Monte Carlo Tutorial

Stefan Gieseke

Universität Karlsruhe (TH)
Institut für Theoretische Physik

- Quick tour of a MC event.
- Lecture I: Parton Shower Formalism (blackboard).
- Lecture II: Monte Carlo Methods.

Monte Carlos?! Why?

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

↑

Full detector simulation.

↑

Fully exclusive hadronic final state.

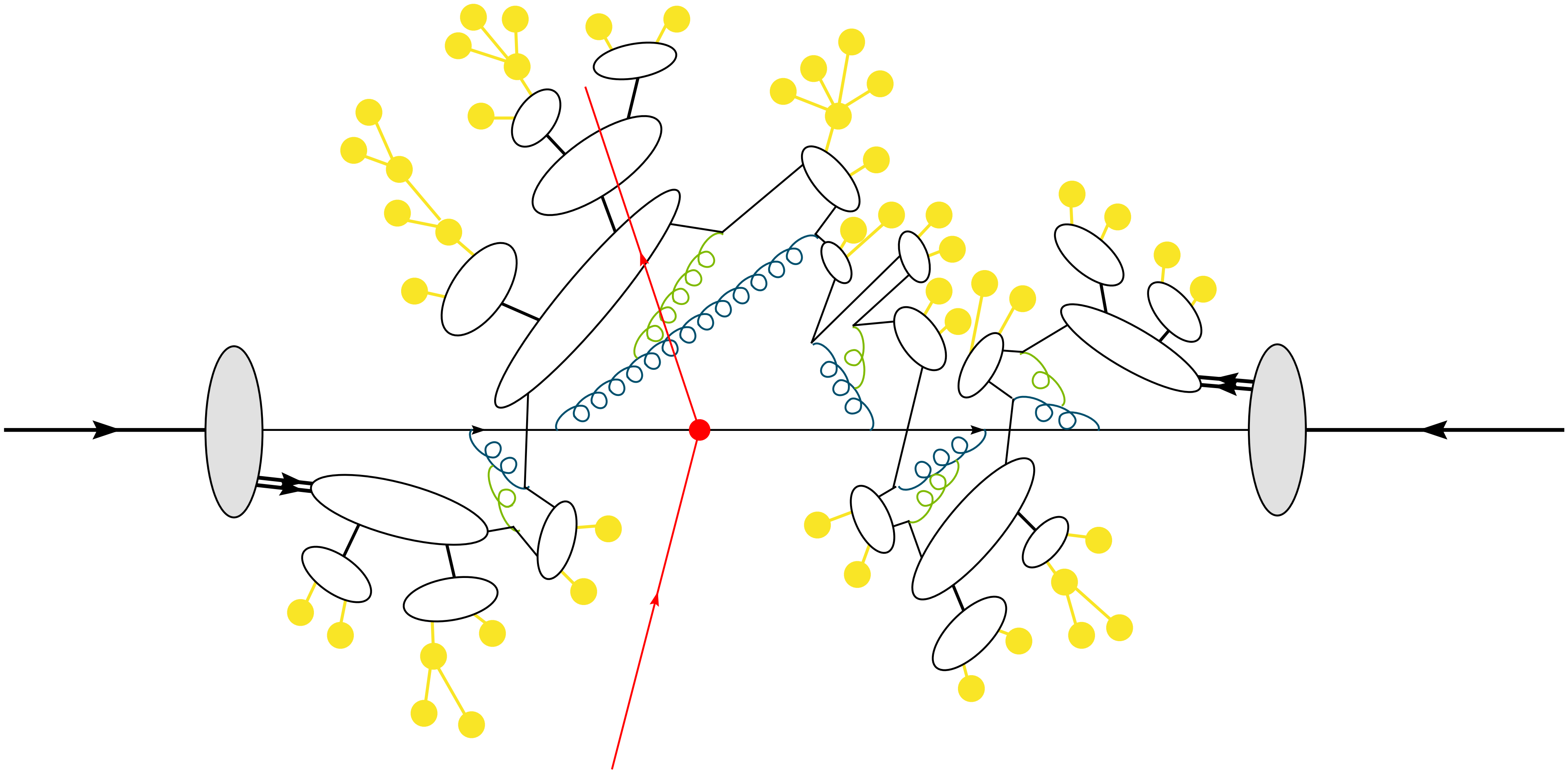
↑

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

↑

Parton level computations.

pp Event Generator



Observable $\Longleftarrow$ Convolution of all simulation stages $\otimes$ detector simulation

That's a *very* complicated integral!

Lecture II: Monte Carlo Methods

Plan

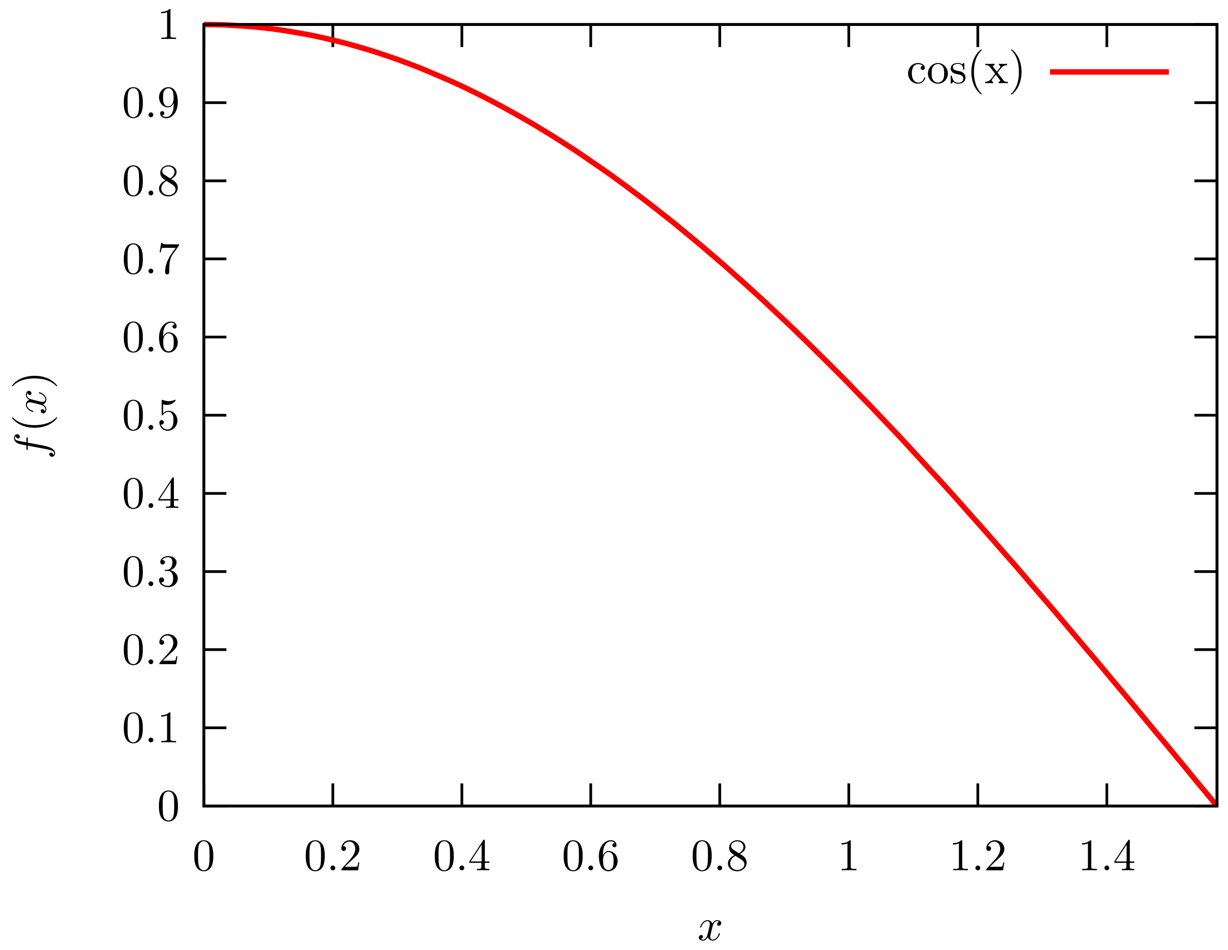
1. Simple MC integration
2. Hit and Miss
3. Importance Sampling
4. Multichannel MC
5. Final Remarks/Real Life MC

Simple MC integration

Probability density:

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

is probability to find value x.



Simple MC integration

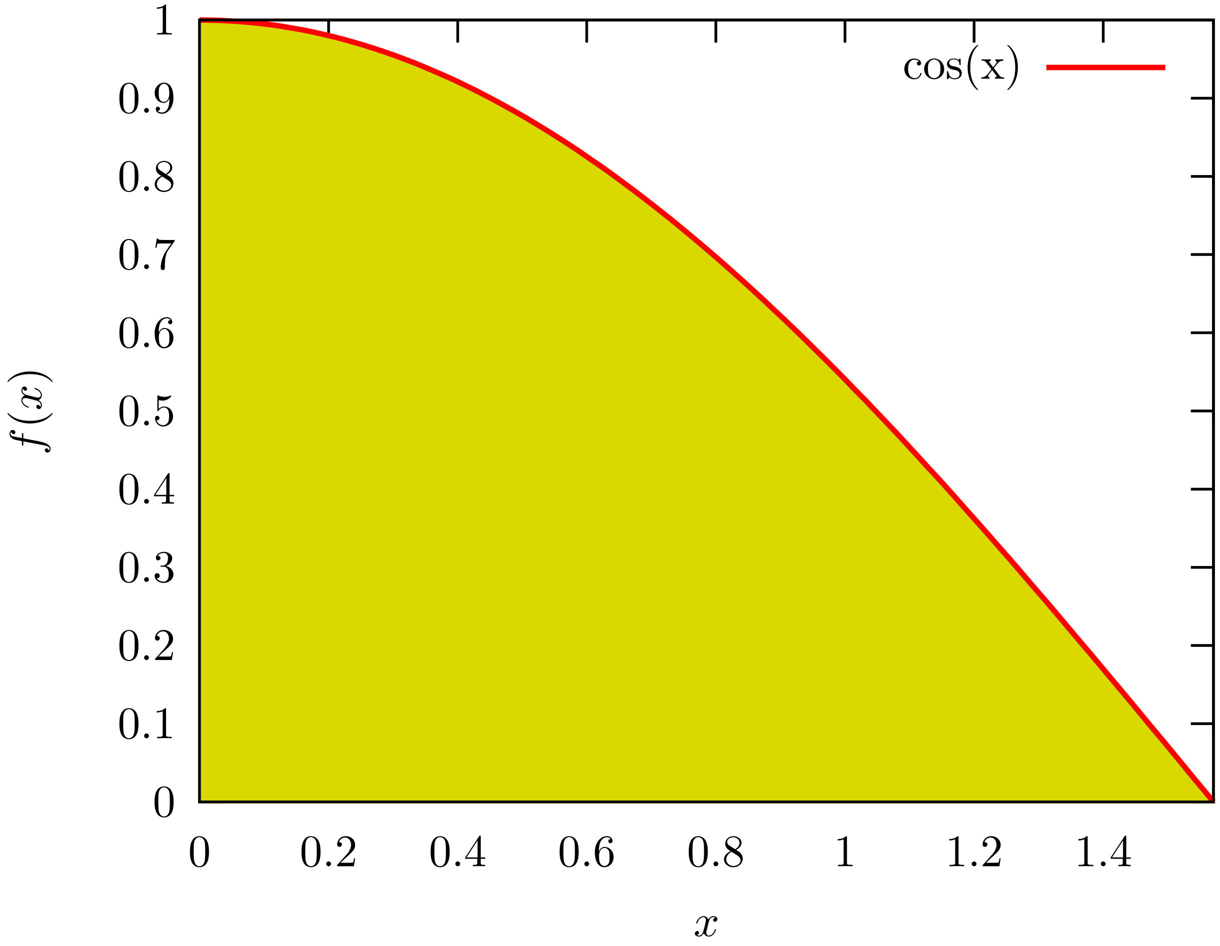
Probability density:

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

is probability to find value x.

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

is called *probability distribution*.



Simple MC integration

Probability density:

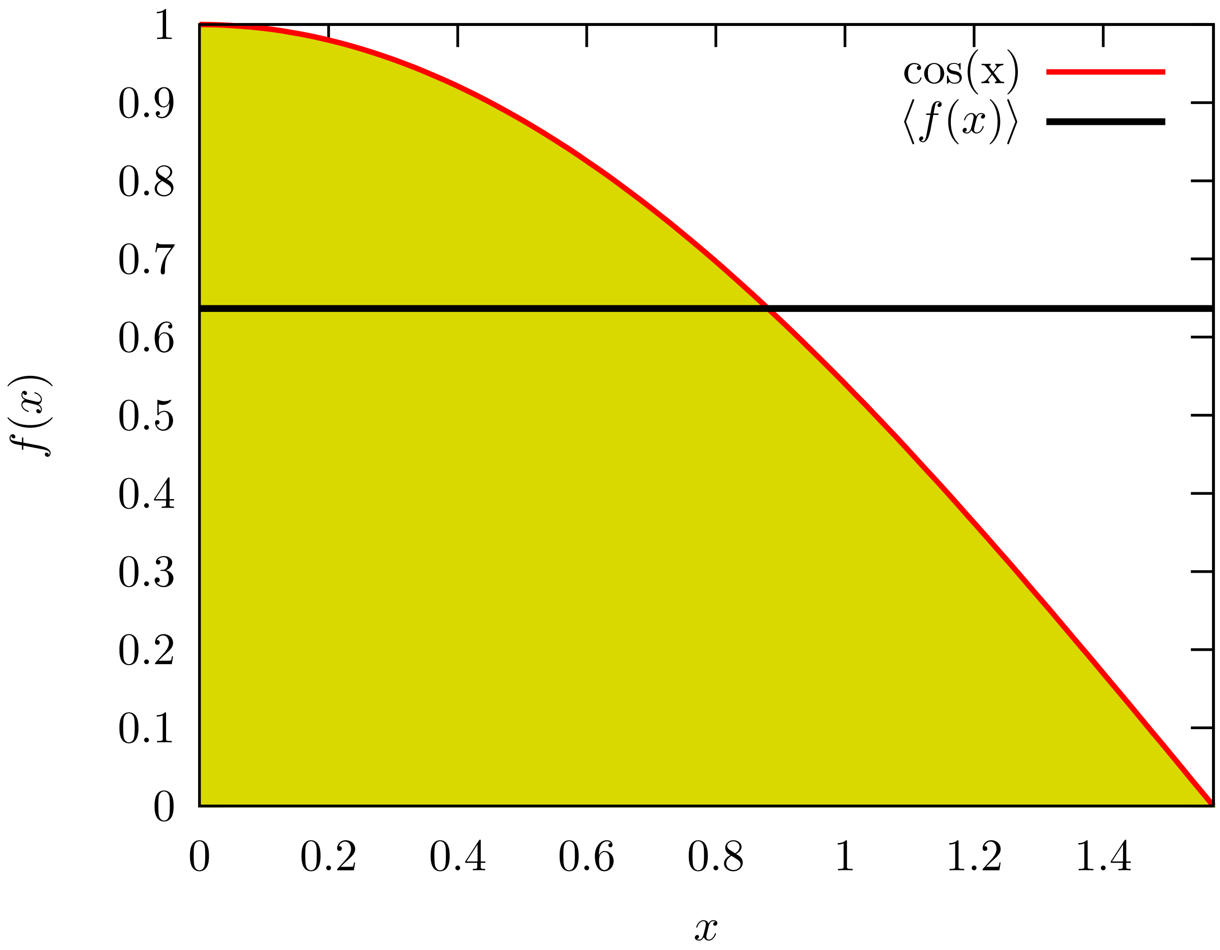
$$dP = f(x)\,dx$$

is probability to find value x.

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

is called *probability distribution*.
Mean value theorem of integration:

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



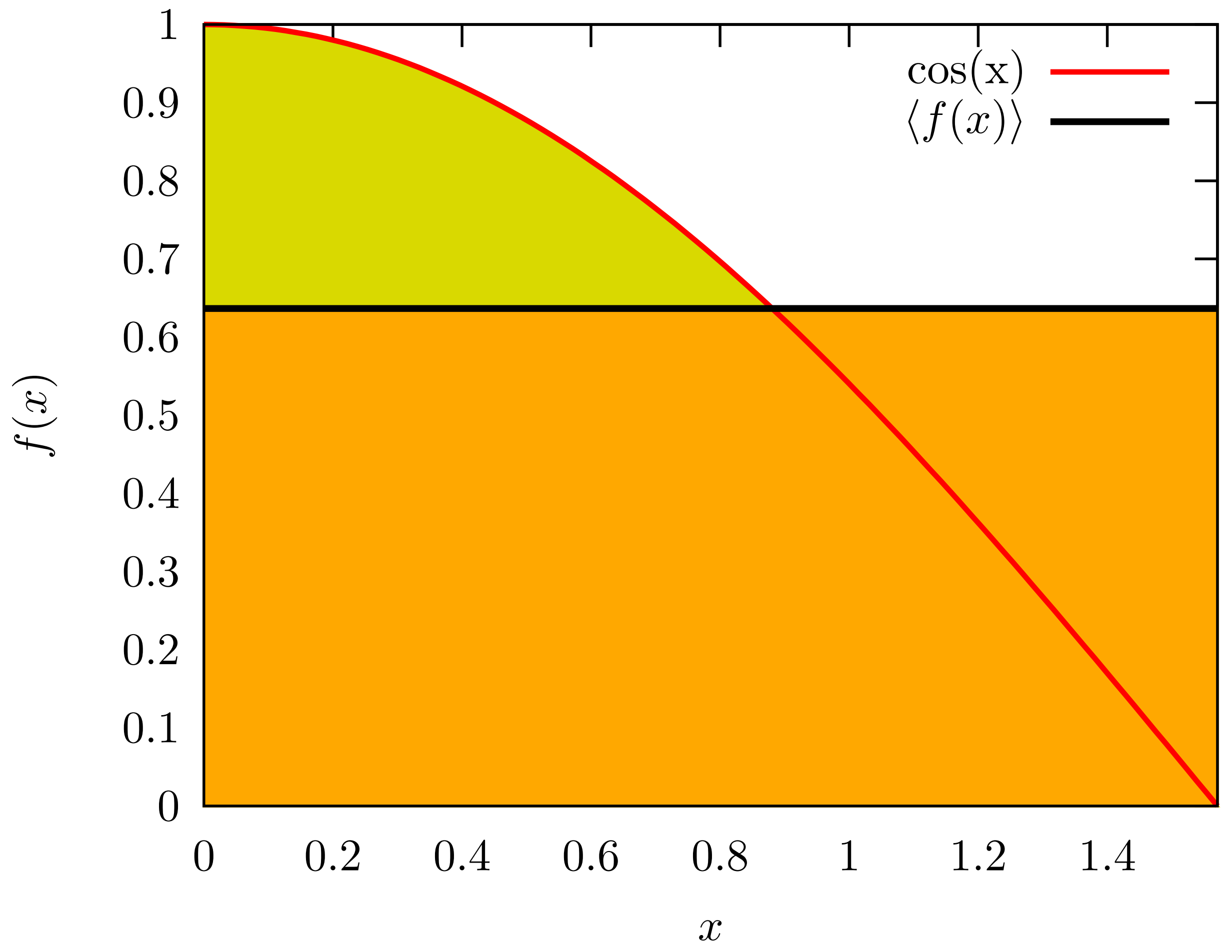
Simple MC integration

Rewrite average as

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

(Riemann integral).
Sum doesn't depend on ordering
$\longrightarrow$ randomize x_i.
In general: *Crude MC*

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



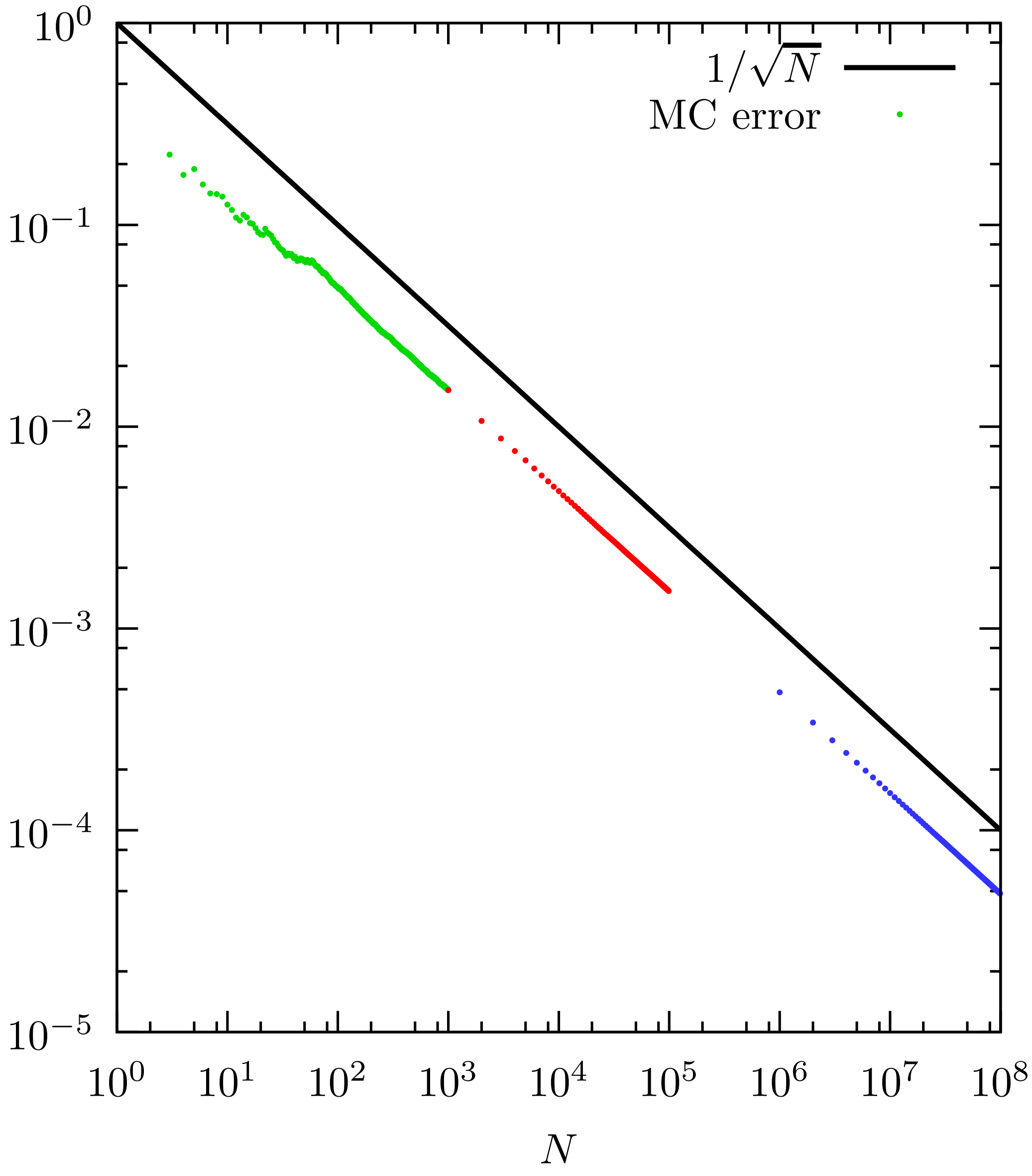
Simple MC integration

Simple check with $\cos(x), 0 \le x \le \pi/2$, compute σ_{MC} from

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

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

Looks like $1/\sqrt{N}$.



Simple MC integration

Look more closely, we can actually compute the error law ourselves:

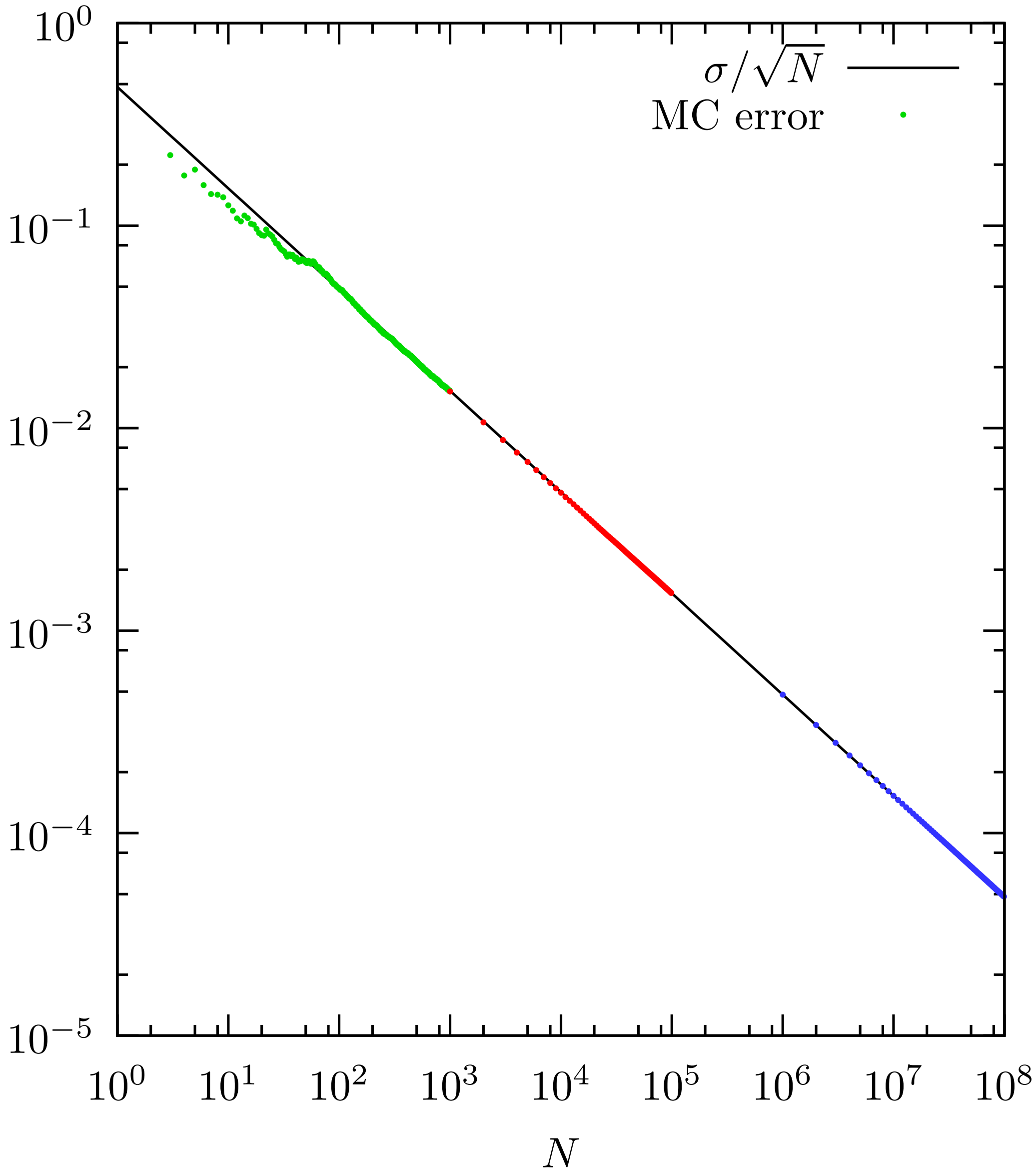
$$\langle f \rangle = \int_0^{\pi/2} \cos(x)\, dx = 1$$

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

then

$$\sigma = \sqrt{\frac{\pi^2}{8} - 1} \approx 0.4834.$$

Spot on.



Hit and Miss

Hit and miss method:

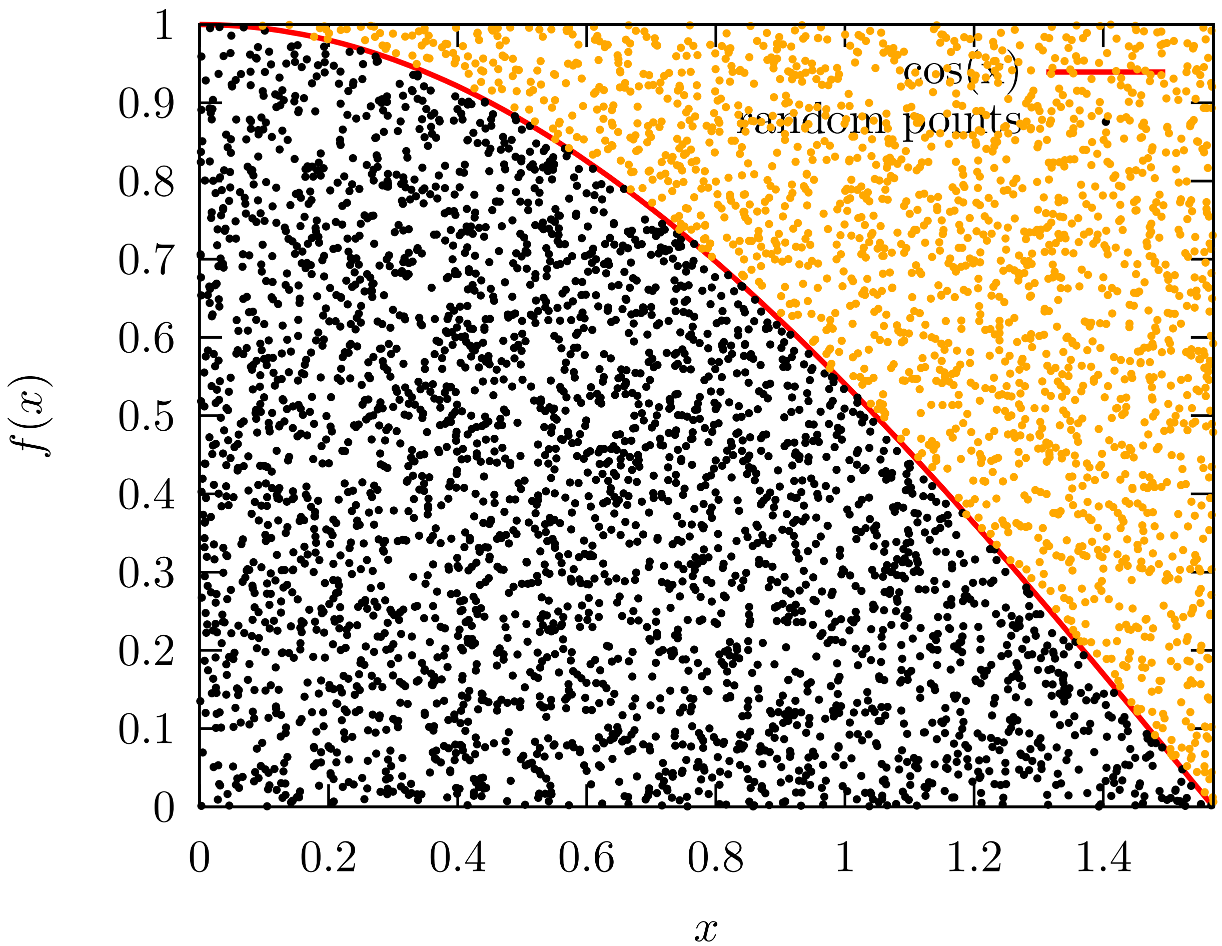
- throw N random points (x, y) into region.
- Count hits N_{hit}, i.e. whenever $y < f(x)$.

Then

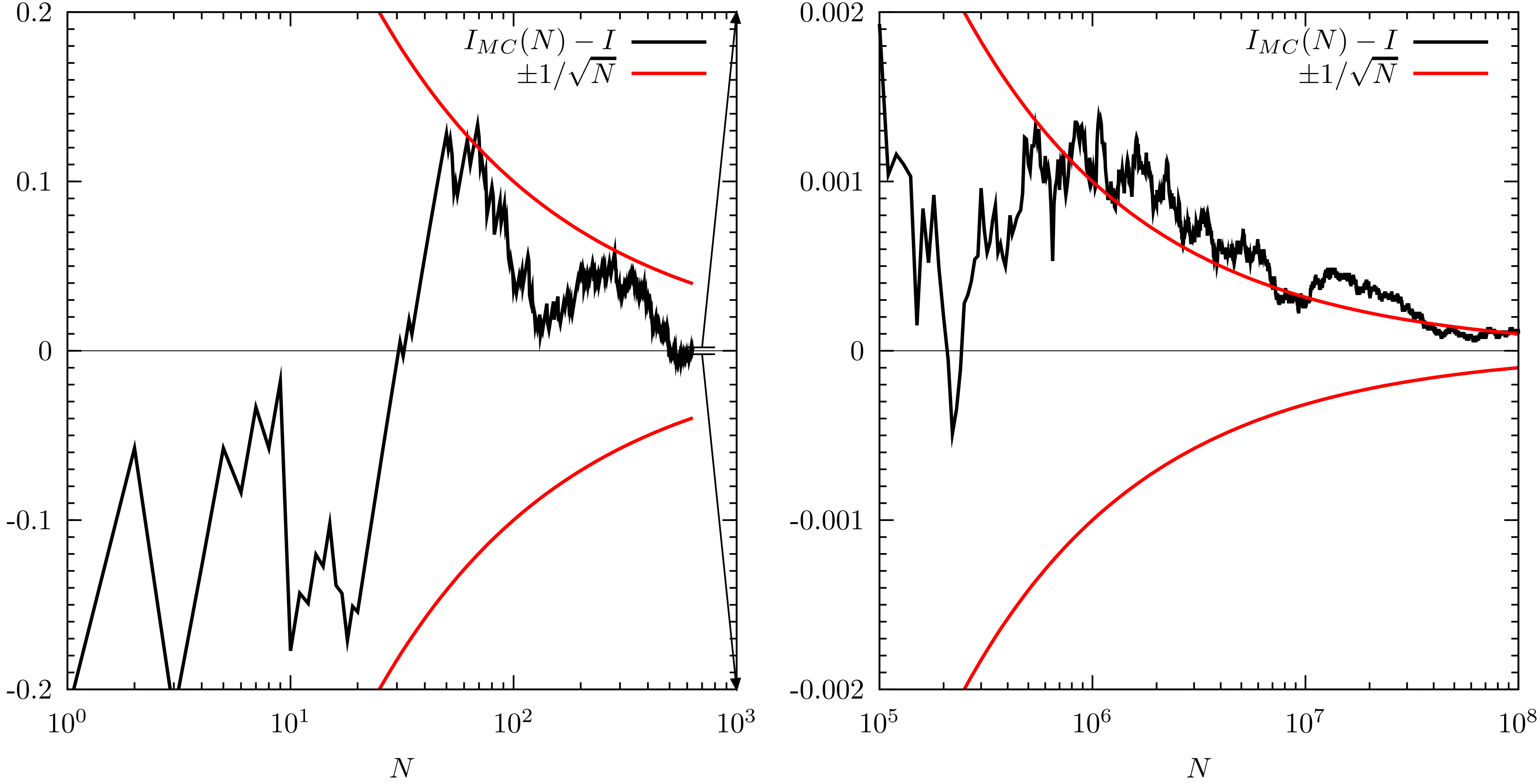
$$I \approx V \frac{N_{\mathrm{hit}}}{N}.$$

approaches 1 again in our example.

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



Apparently, error goes like $1/\sqrt{N}$ again.

Hit and Miss

This method is core of many event generators. However, it is not sufficient as such.

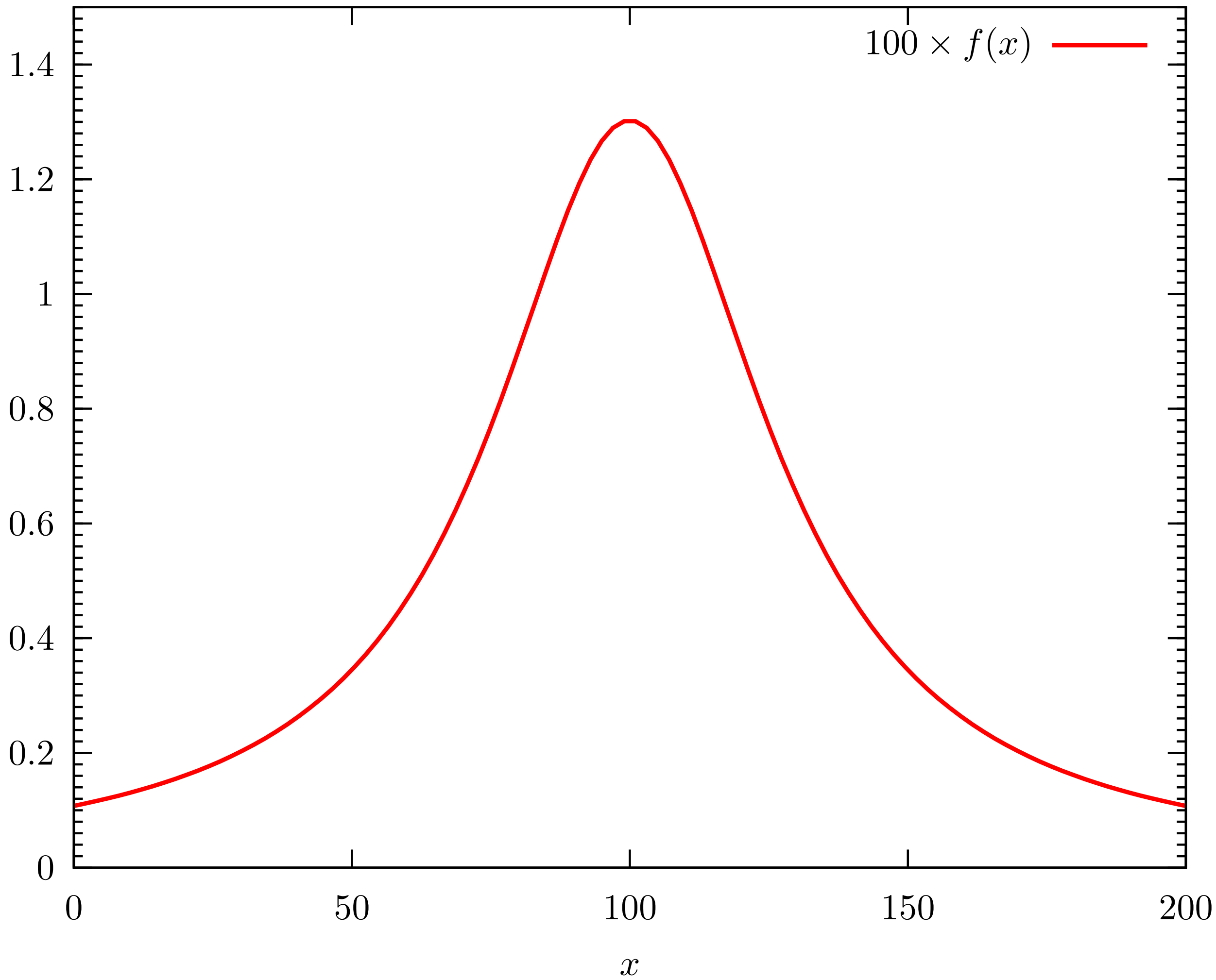
✓ Can handle any density $f(x)$, however wild and unknown it is.
✗ $f(x)$ should be bounded from above.
✗ Sampling will be very *inefficient* whenever $\mathrm{Var}\, f$ is large.

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

Inverting the integral

Another sampling method, that's needed for the following.

- Probability density $f(x)$. Not necessarily with given normalization.

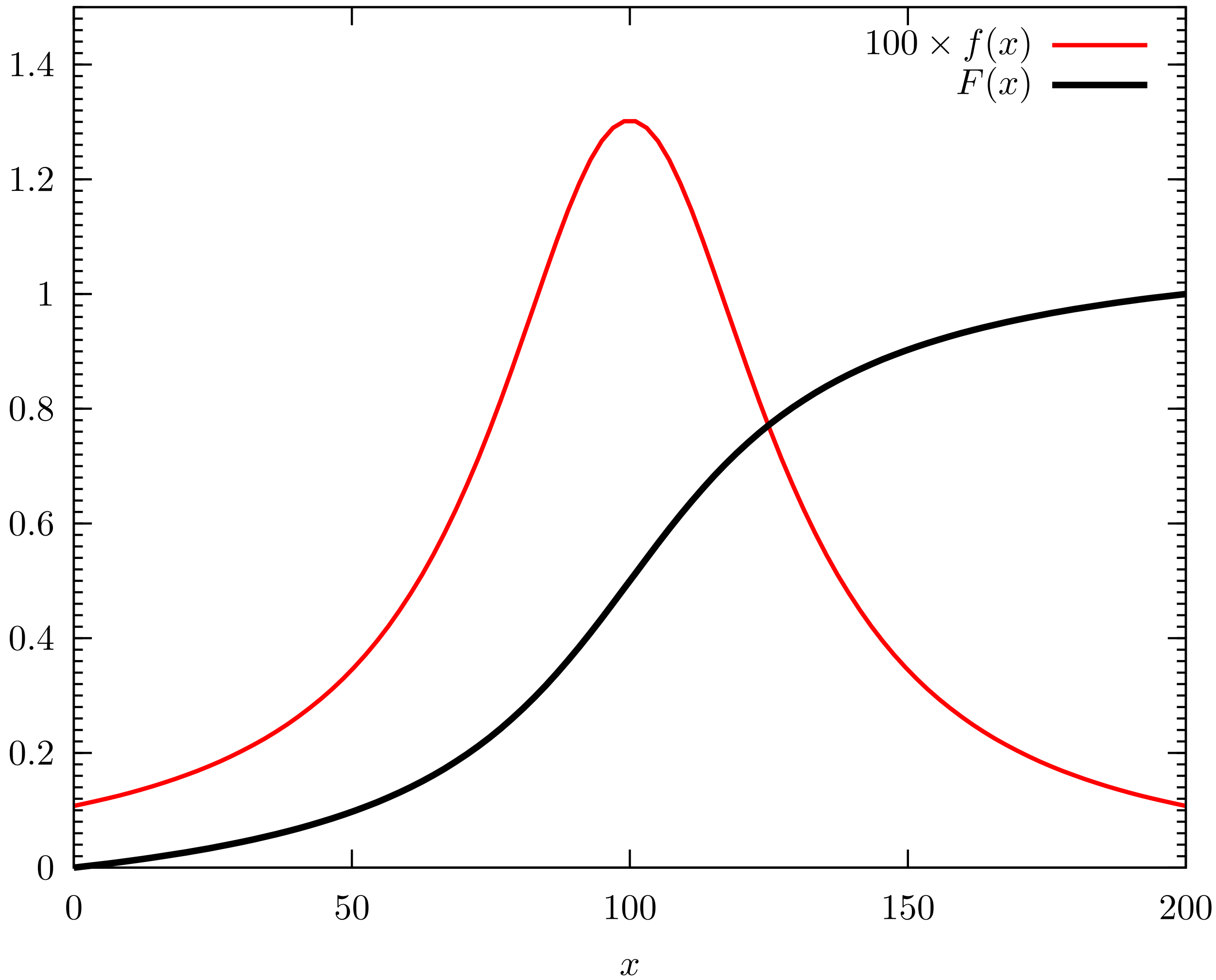


Inverting the integral

Another sampling method, that's needed for the following.

- Probability density $f(x)$. Not necessarily with given normalization.
- Integral known,

$$F(x) = \int_{-\infty}^{x} f(x)\, dx \ .$$



Inverting the integral

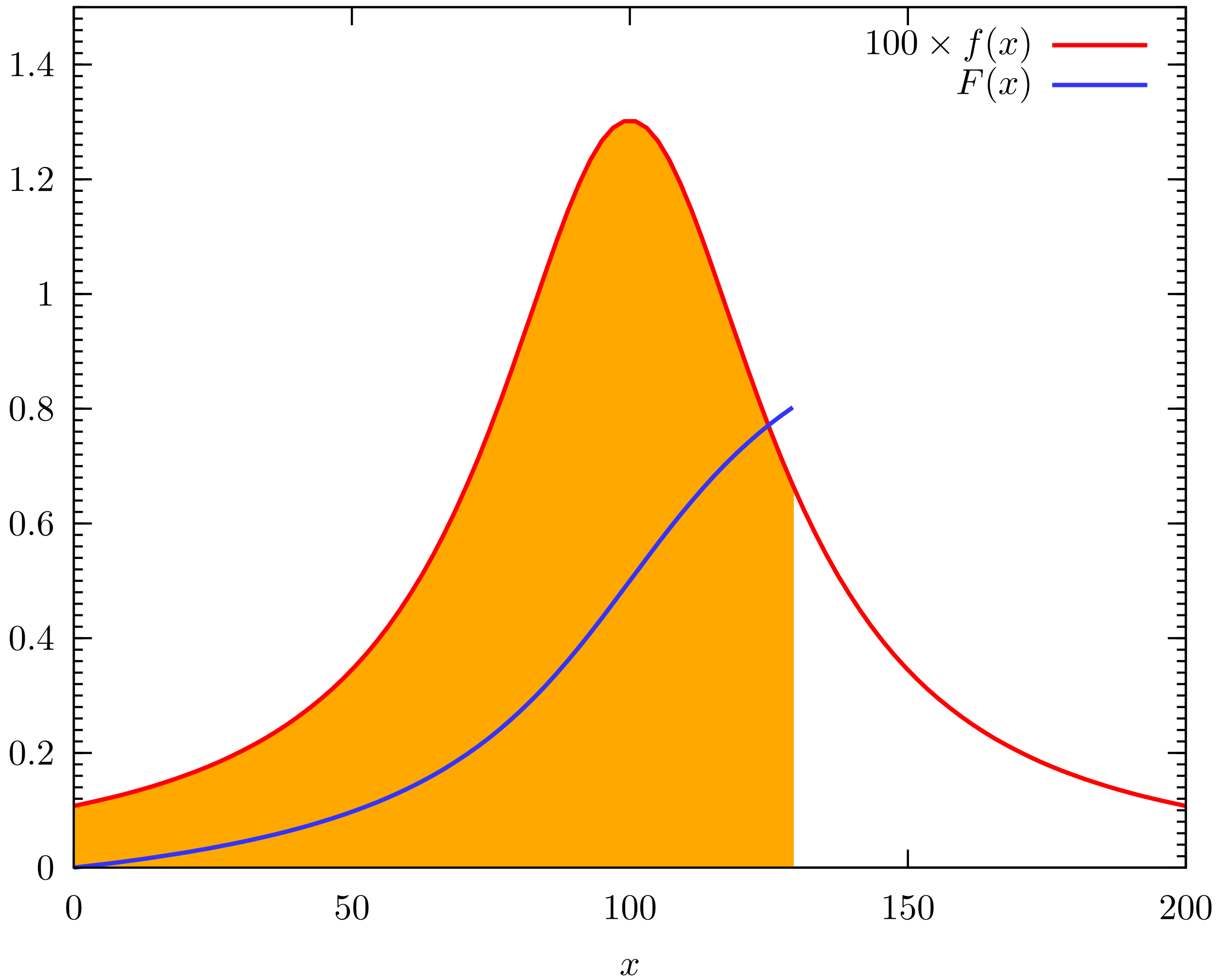
Another sampling method, that's needed for the following.

- Probability density $f(x)$. Not necessarily with given normalization.
- Integral known,

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \ .$$

- $P(x < x_s) = F(x_s)$.
- Probability = 'area', distributed evenly,

$$\int_0^r dP = r$$



Inverting the integral

Another sampling method, that's needed for the following.

- Probability density $f(x)$. Not necessarily with given normalization.
- Integral known,

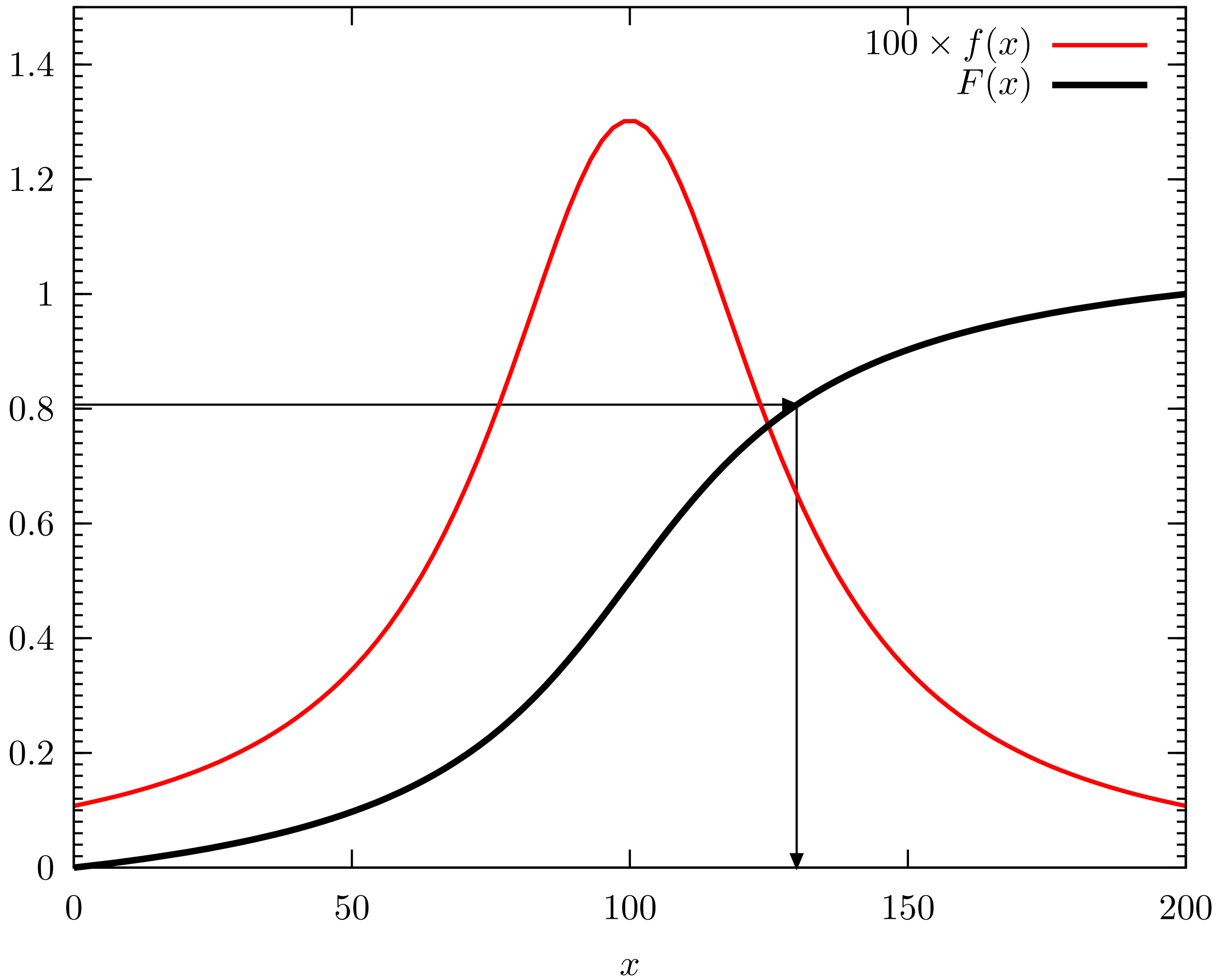
$$F(x) = \int_{-\infty}^{x} f(x)\, dx \ .$$

- $P(x < x_s) = F(x_s)$.
- Probability = 'area', distributed evenly,

$$\int_0^r dP = r$$

Sample x according to $f(x)$ with

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



Importance Sampling

Error on Crude MC $\sigma_{MC} = \sigma/\sqrt{N}$. $\Longrightarrow$ Reduce variance of integand.

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

where we have chosen $\int p\,dV = 1$ for convenience. Consider error term:

$$E = \left\langle \frac{f^2}{p^2} \right\rangle - \left\langle \frac{f}{p} \right\rangle^2 = \int \frac{f^2}{p^2}\,p dV - \left[\int \frac{f}{p}\,p dV\right]^2 = \int \frac{f^2}{p}\,dV - \left[\int f\,dV\right]^2 \ .$$

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

$$0 = \delta E = \delta \left(\int \frac{f^2}{p}\,dV - \left[\int f\,dV\right]^2 + \lambda \int p\,dV \right) = \int \left(-\frac{f^2}{p^2} + \lambda \right) dV\,\delta p \ ,$$

hence

$$p = \frac{|f|}{\sqrt{\lambda}} = \frac{|f|}{\int |f|\,dV} \ .$$

Choose p as close to f as possible.

Importance Sampling

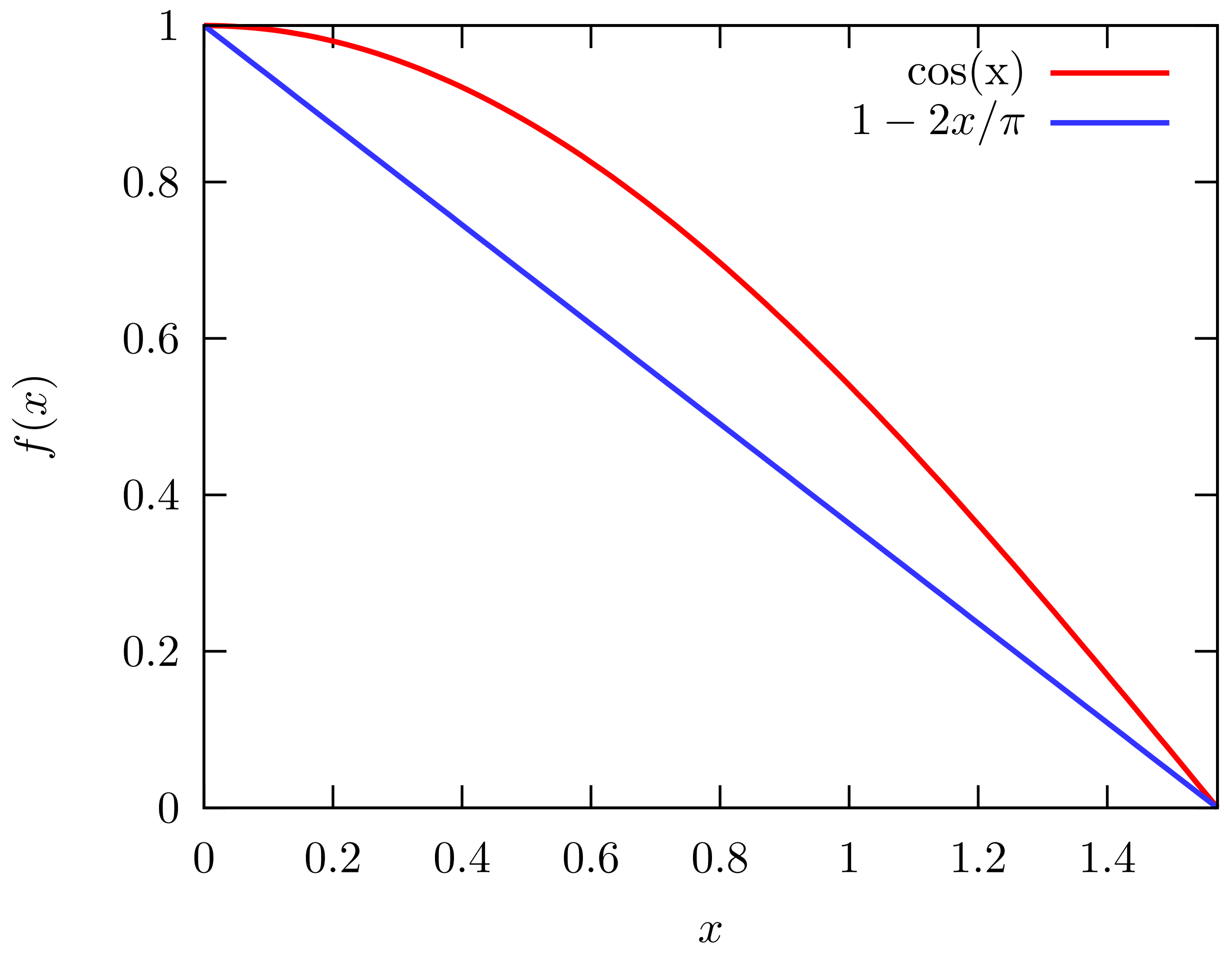
Improving $\cos(x)$ sampling,

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

Sample x with *inverting the integral* technique,

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

with flat random number ρ.



Importance Sampling

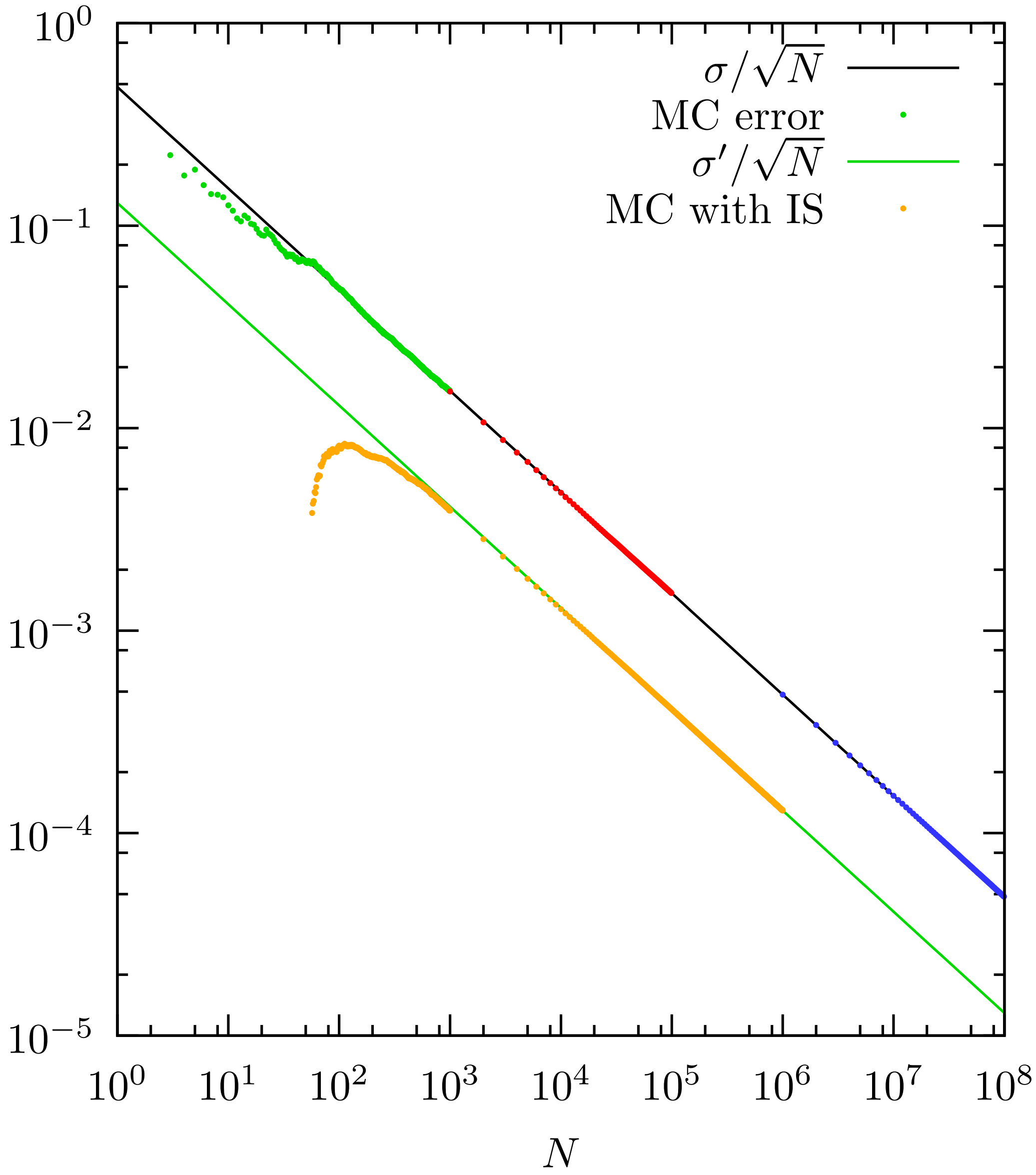
Improving $\cos(x)$ sampling,

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

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

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

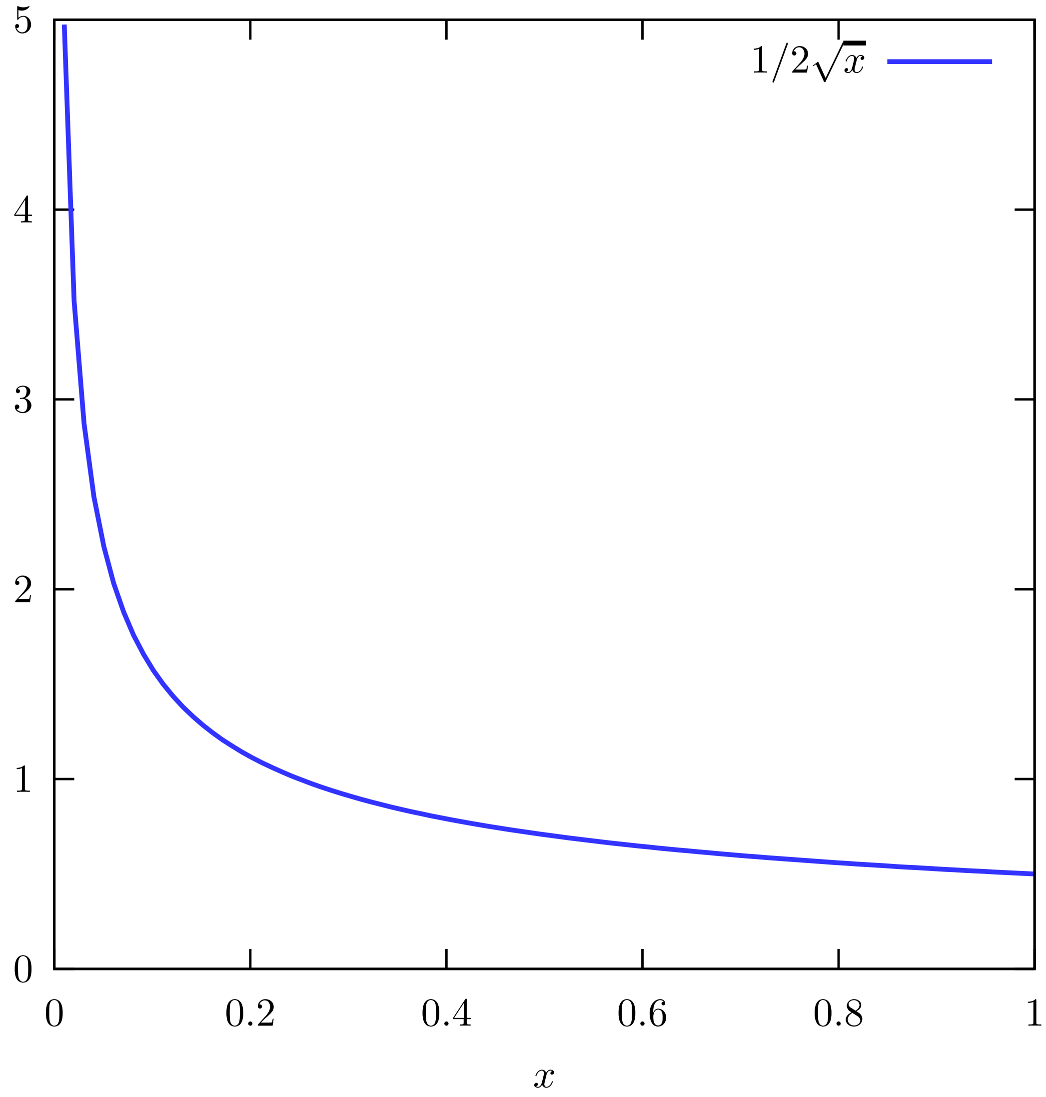
Much better convergence
and — important for us — about 80% "accepted events".
Reduced variance ⇒ better efficiency.



Importance Sampling

More interesting for divergent integrands, eg

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



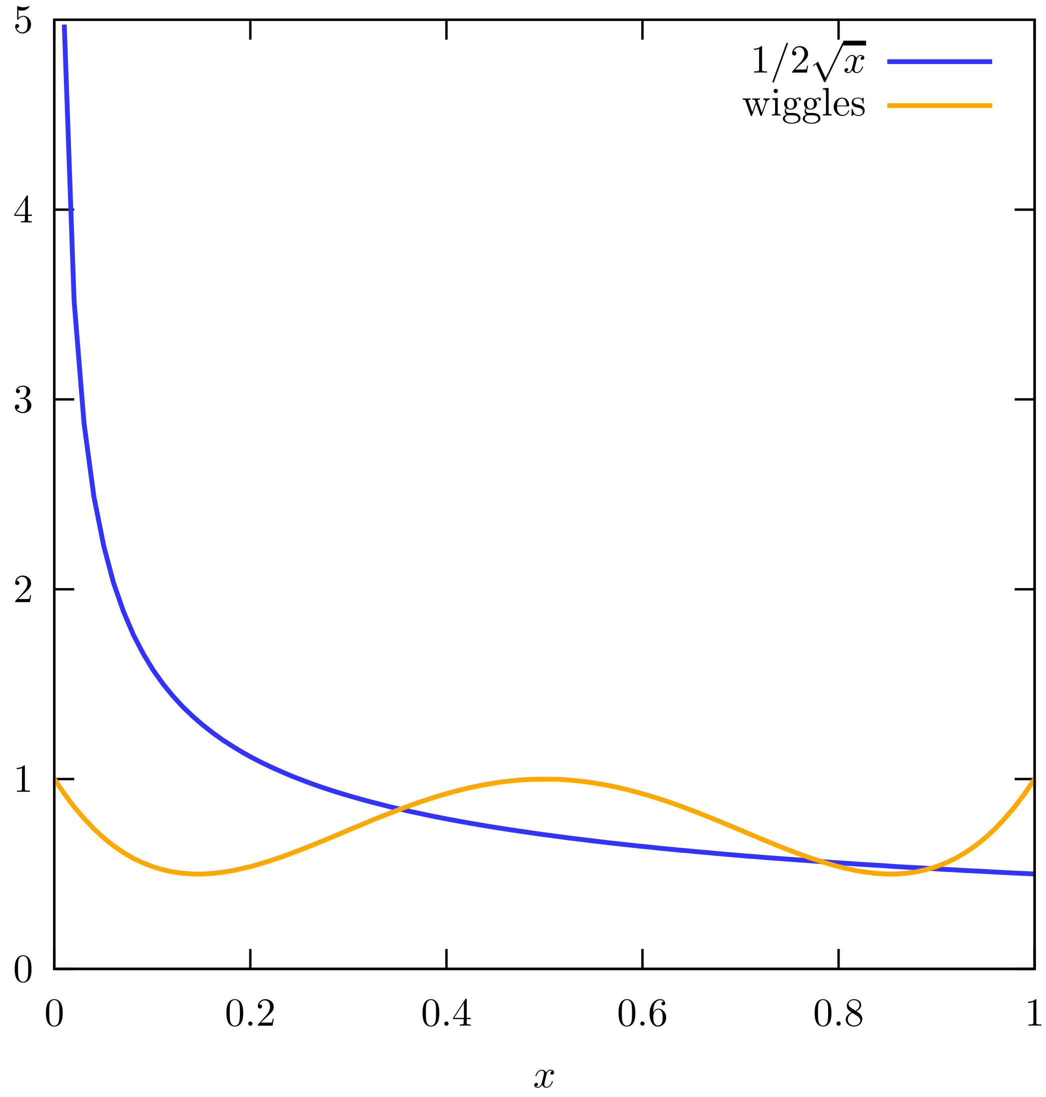
Importance Sampling

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

More interesting for divergent integrands, eg

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

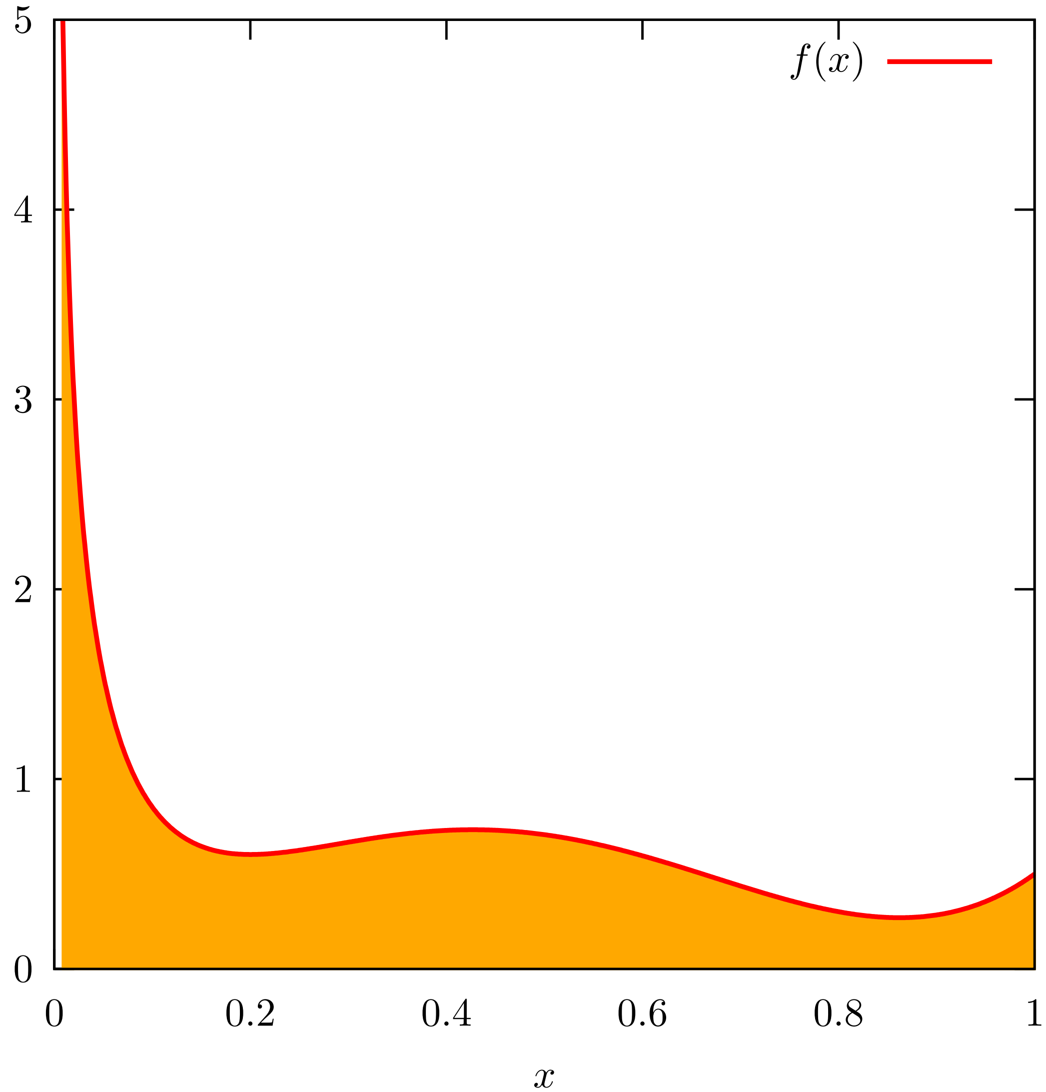
with some wiggles,

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

— we want to integrate

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

Note, that integral is finite ("cutoff" due to plot program $\longrightarrow$).

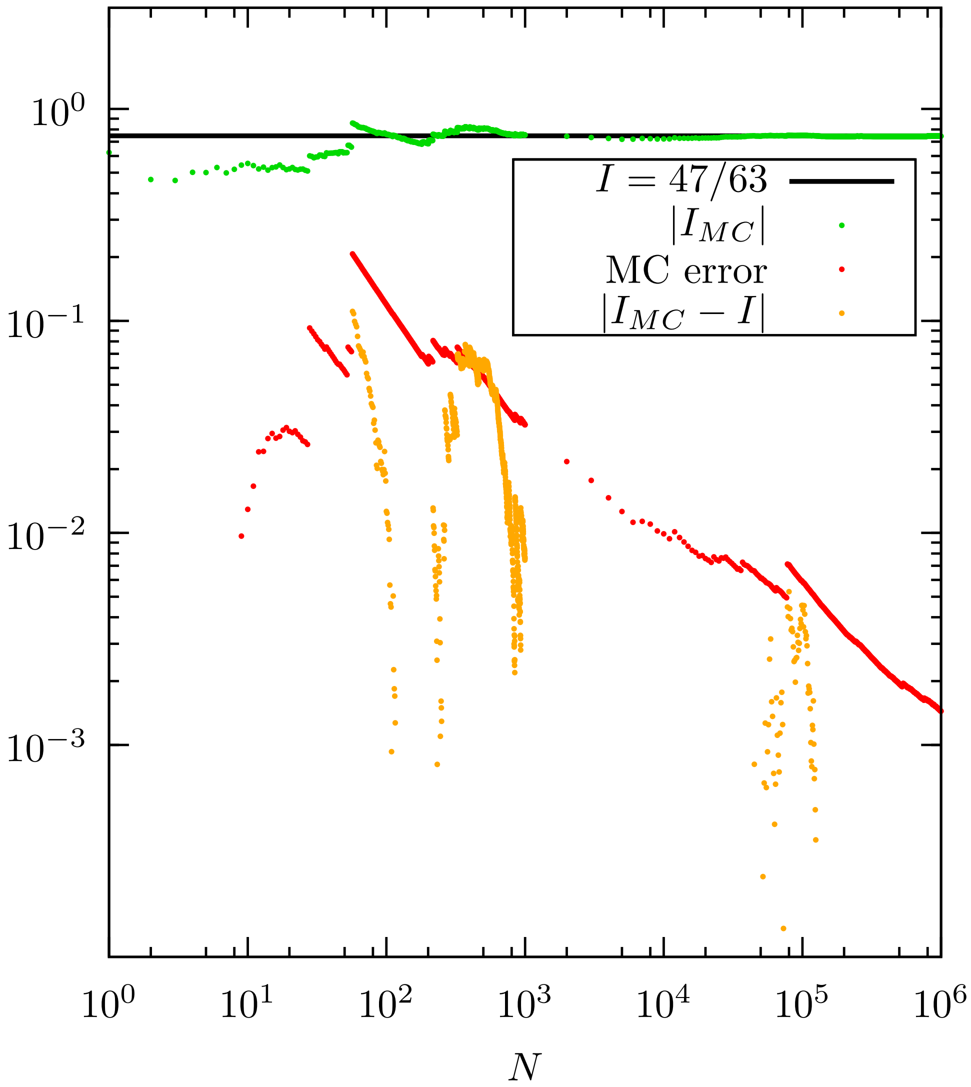


Importance Sampling

- 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.

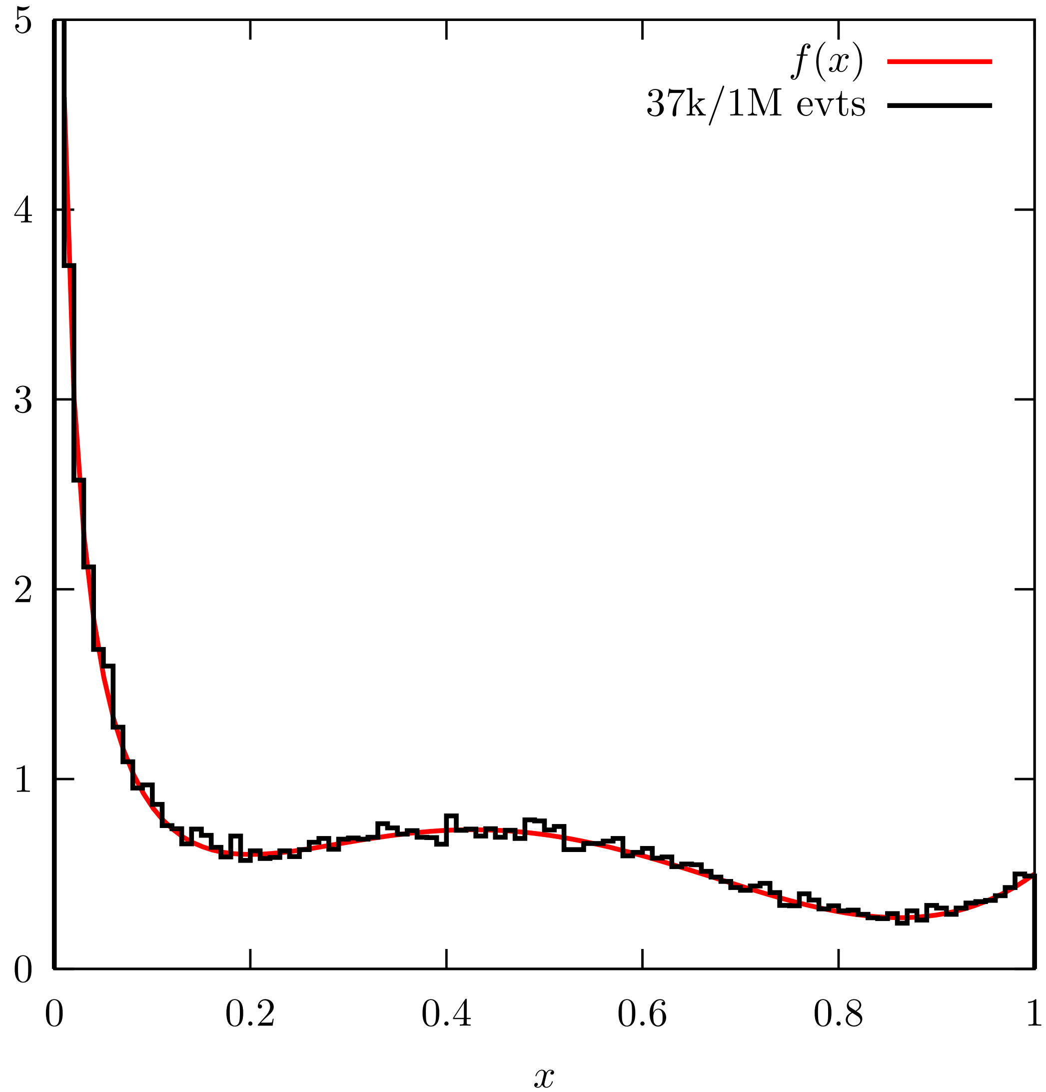
Use hit+mass variant here:

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



Importance Sampling

- 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.
- Efficiency for MC events only 3.7%.
- Note the wiggly histogram.



Importance Sampling

Now importance sampling,

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

so,

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

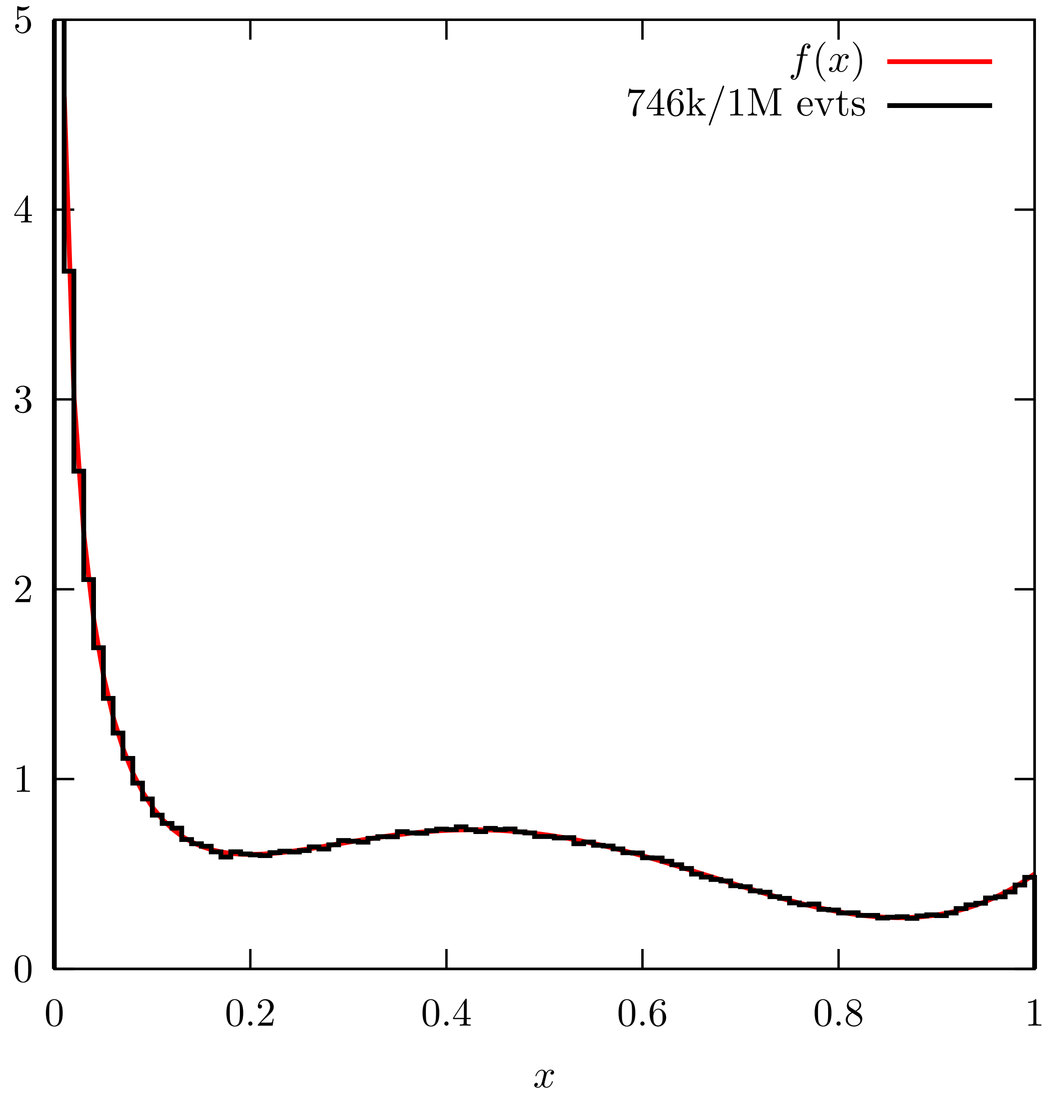
x sampled with *inverting the integral* from flat random numbers ρ, $x = \rho^2$.

Importance Sampling

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

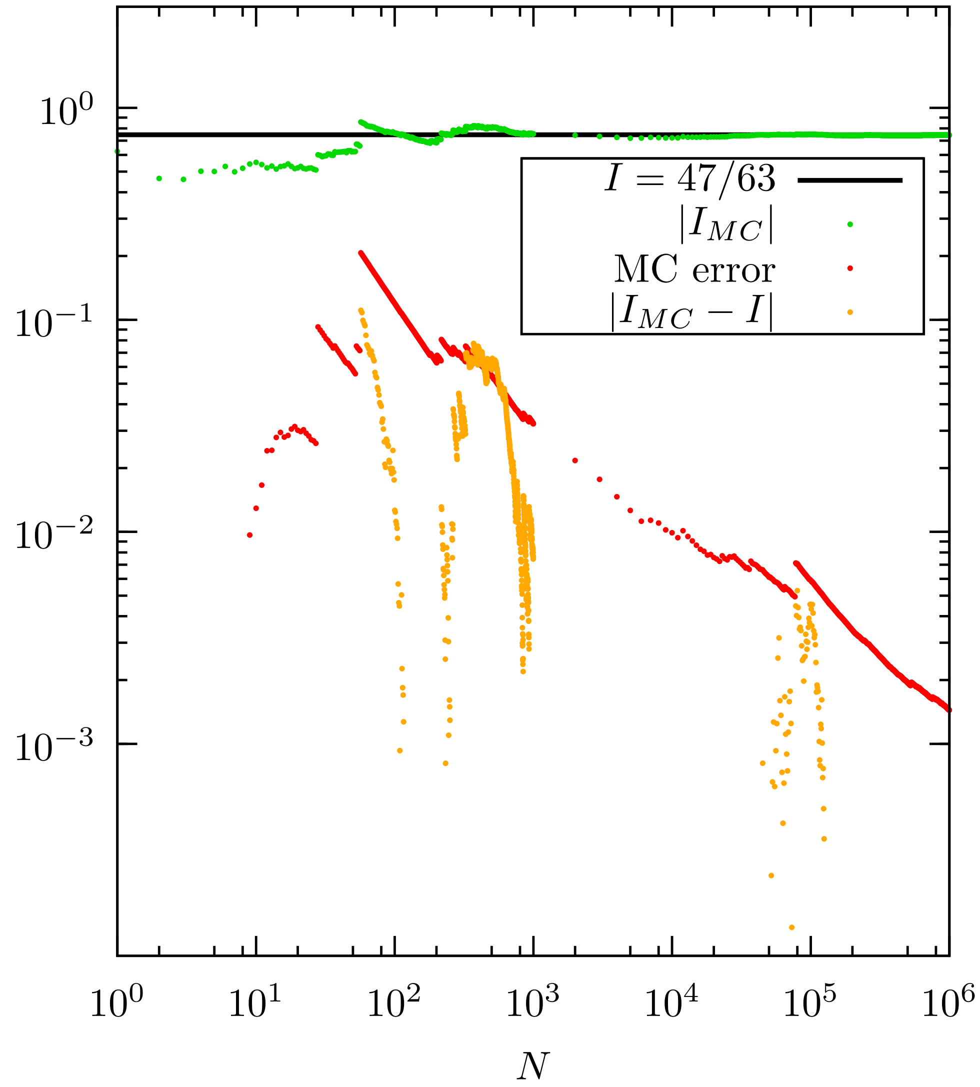
$$\rho = \sqrt{x}, \qquad d\rho = \frac{dx}{2\sqrt{x}}$$

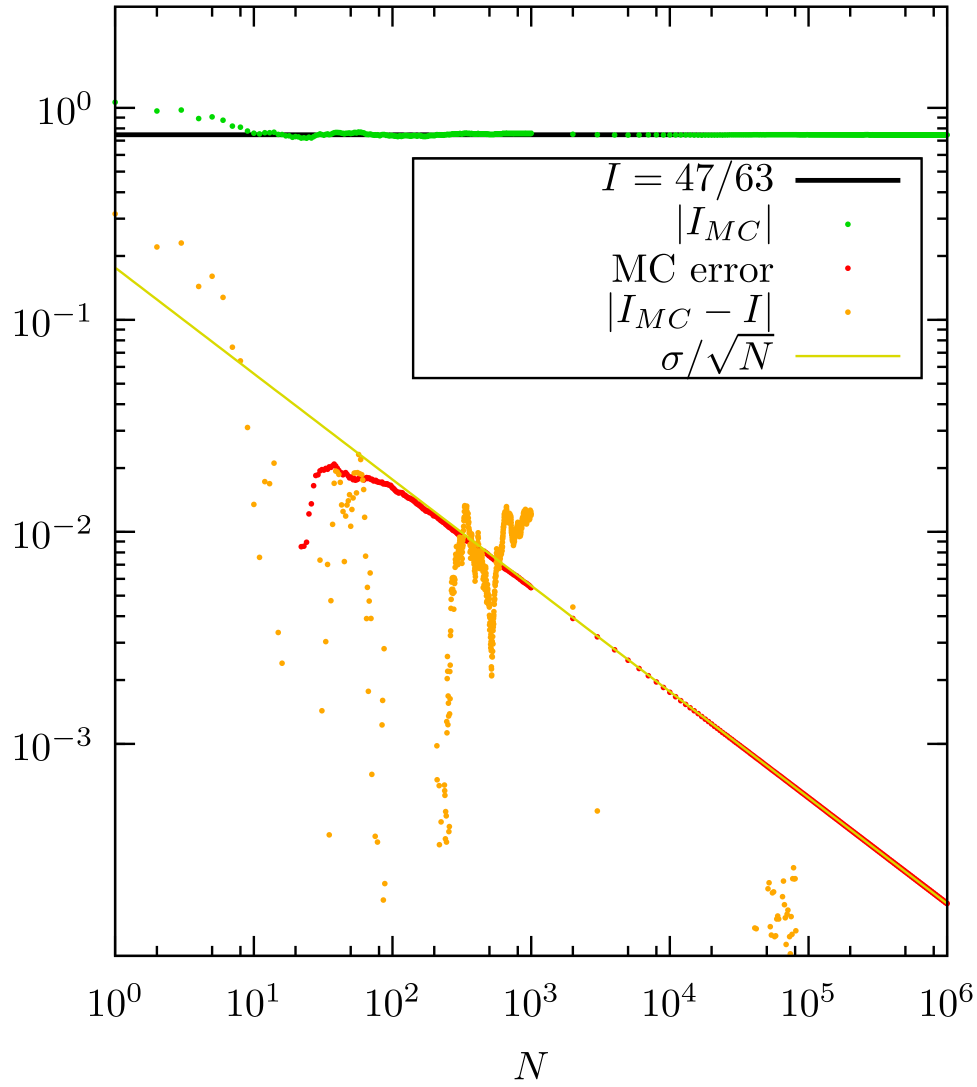
Events generated with $w_{\max} = 1$, as $p(x) \le 1$, no guesswork needed here! Now, we get 74.6% MC efficiency.



Importance Sampling

Crude MC vs Importance sampling. $100\times$ more events needed to reach same accuracy.





Useful example

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

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

$$= \frac{1}{m\Gamma} \int_{y_0}^{y_1} \frac{dy}{y^2+1} \qquad (y = \frac{s-m^2}{m\Gamma})$$

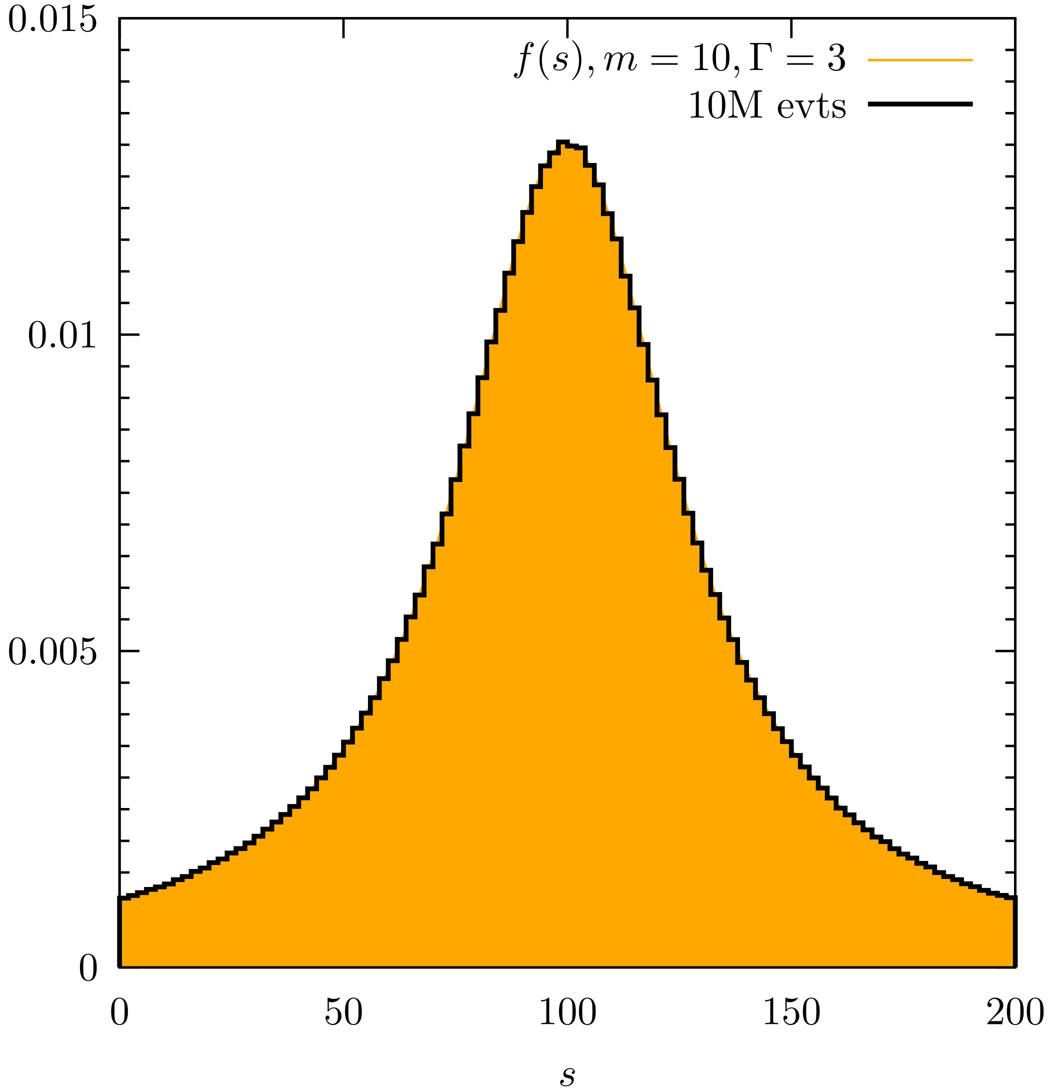
$$= \frac{1}{m\Gamma} \arctan \frac{s-m^2}{m\Gamma} \Big|_{s_0}^{s_1}$$

Inverting the integral gives (with convenient normalization ($\neq 1$))

$$f(s) = \frac{m\Gamma}{(s-m^2)^2 + m^2\Gamma^2} ,$$

$$F(s) = \arctan \frac{s-m^2}{m\Gamma} = \rho ,$$

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



Multichannel MC

$f(s)$ has multiple peaks ($\times$ wiggles from ME). Peak structure known quite well (BW peaks or just poles from Feynman diagrams). Encode this in sum of sample functions $g_i(s)$ with arbitrary (to begin with) weights α_i, $\sum_i \alpha_i = 1$.

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

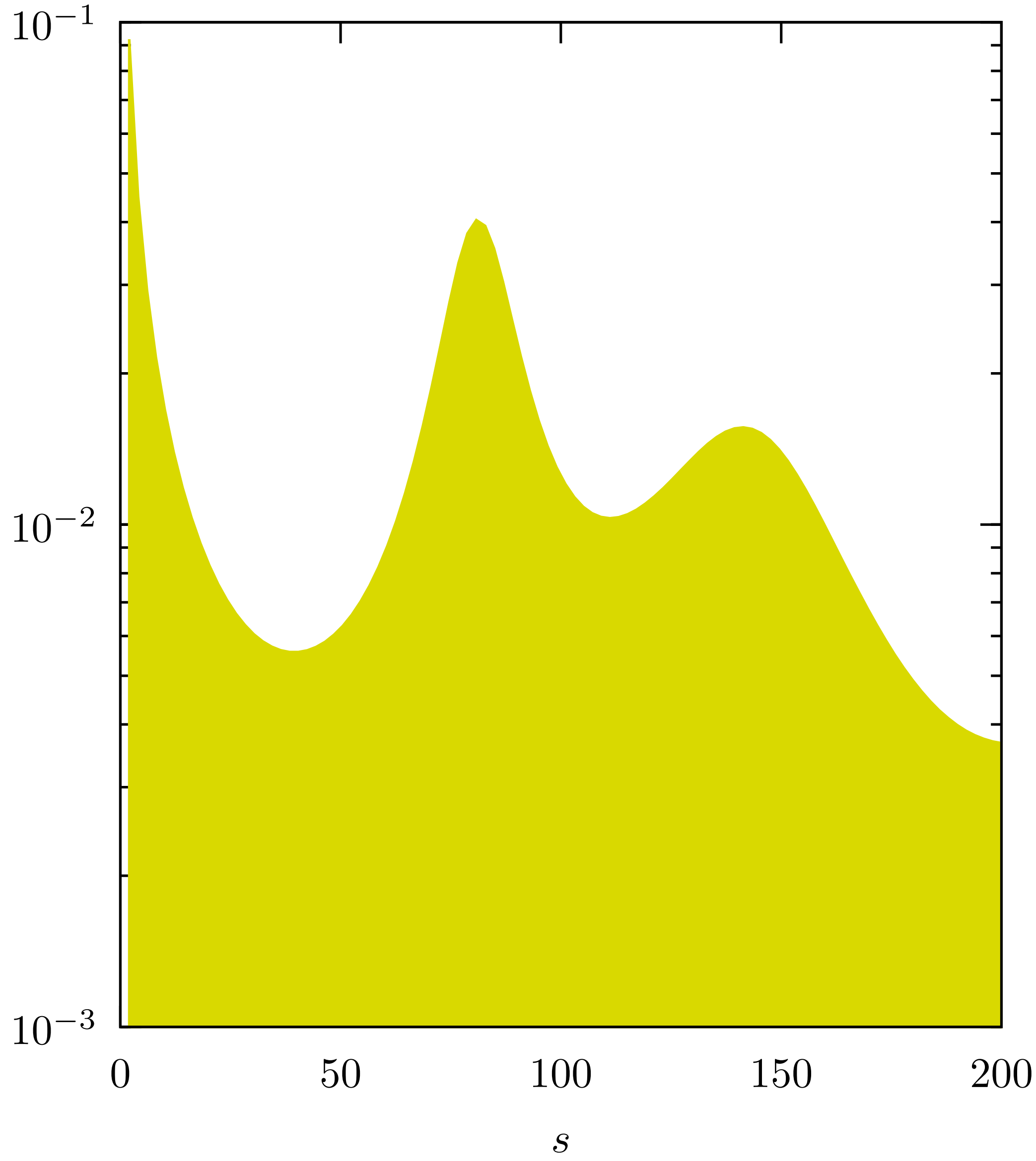
Then

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

Now $g_i(s)\, ds = d\rho_i$ (inverting the integral). Select the distribution (i) you'd like to sample next event from acc to weights α_i. α_i can be optimized after a number of trial events (increase α_i for channels with large variance).

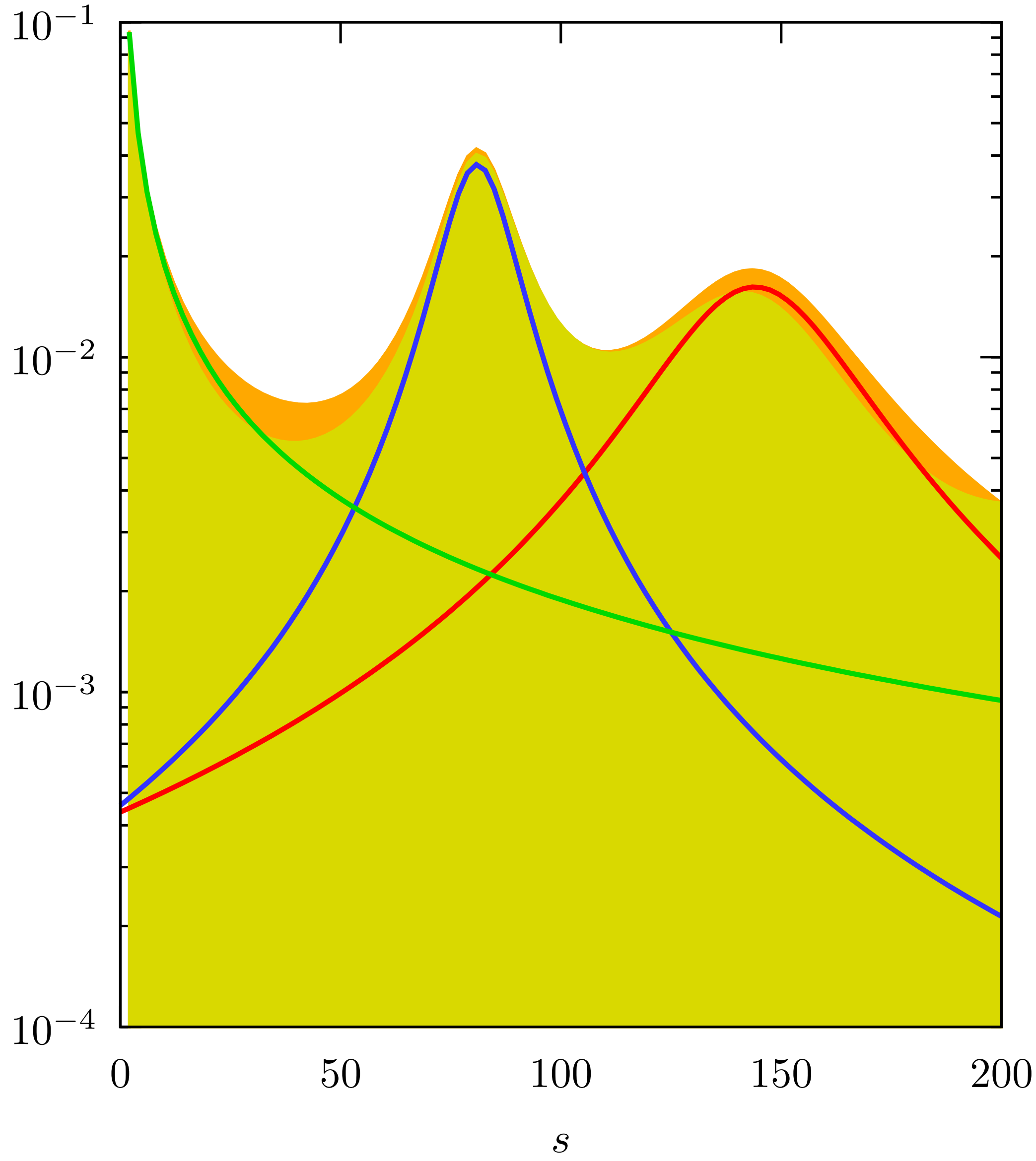
Multichannel MC

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



Multichannel MC

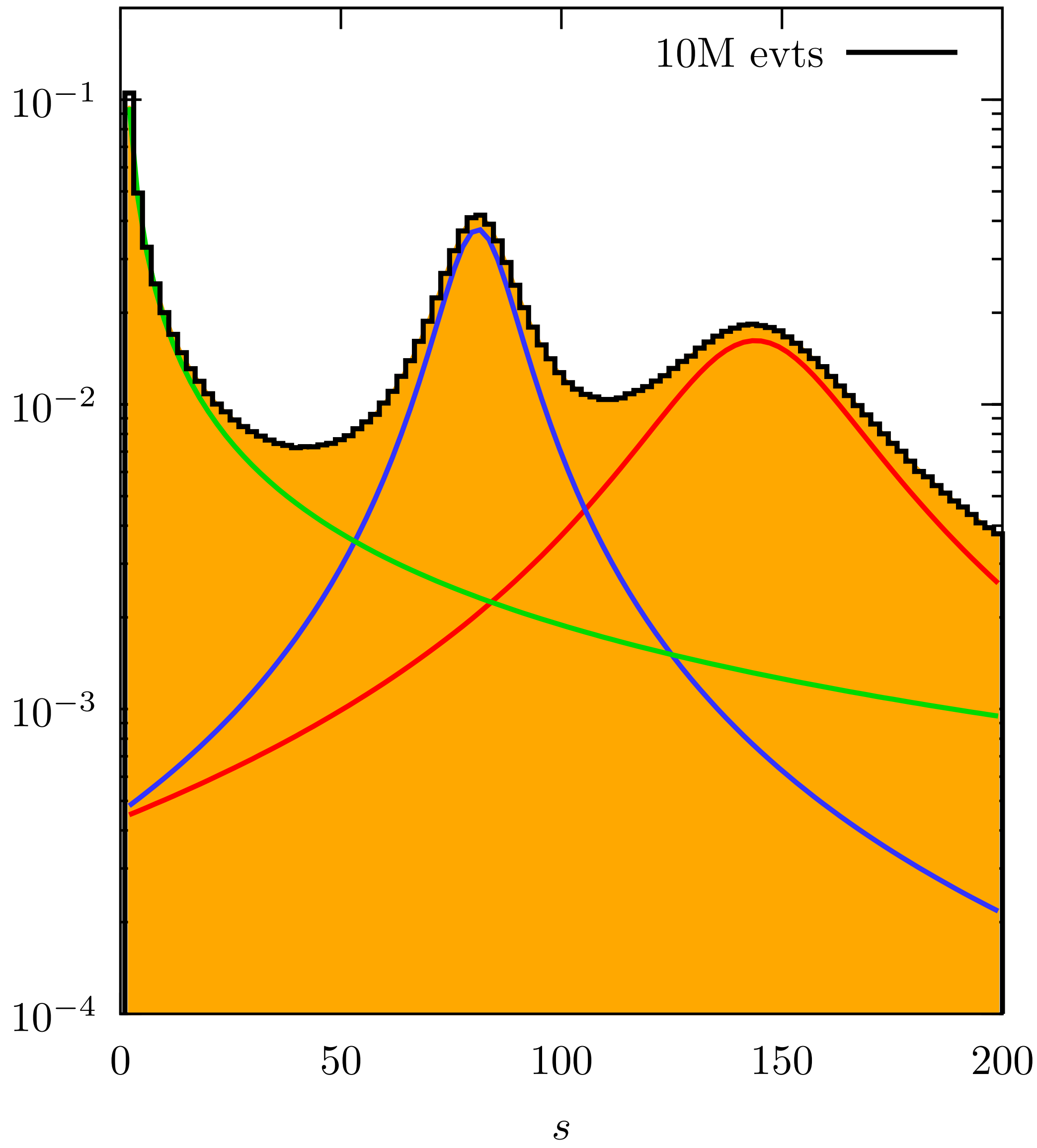
- $f(s)$ has multiple peaks ($\times$ wiggles from ME).
- Peak structure known quite well (BW peaks or just poles from Feynman diagrams).



Multichannel MC

- $f(s)$ has multiple peaks ($\times$ wiggles from ME).
- Peak structure known quite well (BW peaks or just poles from Feynman diagrams).
- Encode this in sum of sample functions $g_i(s)$ with arbitrary (to begin with) weights α_i, $\sum_i \alpha_i = 1$.

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



Multichannel MC

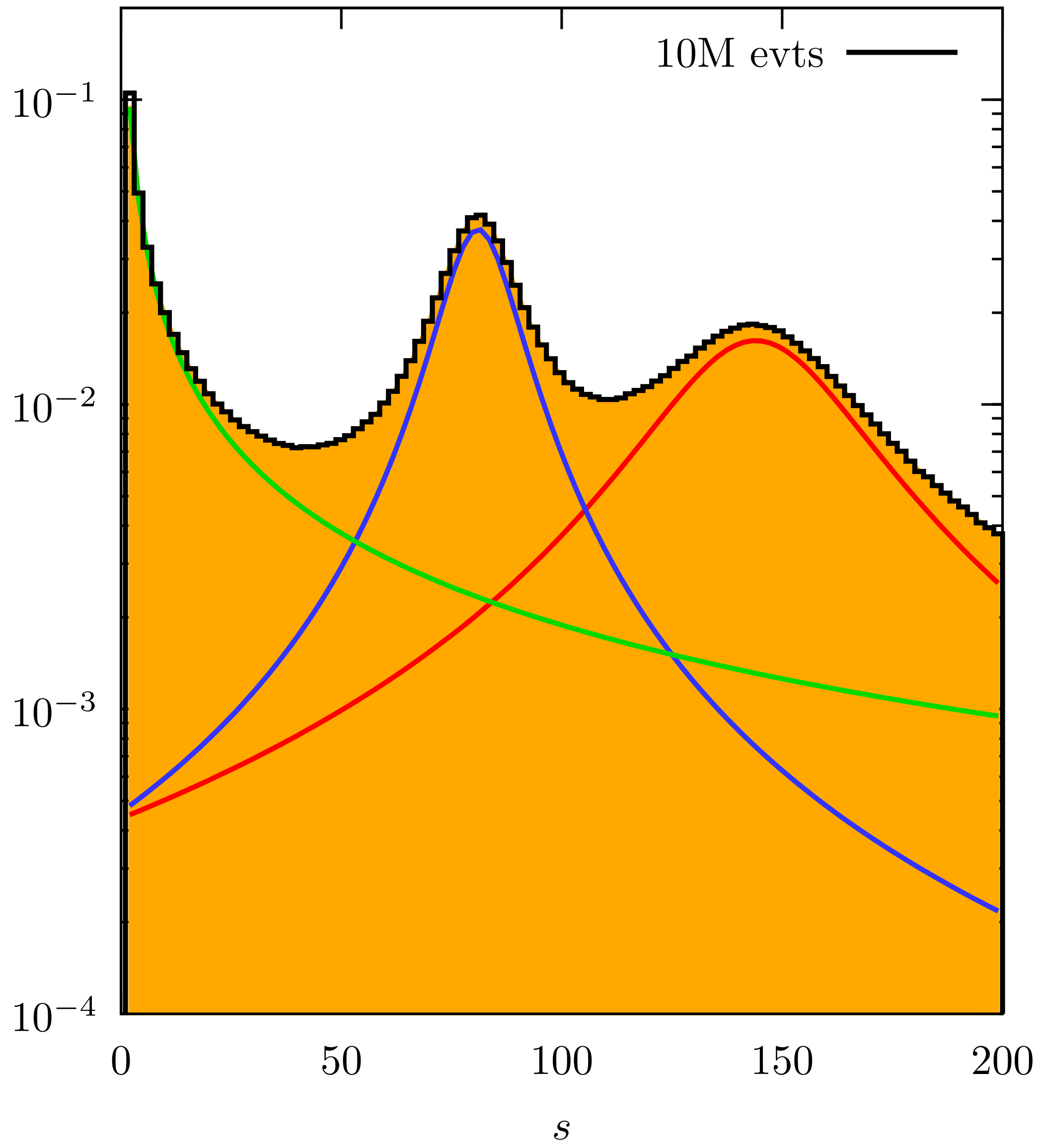
- Encode this in sum of sample functions $g_i(s)$ with arbitrary (to begin with) weights α_i, $\sum_i \alpha_i = 1$.

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

Then

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

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



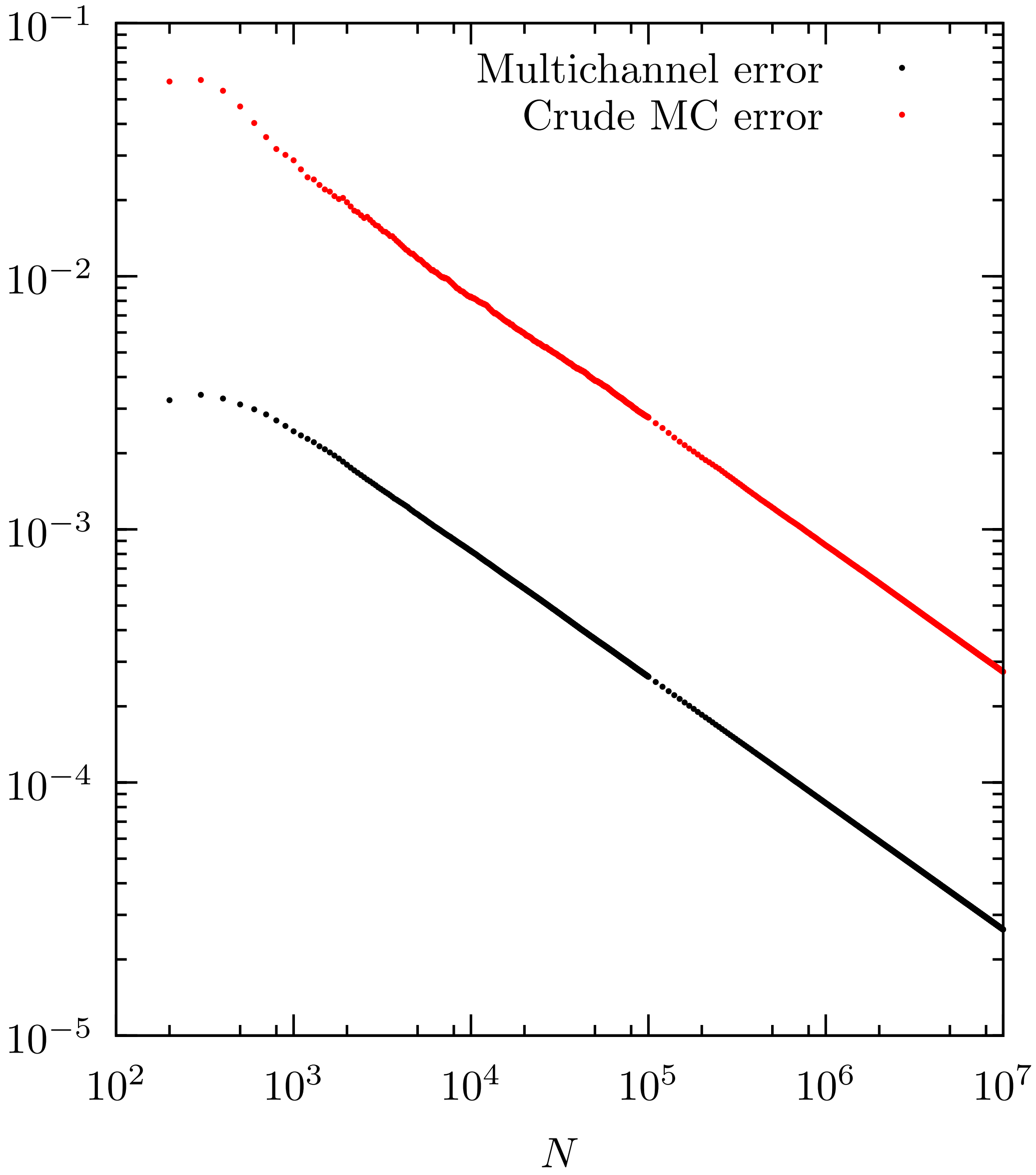
Multichannel MC

- Encode this in sum of sample functions $g_i(s)$ with arbitrary (to begin with) weights α_i, $\sum_i \alpha_i = 1$.

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

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

- Now $g_i(s)\, ds = d\rho_i$ (inverting the integral).
- Select the distribution (i) you'd like to sample next event from acc to weights α_i.
- α_i can be optimized after a number of trial events.



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 $\mathrm{MCError} \sim \frac{1}{\sqrt{N}}$,
  compare eg Trapezium rule $\mathrm{Error} \sim \frac{1}{N^{2/D}}$.
- Many important techniques covered here in detail! Should be good starting point.