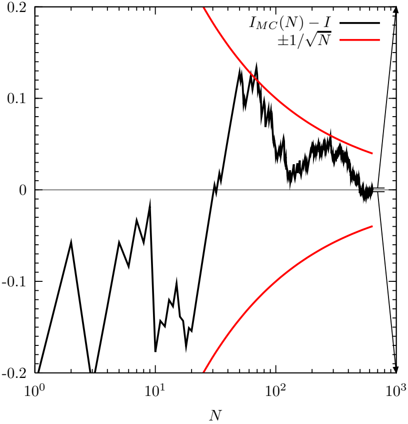
approaches 1 again in our example.

Example: $f(x) = \cos(x)$.



Every accepted value of x can be considered an event in this picture. As $f(x)$ is the 'histogram' of x, it seems obvious that the x values are distributed as $f(x)$ from this picture.

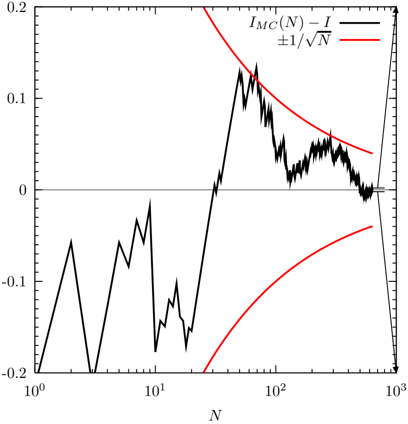
Hit and Miss



How well does it converge?

Error $1/\sqrt{N}$.

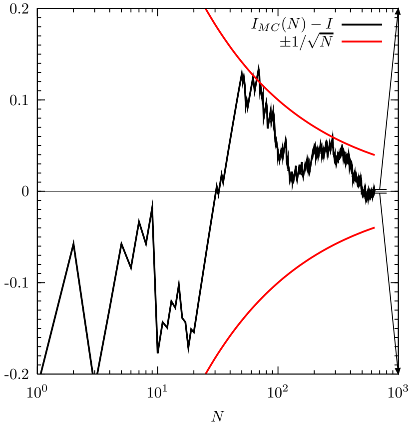
Hit and Miss

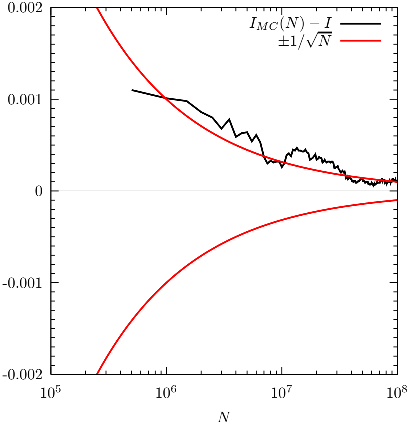


More points, zoom in…

Error $1/\sqrt{N}$.

Hit and Miss





Error $1/\sqrt{N}$.

Hit and Miss

This method is used in many event generators. However, it is not sufficient as such.

- Can handle any density $f(x)$,
  however wild and unknown it is.
- $f(x)$ should be bounded from above.
- Sampling will be very *inefficient* whenever $\mathrm{Var}(f)$ is large.

Improvements go under the name variance reduction as they improve the error of the crude MC at the same time.

Simple MC integration

Mean value theorem of integration:

$$
\begin{aligned}
I &= \int_{x_0}^{x_1} f(x)\,dx \\
&= (x_1 - x_0)\langle f(x)\rangle
\end{aligned}
$$

(Riemann integral).

Simple MC integration

Mean value theorem of integration:

$$
\begin{aligned}
I &= \int_{x_0}^{x_1} f(x)\,dx \\
&= (x_1 - x_0)\langle f(x)\rangle \\
&\approx (x_1 - x_0)\frac{1}{N}\sum_{i=1}^{N} f(x_i)
\end{aligned}
$$

(Riemann integral).

Simple MC integration

Mean value theorem of integration:

$$
\begin{aligned}
I &= \int_{x_0}^{x_1} f(x)\,dx \\
&= (x_1 - x_0)\langle f(x)\rangle \\
&\approx (x_1 - x_0)\frac{1}{N}\sum_{i=1}^{N} f(x_i)
\end{aligned}
$$

(Riemann integral).

Sum doesn't depend on ordering
$\longrightarrow$ randomize x_i.

Simple MC integration

Mean value theorem of integration:

$$
\begin{aligned}
I &= \int_{x_0}^{x_1} f(x)\,dx \\
&= (x_1 - x_0)\langle f(x)\rangle \\
&\approx (x_1 - x_0)\frac{1}{N}\sum_{i=1}^{N} f(x_i)
\end{aligned}
$$

(Riemann integral).

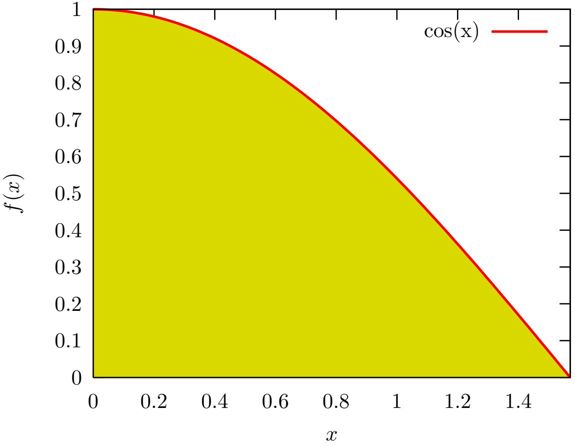
Sum doesn't depend on ordering
$\longrightarrow$ randomize x_i.

Yields a flat distribution of events x_i,
but weighted with *weight* $f(x_i)$ ($\to$ unweighting).

Simple MC integration

Pictorially:

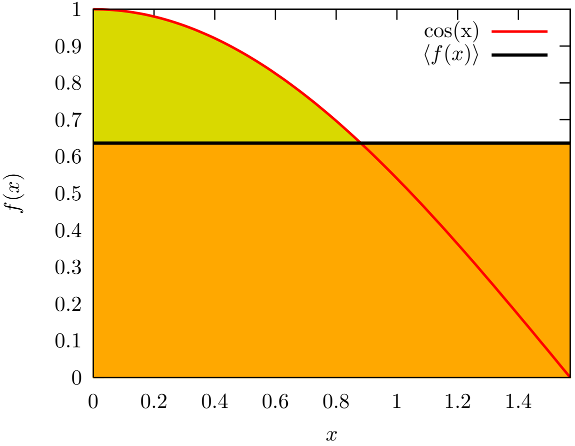
$$I = \int_{x_0}^{x_1} f(x)\,dx$$
$$= (x_1 - x_0)\langle f(x)\rangle$$



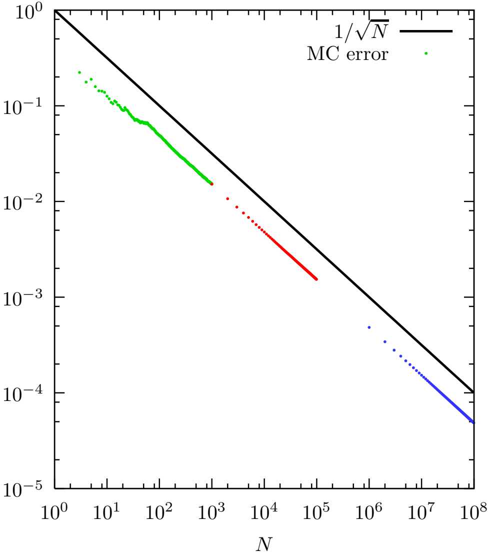
Simple MC integration

Pictorially:

$$I = \int_{x_0}^{x_1} f(x)\,dx$$
$$= (x_1 - x_0)\langle f(x)\rangle$$



What's the error?



Again, looks like

$$\sigma \sim \frac{1}{\sqrt{N}}$$

Simple MC integration

What's the error?

We can calculate it (central limit theorem for the average):

In general: *Crude MC*

$$
\begin{aligned}
I &= \int f dV \\
&\approx V\langle f\rangle \pm V\sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N}} \\
&\approx V\langle f\rangle \pm V\frac{\sigma}{\sqrt{N}}
\end{aligned}
$$

Simple MC integration

What's the error?

We can calculate it (central limit theorem for the average):

Our example: $\cos(x), 0 \leq x \leq \pi/2$,
compute σ_{MC} from

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

$$\langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} f^2(x_i).$$

What's the error?

We can calculate it (central limit theorem for the average):

Compute σ directly ($V = \pi/2$):

$$V\langle f\rangle = \int_0^{\pi/2} \cos(x)\,dx = 1$$
$$V\langle f^2\rangle = \int_0^{\pi/2} \cos^2(x)\,dx = \frac{\pi}{4}$$

then

$$\sigma = \sqrt{\frac{\pi}{2}\frac{\pi}{4} - 1^2} \approx 0.4834.$$

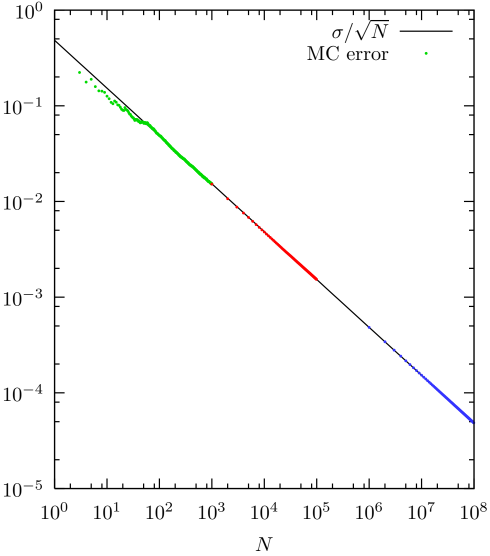
Simple MC integration

What's the error?

Now, compare

$$\sigma_{MC} = \frac{0.4834}{\sqrt{N}}$$

with error estimate from MC.



Simple MC integration

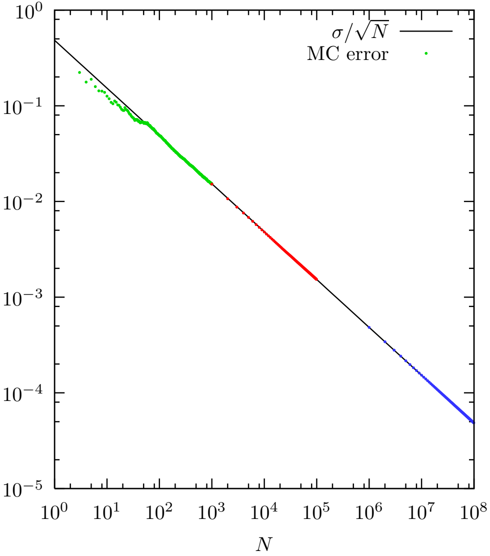
What's the error?

Now, compare

$$\sigma_{MC} = \frac{0.4834}{\sqrt{N}}$$

with error estimate from MC.

Spot on.

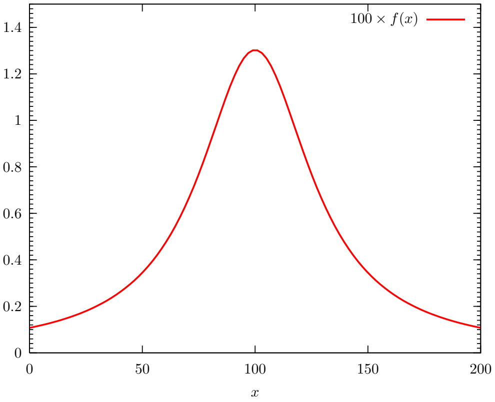


Another basic MC method, based on the observation that

Probability $\sim$ *Area*

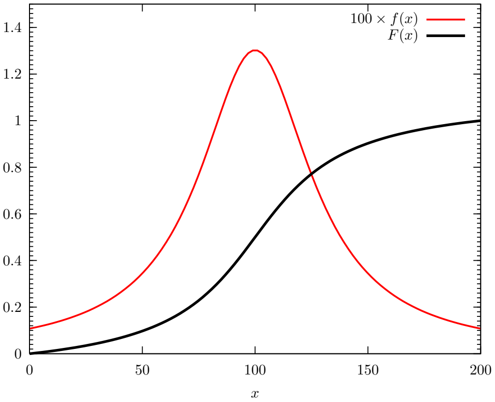
Inverting the Integral

- Probability density $f(x)$. Not necessarily normalized.



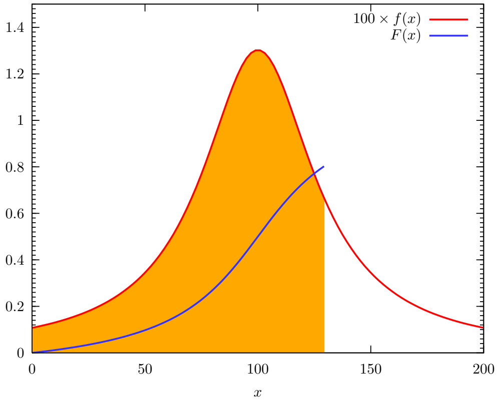
Inverting the Integral

- Probability density $f(x)$. Not necessarily normalized.
- Integral $F(x)$ known,



Inverting the Integral

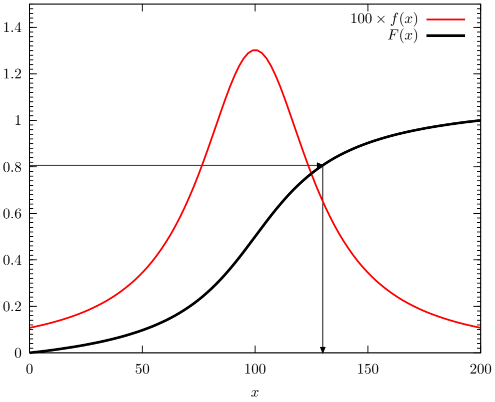
- Probability density $f(x)$. Not necessarily normalized.
- Integral $F(x)$ known,
- $P(x < x_s) = F(x_s)$.



Inverting the Integral

- Probability density $f(x)$. Not necessarily normalized.
- Integral $F(x)$ known,
- $P(x < x_s) = F(x_s)$.
- Probability = 'area', distributed evenly,

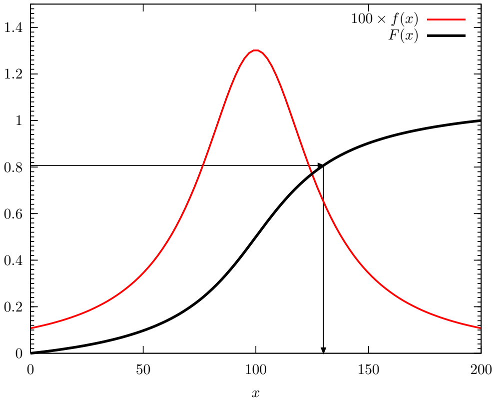
$$\int_0^r dP = r$$



Inverting the Integral

- Probability density $f(x)$. Not necessarily normalized.
- Integral $F(x)$ known,
- $P(x < x_s) = F(x_s)$.
- Probability = 'area', distributed evenly,

$$\int_0^r dP = r$$



Sample x according to $f(x)$ with

$$x = F^{-1}\Big[F(x_0) + r\big(F(x_1) - F(x_0)\big)\Big] .$$

Inverting the Integral

Another basic MC method, based on the observation that

Probability $\sim$ Area

Sample x according to $f(x)$ with

$$x = F^{-1}\Big[F(x_0) + r\big(F(x_1) - F(x_0)\big)\Big] \, .$$

Optimal method, but we need to know

- The integral $F(x)$,
- It's inverse $F^{-1}(y)$.

That's rarely the case for real problems.

But very powerful in combination with other techniques.

Importance sampling

Error on Crude MC $\sigma_{MC} = \sigma/\sqrt{N}$.
$\Longrightarrow$ Reduce error by reducing variance of integrand.

Importance sampling

Error on Crude MC $\sigma_{MC} = \sigma/\sqrt{N}$.
$\Longrightarrow$ Reduce error by reducing variance of integrand.

Idea: *Divide out the singular structure.*

$$I = \int f \, \mathrm{d}V = \int \frac{f}{p} p \mathrm{d}V \approx \left\langle \frac{f}{p} \right\rangle \pm \sqrt{\frac{\langle f^2/p^2 \rangle - \langle f/p \rangle^2}{N}} \, .$$

where we have chosen $\int p \, \mathrm{d}V = 1$ for convenience.

Note: need to sample flat in $p \, \mathrm{d}V$, so we better know $\int p \mathrm{d}V$ and it's inverse.

Importance sampling

Consider error term:

$$E = \left\langle \frac{f^2}{p^2} \right\rangle - \left\langle \frac{f}{p} \right\rangle^2 = \int \frac{f^2}{p^2} p \mathrm{d}V - \left[\int \frac{f}{p} p \mathrm{d}V \right]^2$$

$$= \int \frac{f^2}{p} \mathrm{d}V - \left[\int f \mathrm{d}V \right]^2 .$$

Importance sampling

Consider error term:

$$E = \int \frac{f^2}{p} \, \mathrm{d}V - \left[\int f \, \mathrm{d}V \right]^2 .$$

Best choice of p? Minimises $E \to$ functional variation of error term with (normalized) p:

$$\begin{aligned} 0 = \delta E &= \delta \left(\int \frac{f^2}{p} \, \mathrm{d}V - \left[\int f \, \mathrm{d}V \right]^2 + \lambda \int p \, \mathrm{d}V \right) \\ &= \int \left(-\frac{f^2}{p^2} + \lambda \right) \mathrm{d}V \delta p , \end{aligned}$$

Importance sampling

Consider error term:

$$E = \int \frac{f^2}{p} \, \mathrm{d}V - \left[\int f \, \mathrm{d}V \right]^2 .$$

Best choice of p? Minimises E $\rightarrow$ functional variation of error term with (normalized) p:

$$0 = \delta E = \int \left(-\frac{f^2}{p^2} + \lambda \right) \mathrm{d}V \delta p ,$$

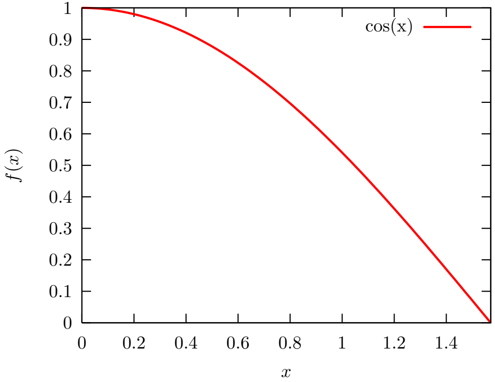
hence

$$p = \frac{|f|}{\sqrt{\lambda}} = \frac{|f|}{\int |f| \, \mathrm{d}V} .$$

Choose p as close to f as possible.

Importance sampling — example

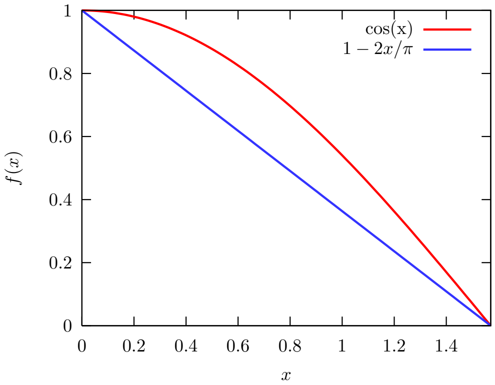
Improving $\cos(x)$ sampling,



Importance sampling — example

Improving $\cos(x)$ sampling,

$$I = \int_0^{\pi/2} \cos(x)\, dx$$
$$= \int_0^{\pi/2} \frac{\cos(x)}{1-\frac{2}{\pi}x} \left(1-\frac{2}{\pi}x\right) dx$$
$$= \int_0^{\pi/4} \left.\frac{\cos(x)}{1-\frac{2}{\pi}x}\right|_{x=x(\rho)} d\rho\ .$$



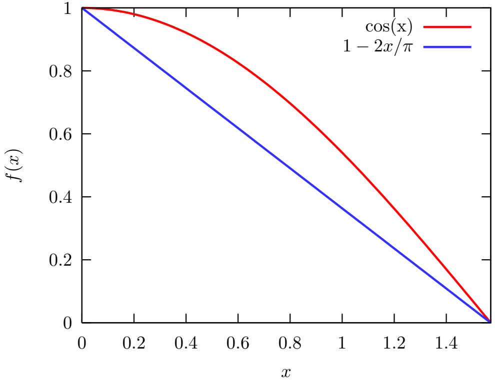
Importance sampling — example

Improving $\cos(x)$ sampling,

$$I = \int_0^{\pi/2} \cos(x)\, dx$$

$$= \int_0^{\pi/2} \frac{\cos(x)}{1-\frac{2}{\pi}x} \left(1-\tfrac{2}{\pi}x\right) dx$$

$$= \int_0^{\pi/4} \left.\frac{\cos(x)}{1-\frac{2}{\pi}x}\right|_{x=x(\rho)} d\rho\ .$$



Sample x with *inverting the integral* technique (flat random number ρ),

$$x = \frac{\pi}{2} - \sqrt{\frac{\pi^2}{4} - \pi\rho}\ .$$

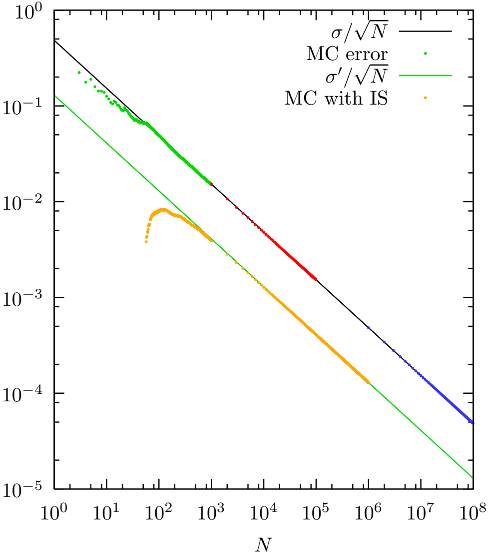
Importance sampling — example

Improving $\cos(x)$ sampling,

much better convergence,

about 80% “accepted events”.

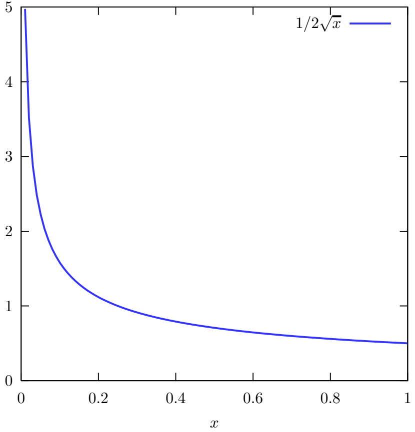
Reduced variance $\Rightarrow$ better efficiency.



Importance sampling — better example

More interesting for divergent integrands, eg

$$\frac{1}{2\sqrt{x}} \ ,$$



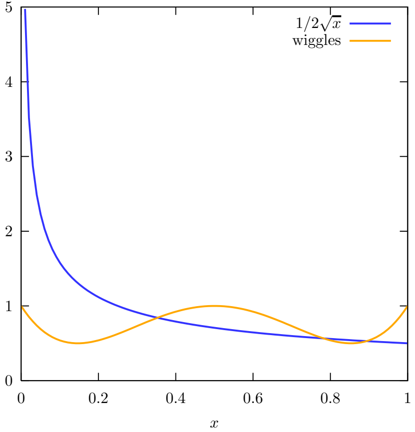
Importance sampling — better example

More interesting for divergent integrands, eg

$$\frac{1}{2\sqrt{x}}\ ,$$

with some wiggles,

$$p(x) = 1 - 8x + 40x^2 - 64x^3 + 32x^4\ .$$



Importance sampling — better example

More interesting for divergent integrands, eg

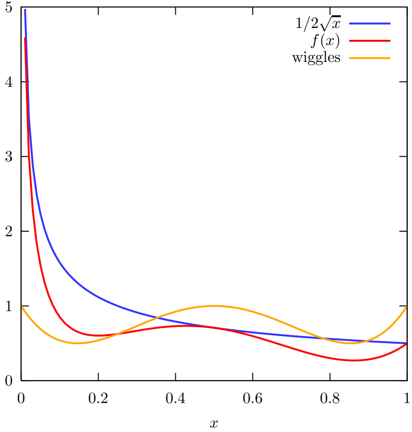
$$\frac{1}{2\sqrt{x}} \,,$$

with some wiggles,

$$p(x) = 1 - 8x + 40x^2 - 64x^3 + 32x^4 \,.$$

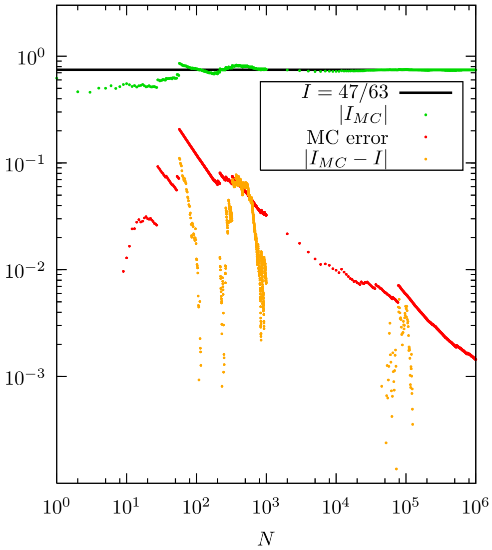
i.e. we want to integrate

$$f(x) = \frac{p(x)}{2\sqrt{x}} \,.$$



Importance sampling — better example

- Crude MC gives result in reasonable 'time'.
- Error a bit unstable.
- Event generation with maximum weight $w_{\max} = 20$. (that's arbitrary.)
- hit/miss/events with $(w > w_{\max}) =$ 36566/963434/617 with 1M generated events.

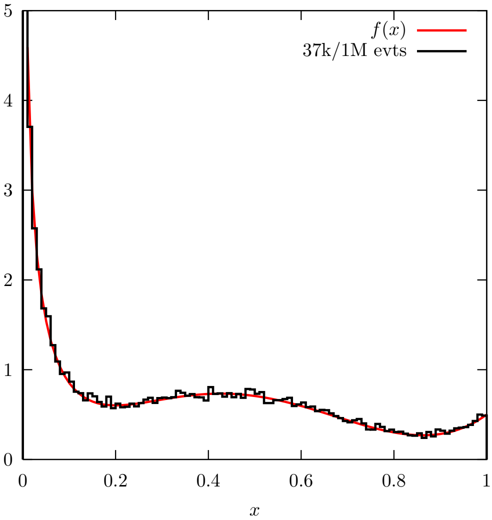


Importance sampling — better example

Want events:

use hit+mass variant here:

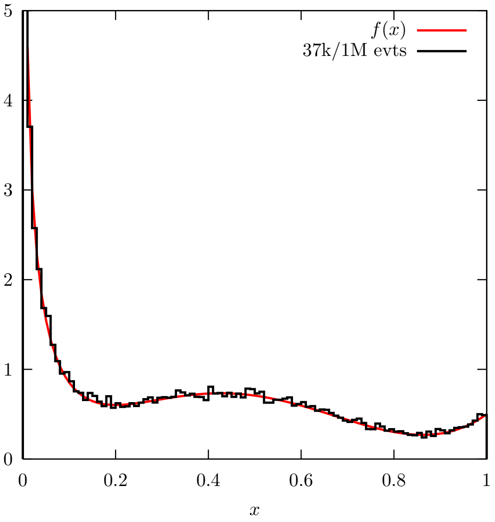
- Choose new random number r
- $w = f(x)$ in this case.
- if $r < w/w_{\max}$ then "hit".
- MC efficiency = hit$/N$.



Importance sampling — better example

Want events:
use hit+mass variant here:

- Choose new random number r
- $w = f(x)$ in this case.
- if $r < w/w_{\max}$ then "hit".
- MC efficiency = hit$/N$.
- Efficiency for MC events only 3.7%.
- Note the wiggly histogram.



Importance sampling — better example

Now importance sampling, i.e. divide out $1/2\sqrt{x}$.

$$\int_0^1 \frac{p(x)}{2\sqrt{x}}\,dx = \int_0^1 \left(\frac{p(x)}{2\sqrt{x}} \Big/ \frac{1}{2\sqrt{x}} \right) \frac{dx}{2\sqrt{x}}$$
$$= \int_0^1 p(x)\, d\sqrt{x}$$
$$= \int_0^1 p(x(\rho))\, d\rho$$

so,

$$\rho = \sqrt{x}, \qquad d\rho = \frac{dx}{2\sqrt{x}}$$

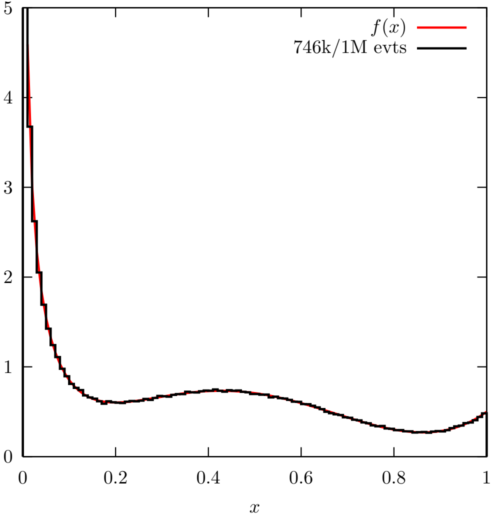
x sampled with *inverting the integral* from flat random numbers ρ, $x = \rho^2$.

Importance sampling — better example

$$\int_0^1 \frac{p(x)}{2\sqrt{x}}\,dx = \int_0^1 p(x(\rho))\,d\rho$$

with

$$\rho = \sqrt{x}, \qquad d\rho = \frac{dx}{2\sqrt{x}}$$



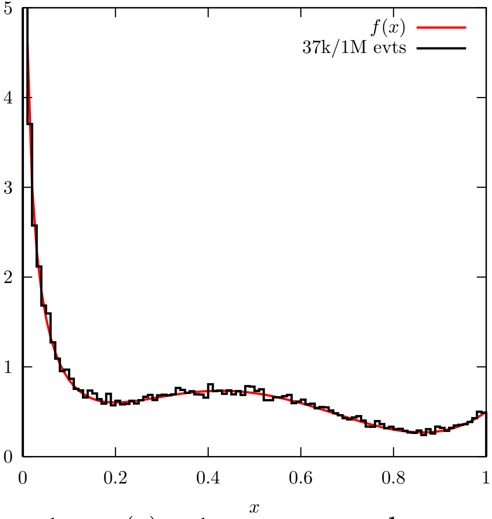
Events generated with $w_{\max} = 1$, as $p(x) \leq 1$, no guesswork needed here! Now, we get 74.6% MC efficiency.

Importance sampling — better example

$$\int_0^1 \frac{p(x)}{2\sqrt{x}}\,dx = \int_0^1 p(x(\rho))\,d\rho$$

with

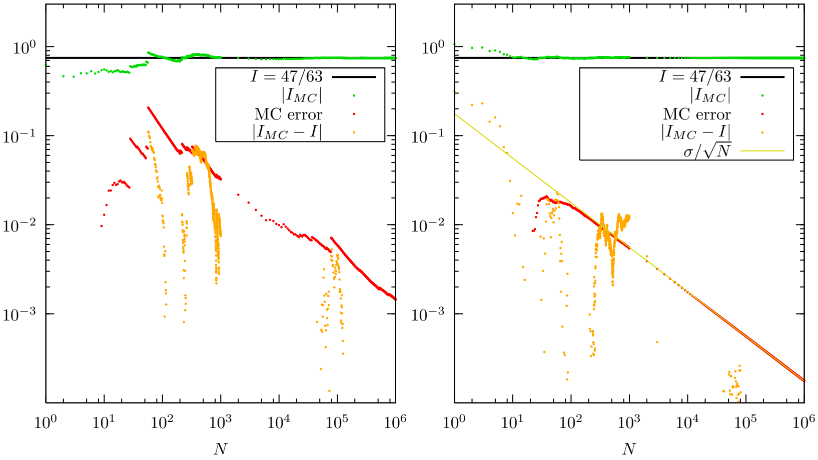
$$\rho = \sqrt{x}, \qquad d\rho = \frac{dx}{2\sqrt{x}}$$



Events generated with $w_{\max} = 1$, as $p(x) \leq 1$, no guesswork needed here! Now, we get 74.6% MC efficiency.
…as opposed to 3.7%.

Importance sampling — better example

Crude MC vs Importance sampling.



100× more events needed to reach same accuracy.

Importance sampling — another useful example

Breit–Wigner peaks appear in many realistic MEs for cross sections and decays.

$$I = \int_{s_0}^{s_1} \frac{ds}{(s-m^2)^2 + m^2\Gamma^2}$$

Importance sampling — another useful example

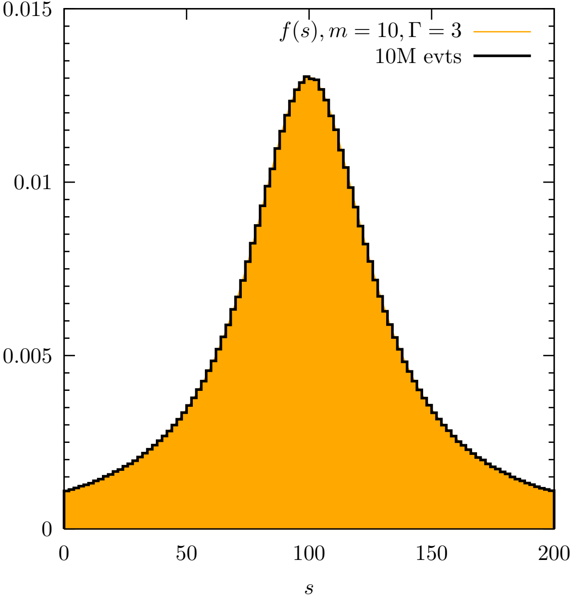
Breit–Wigner peaks appear in many realistic MEs for cross sections and decays.

$$I = \int_{s_0}^{s_1} \frac{ds}{(s-m^2)^2 + m^2\Gamma^2} = \frac{1}{m\Gamma} \int_{y_0}^{y_1} \frac{dy}{y^2+1} \qquad (y = \frac{s-m^2}{m\Gamma})$$
$$= \frac{1}{m\Gamma} \arctan \frac{s-m^2}{m\Gamma} \Big|_{s_0}^{s_1}$$

Inverting the integral gives ("tan mapping").

$$f(s) = \frac{m\Gamma}{(s-m^2)^2 + m^2\Gamma^2} \ ,$$
$$F(s) = \arctan \frac{s-m^2}{m\Gamma} = \rho \ ,$$
$$s = F^{-1}(\rho) = m^2 + m\Gamma \tan\rho \ .$$

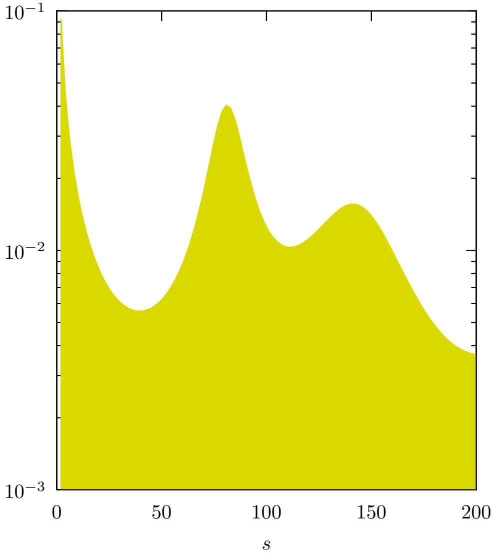
Importance sampling — another useful example



Multichannel MC

Typical problem:

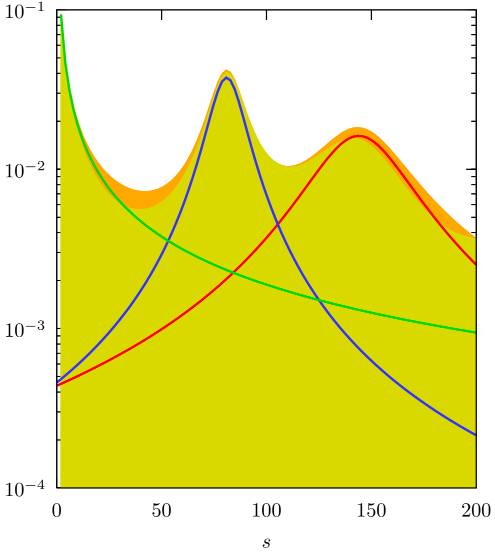
- $f(s)$ has multiple peaks ($\times$ wiggles from ME).



Multichannel MC

Typical problem:

- $f(s)$ has multiple peaks ($\times$ wiggles from ME).
- Usually have some idea of the peak structure.

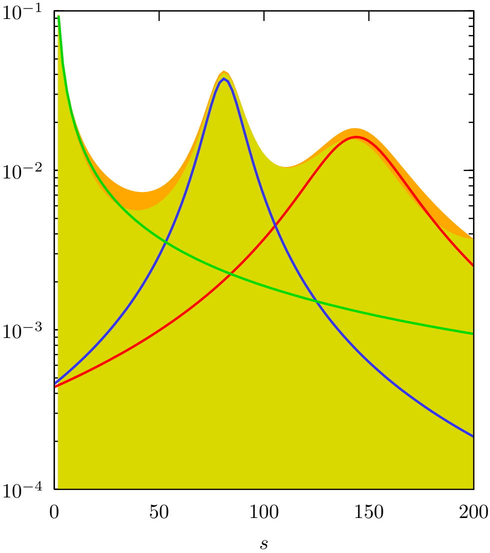


Multichannel MC

Typical problem:

- $f(s)$ has multiple peaks ($\times$ wiggles from ME).
- Usually have some idea of the peak structure.
- Encode this in sum of sample functions $g_i(s)$ with weights $\alpha_i, \sum_i \alpha_i = 1$.

$$g(s) = \sum_i \alpha_i g_i(s) \ .$$



Multichannel MC

Now rewrite

$$\int_{s_0}^{s_1} f(s)\,ds = \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g(s)\,ds$$
$$= \int_{s_0}^{s_1} \frac{f(s)}{g(s)} \sum_i \alpha_i g_i(s)\,ds$$
$$= \sum_i \alpha_i \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g_i(s)\,ds$$

Now $g_i(s)\,ds = d\rho_i$ (inverting the integral).

Now rewrite

$$\begin{aligned}\int_{s_0}^{s_1} f(s)\,ds &= \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g(s)\,ds \\ &= \int_{s_0}^{s_1} \frac{f(s)}{g(s)} \sum_i \alpha_i g_i(s)\,ds \\ &= \sum_i \alpha_i \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g_i(s)\,ds\end{aligned}$$

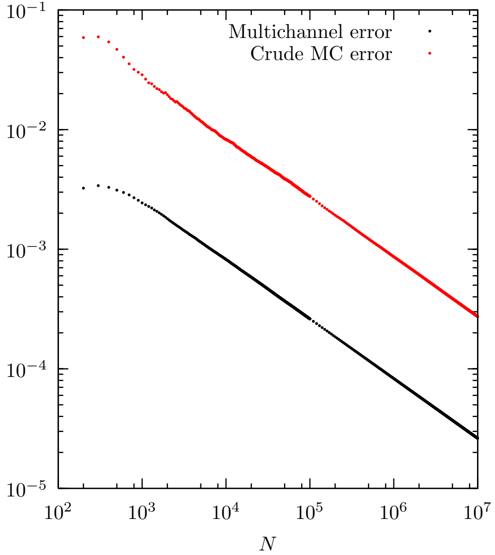
Now $g_i(s)\,ds = d\rho_i$ (inverting the integral).

Select the distribution $g_i(s)$ you'd like to sample next event from acc to weights α_i.

α_i can be optimized after a number of trials.

Multichannel MC

Works quite well:



Final Remarks/Real Life MC

- Didn't discuss random number generators. Please make sure to use 'good' random numbers (eg those that come with `CLHEP`).
- Didn't discuss *stratified sampling* (VEGAS). Sample where variance is biggest.
  (not necessarily where PS is most populated).
- Only discussed one–dimensional case here. N–particle PS has $3N - 4$ dimensions…
- Didn't discuss tools geared towards this, like `RAMBO` (generates flat N particles PS).
- generalisation straightforward, particularly
  MCError $\sim \frac{1}{\sqrt{N}}$,
  compare eg Trapezium rule Error $\sim \frac{1}{N^{2/D}}$.
- Many important techniques covered here in detail! Should be good starting point.

Outline Lecture I

Basics

- Introduction, motivation
- Monte Carlo event generators
- Monte Carlo methods
  - Hit and miss
  - Simple MC integration
  - Variance reduction
  - Multichannel MC

Outline

- Lecture I — Basics:
  - Introduction
  - Monte Carlo techniques
- Lecture II — Perturbative physics
  - Hard scattering
  - Parton showers
- Lecture III — Non–perturbative physics
  - Hadronization
  - Hadronic decays
  - Underlying event
  - MC programs