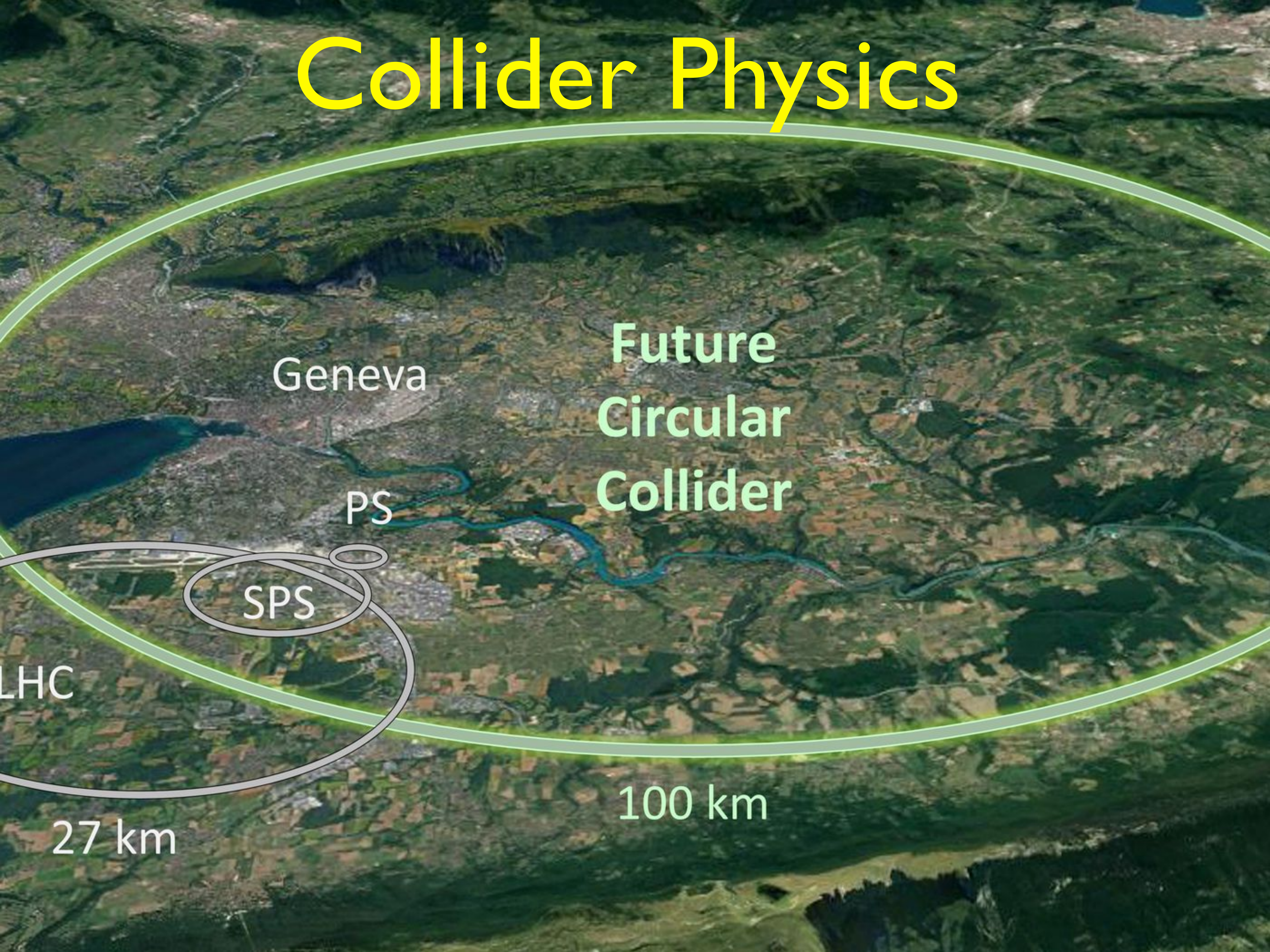
Collider Physics



Monte Carlo Method

Motivation

- Evaluation of cross sections leads to

$$\sigma = \int dx_1 dx_2 \sum_{\text{subp}} f_{a_1/p}(x_1)\, f_{a_2/\bar{p}}(x_2)$$

$$\frac{1}{2\hat{s}(2\pi)^{3n-4}} \int d\Phi_n(x_1 P_A + x_2 P_B;\ p_1 \ldots p_n)\Theta(\text{cuts})\overline{\sum}|\mathcal{M}|^2(a_1 a_2 \to b_1 \ldots b_n)$$

there are 3n-2 integrals. We also need to simulate the detector!

Motivation

- Evaluation of cross sections leads to

$$\sigma = \int dx_1 dx_2 \sum_{\mathrm{subp}} f_{a_1/p}(x_1)\, f_{a_2/\bar{p}}(x_2)$$

$$\frac{1}{2\hat{s}(2\pi)^{3n-4}} \int d\Phi_n(x_1 P_A + x_2 P_B;\ p_1 \dots p_n)\Theta(\mathrm{cuts})\overline{\sum}|\mathcal{M}|^2(a_1 a_2 \to b_1 \dots b_n)$$

there are 3n-2 integrals. We also need to simulate the detector!

- We need effective techniques to perform the calculations!

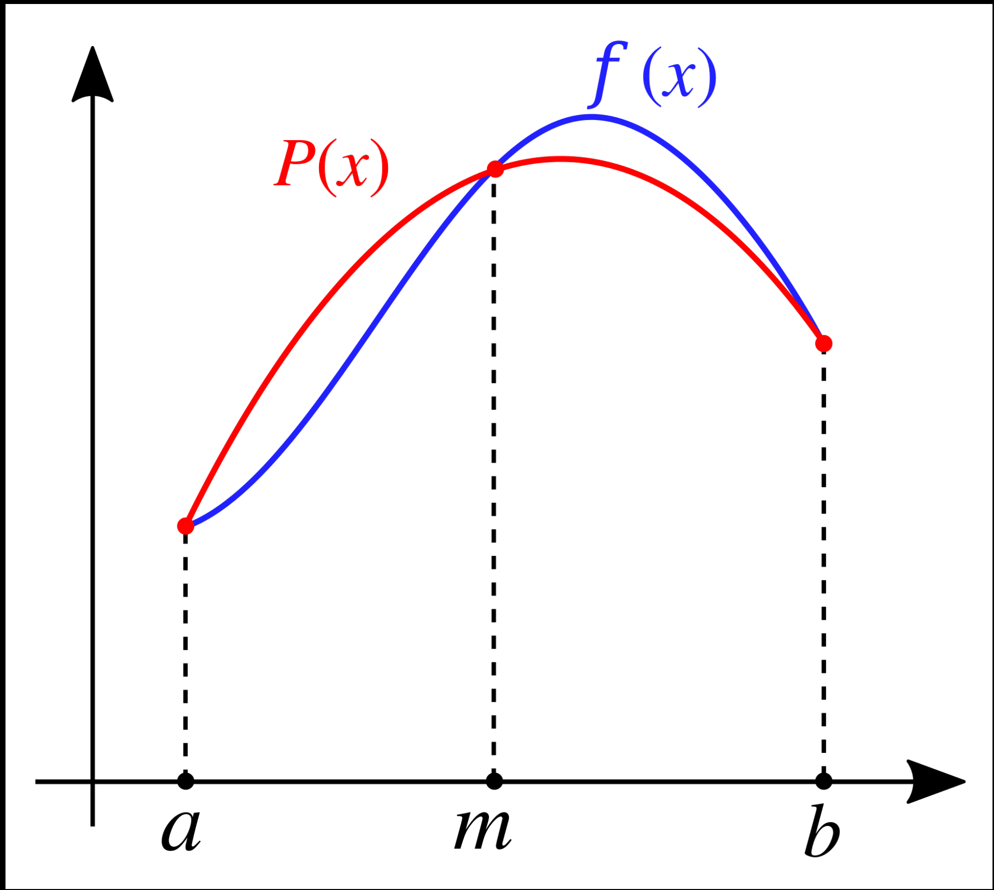
Shortcomings of traditional numerical methods

- Traditional methods work well for low dimensional integrals:

Simpson's rule:

$$\int_{x_0}^{x_2} dx\ f(x) = \frac{\Delta x}{3}\ [f(x_0) + 4f(x_1) + f(x_2)] + \frac{(\Delta x)^5}{90} f^{(4)}(\xi)$$

Notice $\quad \Delta x \propto \frac{1}{N}$



- This can be improved

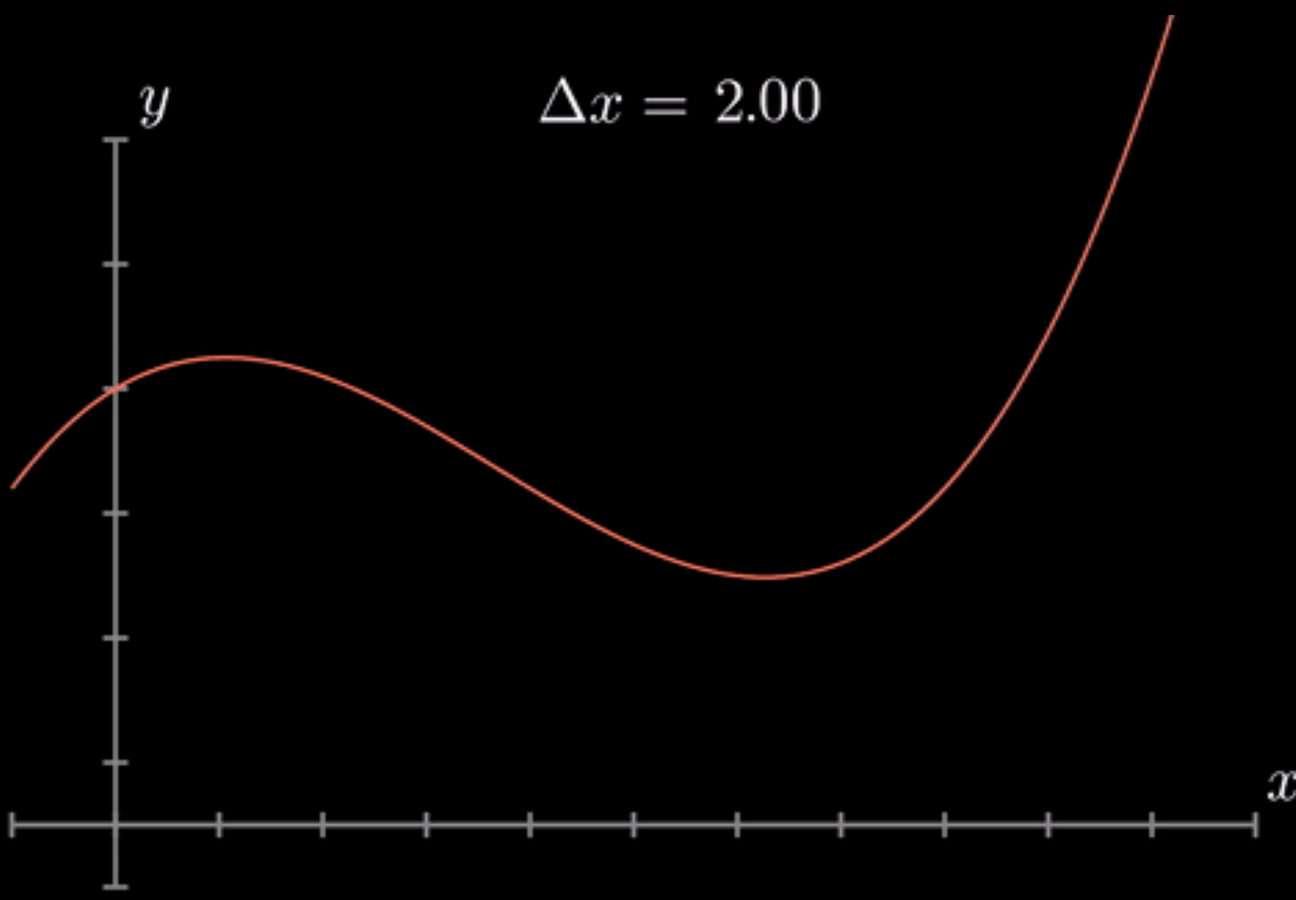
Shortcomings of traditional numerical methods

- Traditional methods work well for low dimensional integrals:

Simpson's rule:

$$\int_{x_0}^{x_2} dx\ f(x) = \frac{\Delta x}{3}\ [f(x_0) + 4f(x_1) + f(x_2)] + \frac{(\Delta x)^5}{90} f^{(4)}(\xi)$$

Notice $\Delta x \propto \frac{1}{N}$



- This can be improved

- Traditional methods work well for low dimensional integrals:

method/uncertainty	1 dimension	d dimensions
Trapedoidal rule	$\frac{1}{n^2}$	$\frac{1}{n^{2/d}}$
Simpson's rule	$\frac{1}{n^4}$	$\frac{1}{n^{4/d}}$
Gauss rule	$\frac{1}{n^{2m-1}}$	$\frac{1}{n^{(2m-1)/d}}$
Monte Carlo	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$

- 1/d factor renders the methods inefficient

Example

$$I = \int_0^1 dx \ \cos\left(\frac{\pi}{2}x\right) = \frac{2}{\pi} \simeq 0.63661977$$

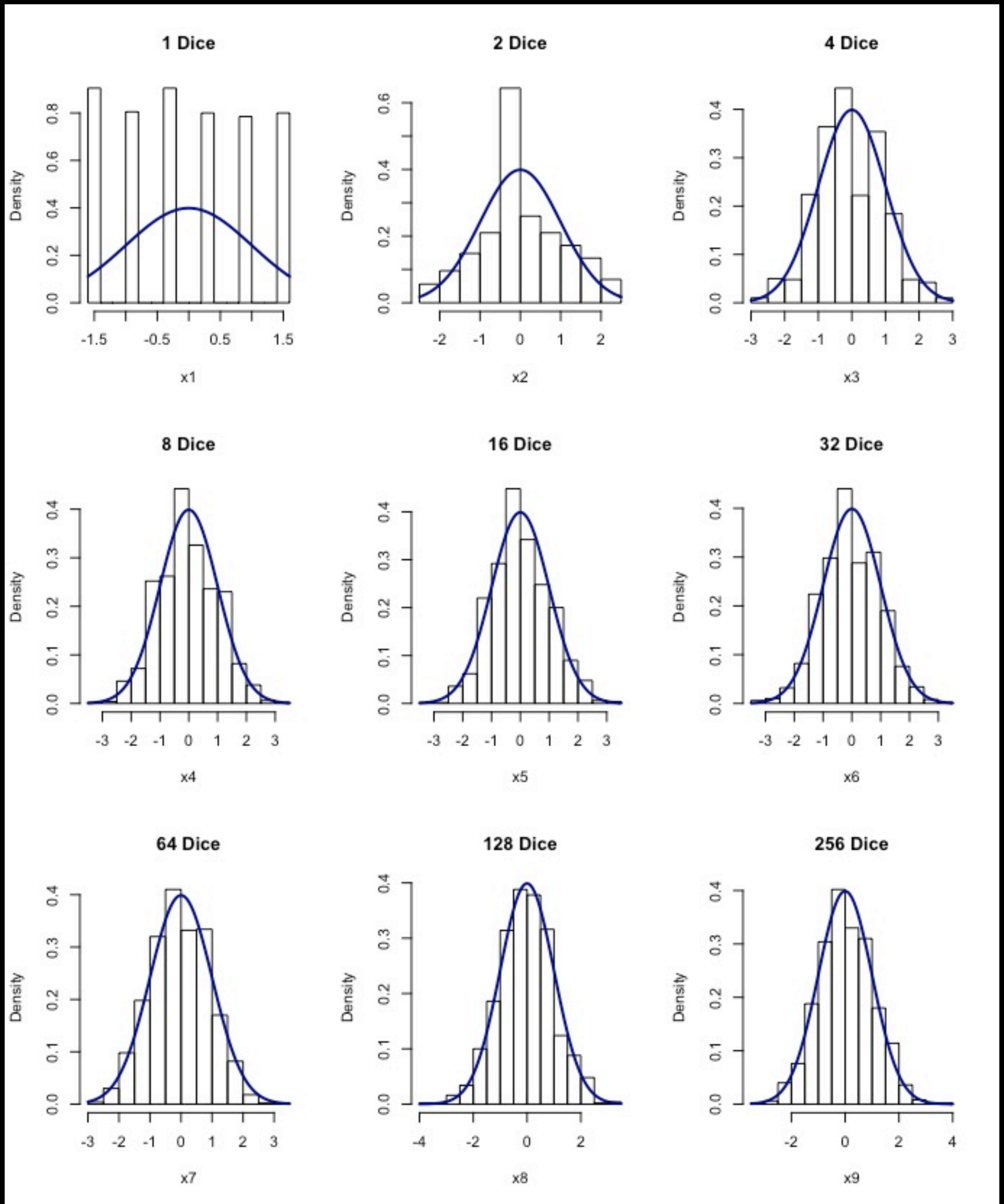
evaluations	Simpson	MC
3	0.638	0.3
5	0.6367	0.8
20	0.63662	0.6
100	0.636619	0.65
1000	0.636619	0.636

Introduction

- MC transforms the problem into a stochastic one.
- MC provides approximate solutions using statistical sampling experiments.
- MC has a wide range of applications from economics to physics
- MC is a statistical method used in simulation of data
- MC uses a sequence of random numbers as data
- MC can be applied to problems with no probabilistic content

Central Limit Theorem

- The sum of a "large number" of random variables is always normally distributed



Basic idea

- MC is the most efficient way to perform multi-dimensional integrals.
- The simplest idea: integrand is a function of a random variable

$$x \in [0,1] \qquad \text{and} \qquad \langle f \rangle = \int_0^1 dx\, f(x) \simeq \frac{1}{N} \sum_{j=1}^{N} f(x_j)$$

x is uniformly distributed [crude MC]

- f(x) is a crude estimator of $\langle f \rangle$
- f(x) is a random variable with variance

$$\sigma_1^2 = \int_0^1 dx\, \left(f - \langle f \rangle\right)^2 \qquad \Longrightarrow \qquad \sigma_N = \frac{\sigma_1}{\sqrt{N}}$$

- We can estimate the probability of the result being correct:

$$\lim_{N\to\infty} \text{Prob}\left(-a\frac{\sigma_1}{\sqrt{N}} \leq \frac{1}{N}\sum_{j=1}^{N} f(x_j) - I \leq b\frac{\sigma_1}{\sqrt{N}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{b} dt\; e^{-\frac{t^2}{2}}$$

- we can estimate the error from the MC simulation

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(f(x_j) - \langle f \rangle)^2$$

Initial remarks

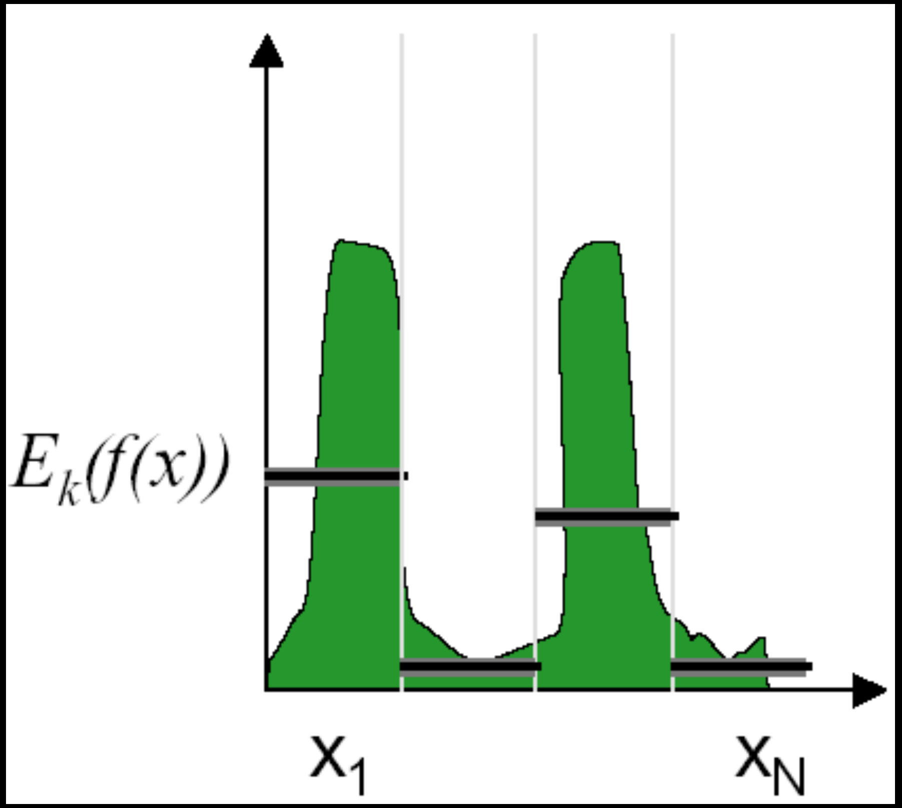
1. MC is exact for f constant. The flatter the better!
2. We should avoid near-singular integrands, e.g.,

$$\int \frac{ds}{(s-M)^2 + M^2\Gamma^2} = \frac{d\theta}{M\Gamma} \qquad \text{with} \qquad s - M^2 = M\Gamma \tan\theta$$

3. Avoid discontinuities of f if possible.
4. MC is a direct simulation of what happens physically.
5. We can also generate events weighted by f(x)
6. The dependence on N is fixed
7. We can improve the method reducing σ_1

$$\sigma_1^2 = \int_0^1 dx\ (f - \langle f \rangle)^2 \qquad \Longrightarrow \qquad \sigma_N = \frac{\sigma_1}{\sqrt{N}}$$

Stratified sampling



- just break the range of integration

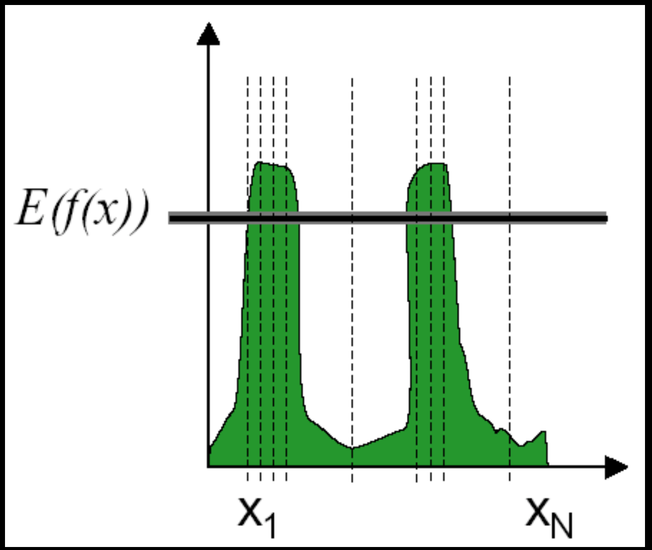
$$0 = \alpha_0 < \alpha_1 \cdots < \alpha_k = 1$$

- apply crude MC to each interval

$$\langle f \rangle \simeq \sum_{j=1}^{k} (\alpha_j - \alpha_{j-1}) \, \frac{1}{n_j} \sum_{i=1}^{n_j} f(\alpha_{j-1} + (\alpha_j - \alpha_{j-1}) \, x_{ij})$$

- variance is reduce for same number of calls of f.

Importance sampling



- use more points where the function is larger

- implementation using g(x) pdf:

$$\langle f \rangle = \int_0^1 dx\ f(x) = \int_0^1 dx\ g(x)\ \frac{f(x)}{g(x)} = \int_0^1 dG\ \frac{f(x)}{g(x)}$$

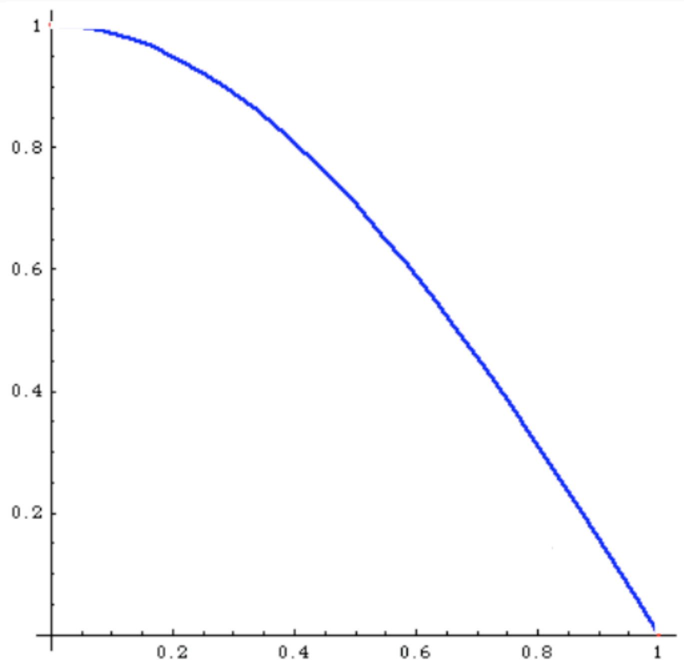
where $$G(x) = \int_0^x dy\ g(y)$$

- Generating random numbers according g(x):

$$\langle f \rangle = \frac{1}{N} \sum_{j=1}^{N} \frac{f(x_j)}{g(x_j)}$$

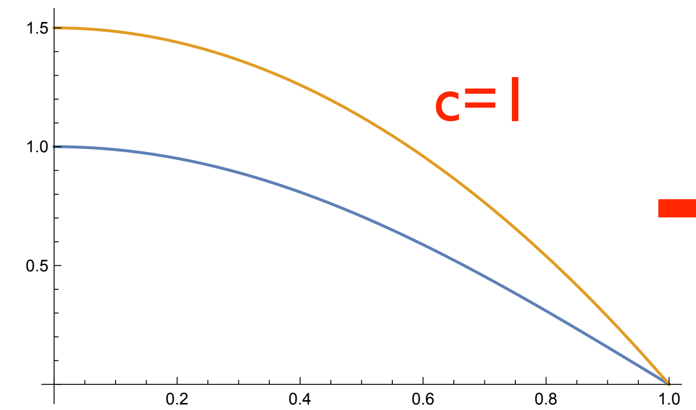
- choosing g(x) we can reduce the variance.
- the variance is $$\sigma^2_{f/g} = \int_0^1 dG \, \left(\frac{f(x)}{g(x)} - \langle f \rangle \right)^2$$
- g should be simple to obtain G explicitly
- if g=cf the variance vanishes
- choose a good function g similar to f
- g(x) approaching zero might jeopardize the variance gain

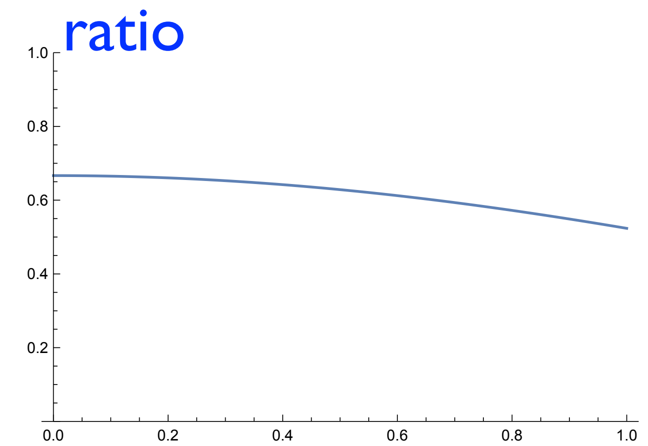
Example



$$I = \int_0^1 dx \cos \frac{\pi}{2} x$$

$$I_N = 0.637 \pm 0.307/\sqrt{N}$$





$$I = \int_0^1 dx (1 - cx^2) \frac{\cos\left(\frac{\pi}{2}x\right)}{(1 - cx^2)} = \int_{\xi_1}^{\xi_2} d\xi \frac{\cos \frac{\pi}{2} x[\xi]}{1 - x[\xi]^2 c}$$

$$\Rightarrow \simeq 1$$

$$I_N = 0.637 \pm 0.031/\sqrt{N}$$

The change reduces the number of evaluations by 100

- Main features of the method:

1. Random variable distribution should be close to the integrated function
2. Basically this is a change of variable to render the integrand flatter
3. Requires the knowledge of the approximate function
4. Requires knowing the integrated function main features

Control Variates

- Use g(x) with a known integral as

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

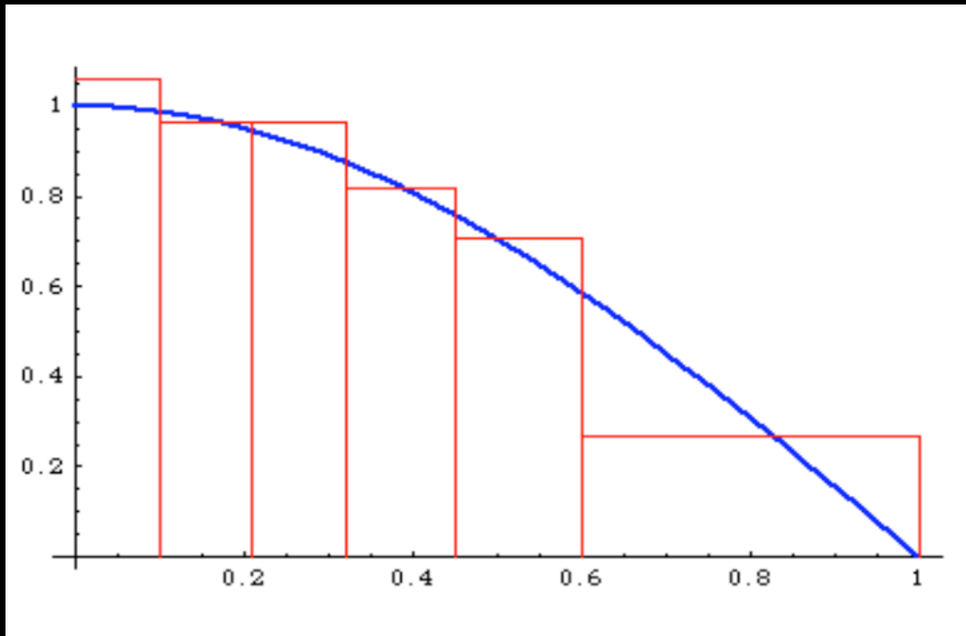
to reduce the variance

Adaptative Monte Carlo

- Basic idea: to create the approximation function on the flight
- VEGAS algorithm combines stratified and importance samplings
- VEGAS creates an approximated version of the ideal function

$$g(x) = \frac{|f(x)|}{\int dx\, |f(x)|}$$

1. start with equal bins
2. rebin to each bin have similar contribution
3. redo the integration with importance sampling and return to 2



1. start with equal bins
2. rebin to each bin have similar contribution
3. redo the integration with importance sampling and return to 2

- At each iteration j

$$E_j = \frac{1}{N_j}\sum_{k=1}^{N_j}\frac{f(x_k)}{g(x_k)} \quad \text{with estimated variance} \quad S_j = \frac{1}{N_j}\sum_{k=1}^{N_j}\left(\frac{f(x_k)}{g(x_k)}\right)^2 - E_j^2$$

the cumulative estimate after m iterations is

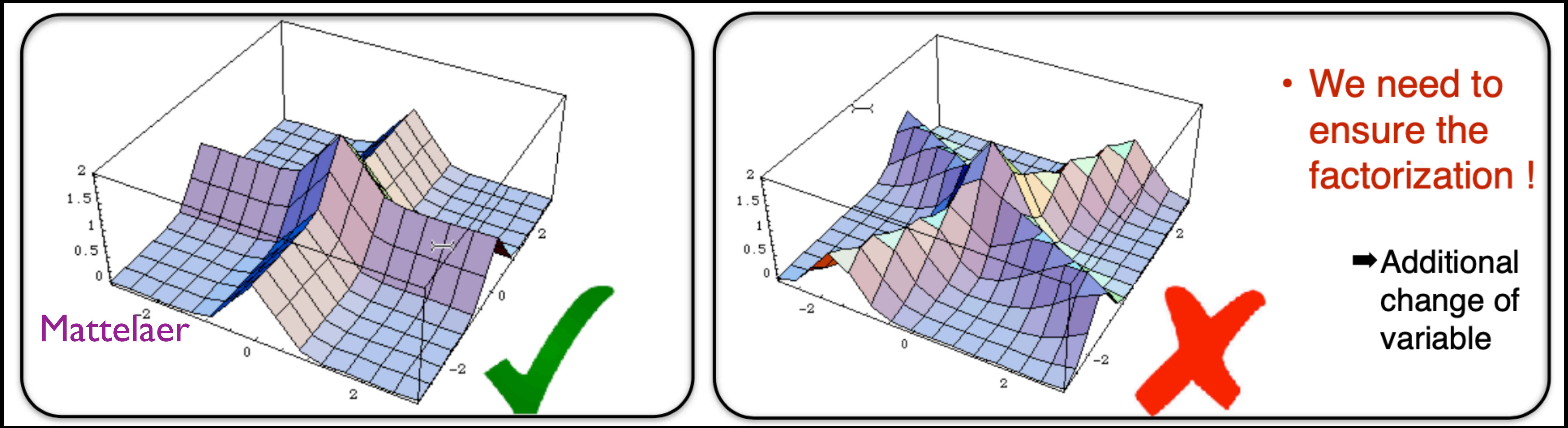
$$E = \left(\sum_{j=1}^{m}\frac{N_j}{S_j^2}\right)^{-1}\sum_{j=1}^{m}\frac{N_j E_j}{S_j^2}$$

and the estimated chi squared $\quad \dfrac{\chi^2}{dof} = \dfrac{1}{m-1}\displaystyle\sum_{j=1}^{m}\frac{(E-E_j)^2}{S_j^2}$

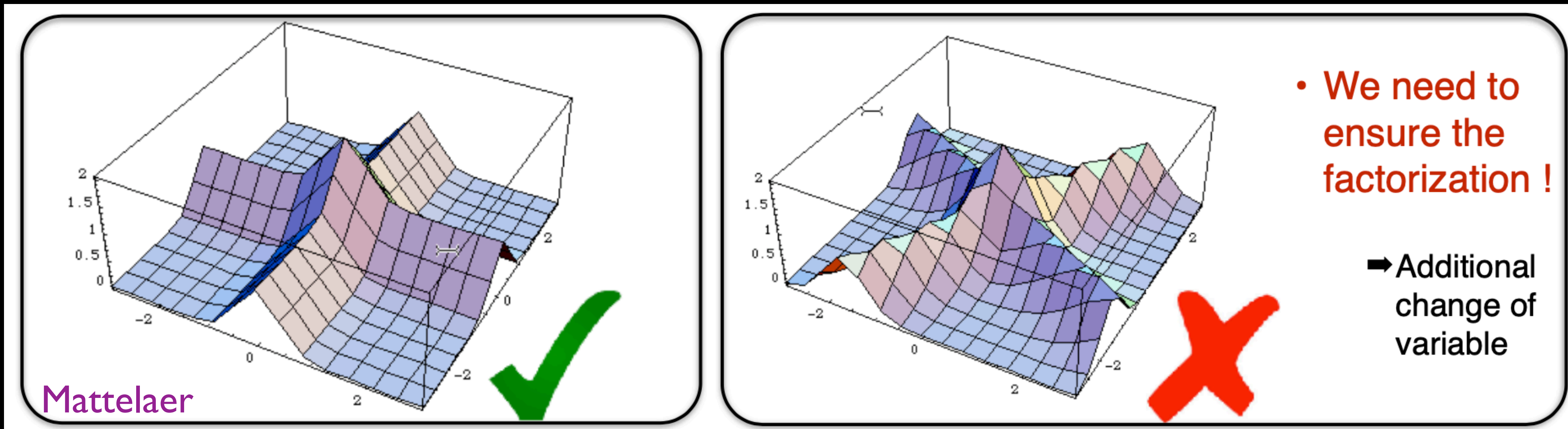
- In more than one dimension, we use a separable probability g(x)

$$g(x) = g_{x_1}(x_1) g_{x_2}(x_2) \ldots g_{x_D}(x_D)$$

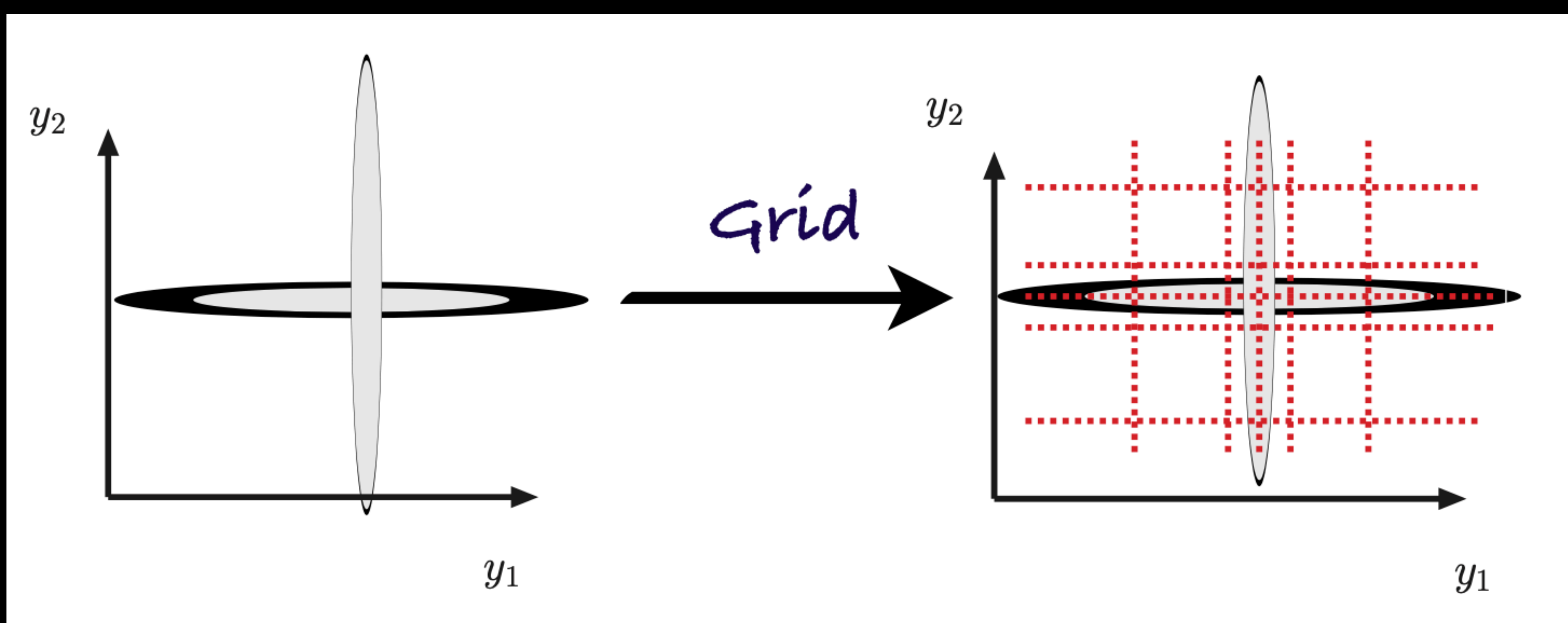
- there are potential problems



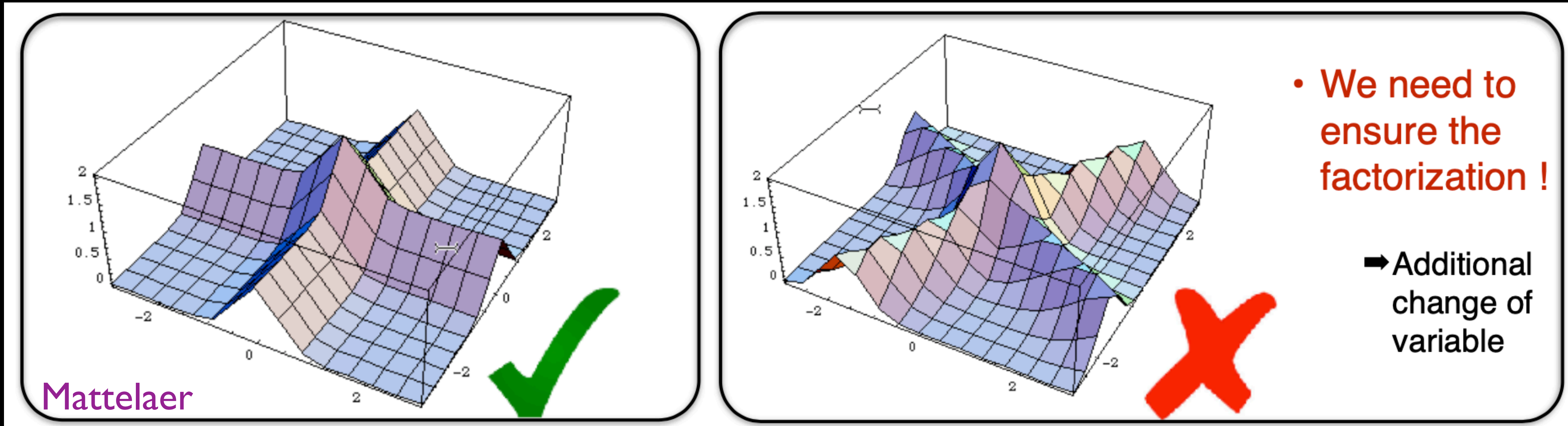
- there are potential problems



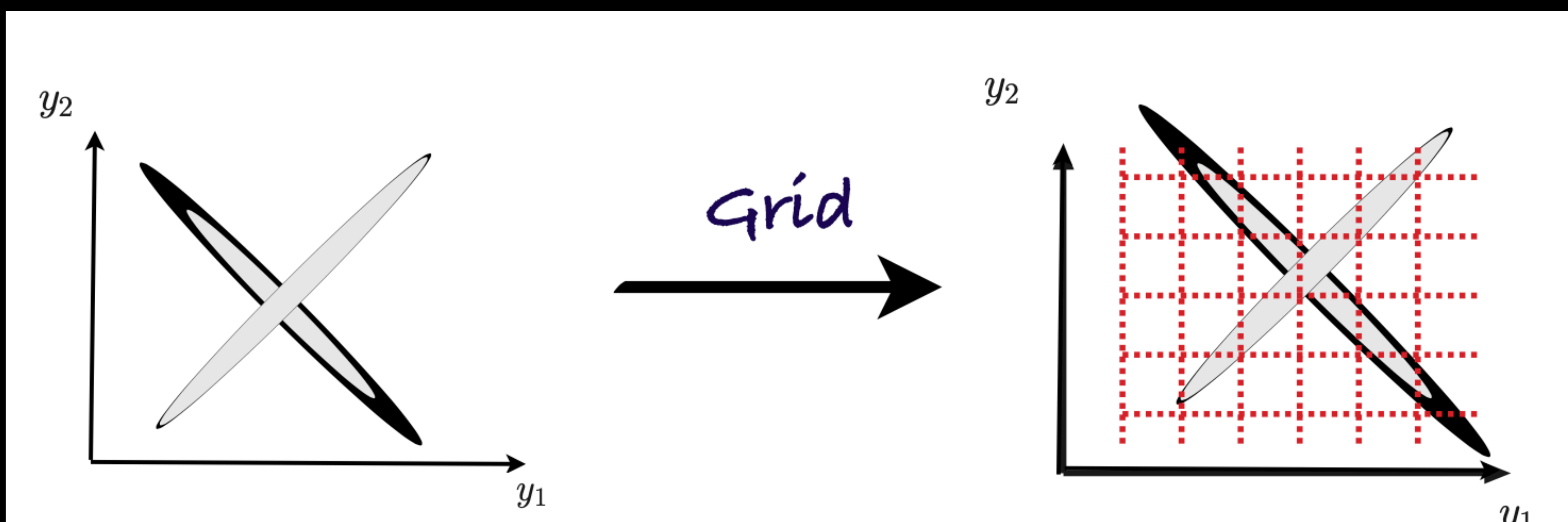
- There can be impact in the efficiency



- there are potential problems

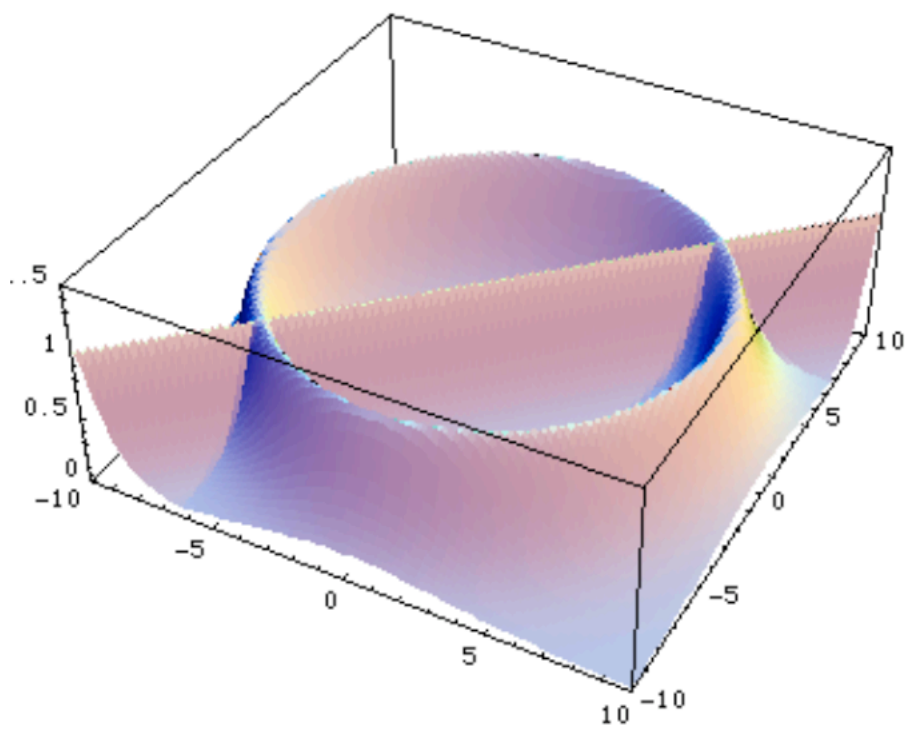


- There can be impact in the efficiency



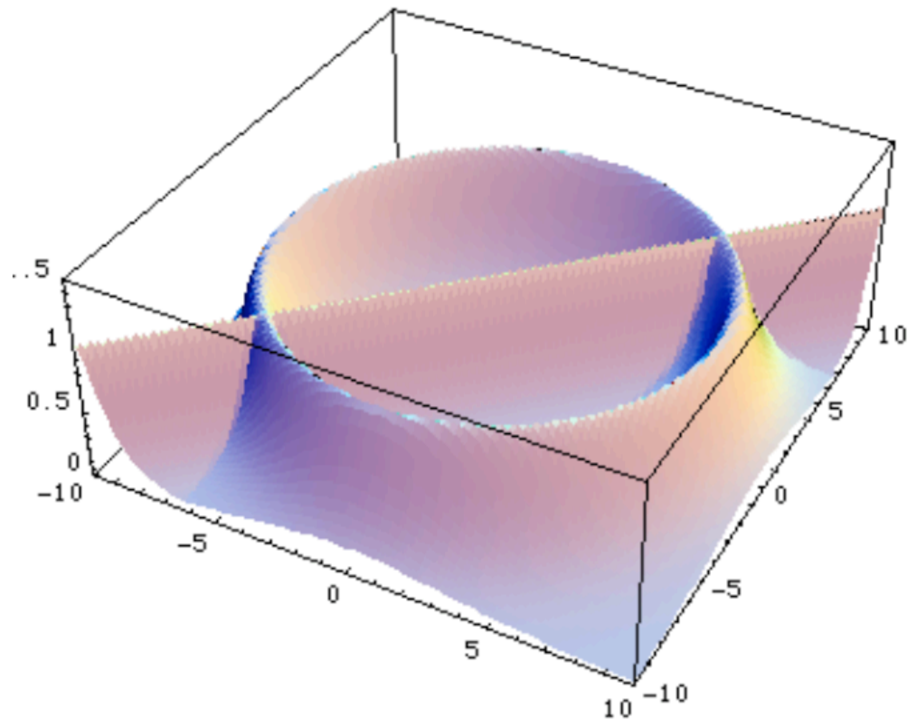
we need a more efficient method

Multi-channel Monte Carlo



What do we do if there is no transformation that aligns all integrand peaks to the chosen axes? Vegas is bound to fail!

[extracted from O. Mattelaer, MC Lecture at IFT Madrid, 2015]



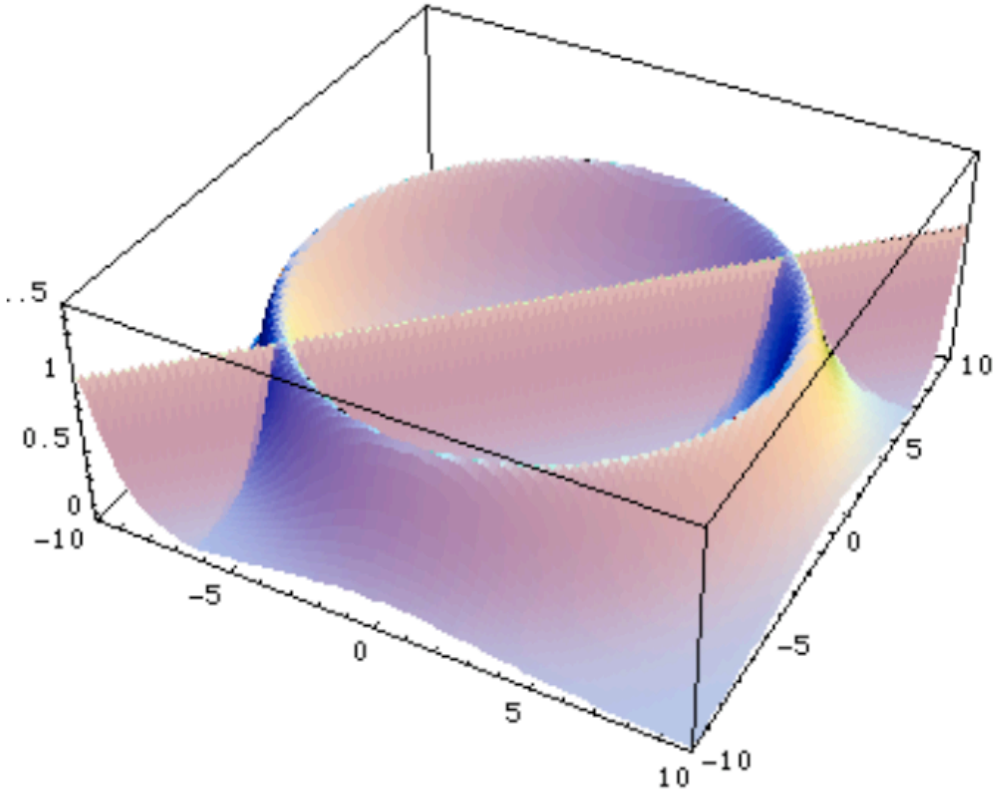
What do we do if there is no transformation that aligns all integrand peaks to the chosen axes? Vegas is bound to fail!

Solution: use different transformations = channels

$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x) \qquad \text{with} \qquad \sum_{i=1}^{n} \alpha_i = 1$$

with each $p_i(x)$ taking care of one "peak" at the time

- Catch: we need to know the integrand! Know thy problem :-)

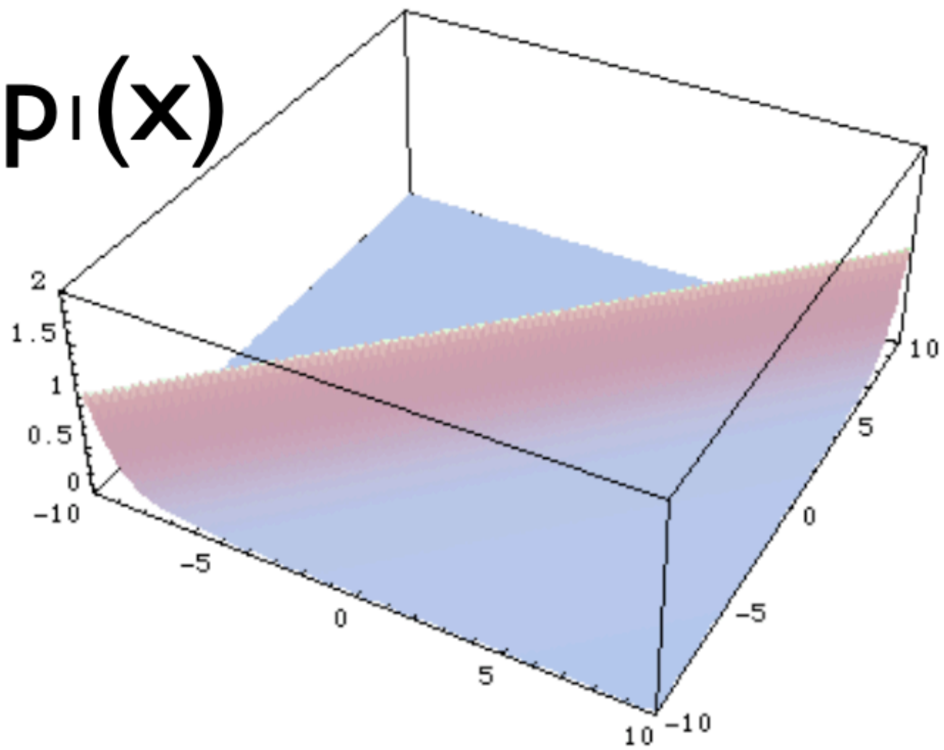


$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x)$$

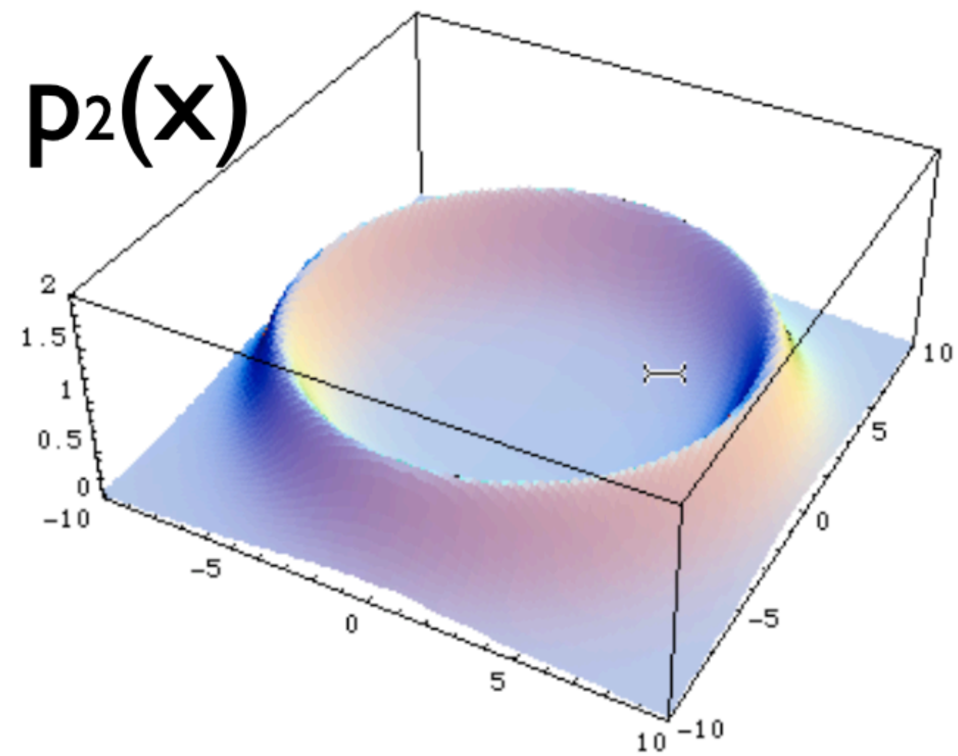
with

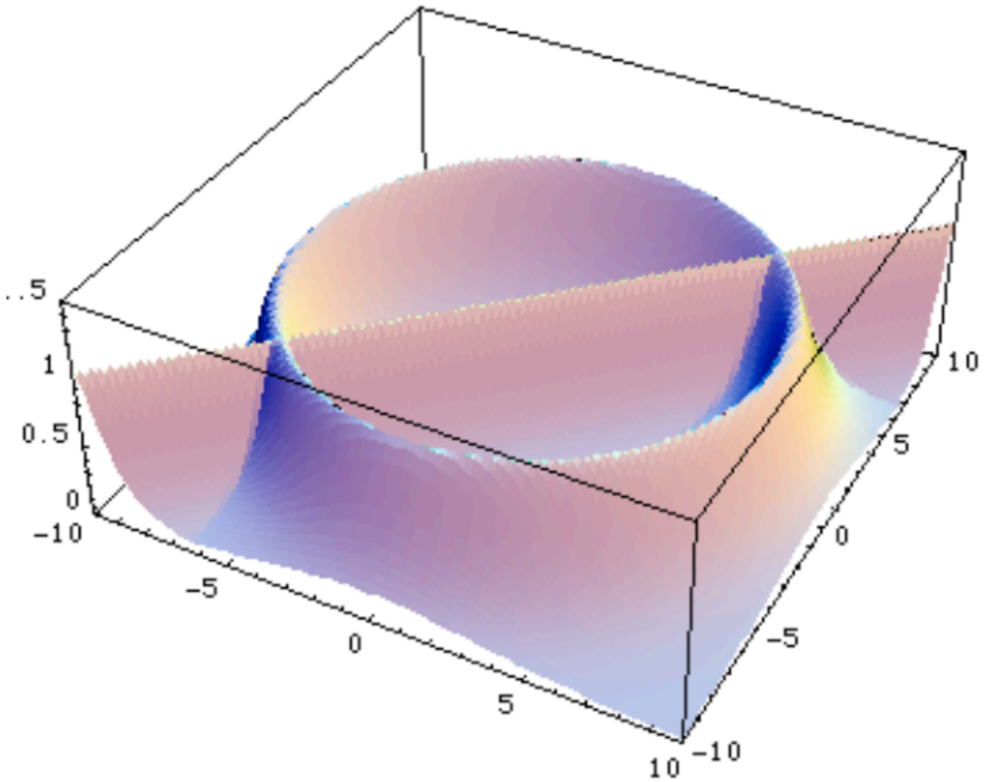
$$\sum_{i=1}^{n} \alpha_i = 1$$

$p_1(x)$



$p_2(x)$





$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x)$$

with

$$\sum_{i=1}^{n} \alpha_i = 1$$

Then,

$$I = \int f(x)dx = \sum_{i=1}^{n} \alpha_i \int \underbrace{\boxed{\frac{f(x)}{p(x)}}}_{\approx 1} p_i(x)dx$$

Weighted to Unweighted Events

- How do we generate x according to the pdf p(x)?
- First method: inverse transform method
- Consider the cumulative distribution function

$$P(x) = \int_{-\infty}^{x} dt\ p(t) \Longrightarrow P(x) \in [0,1]$$

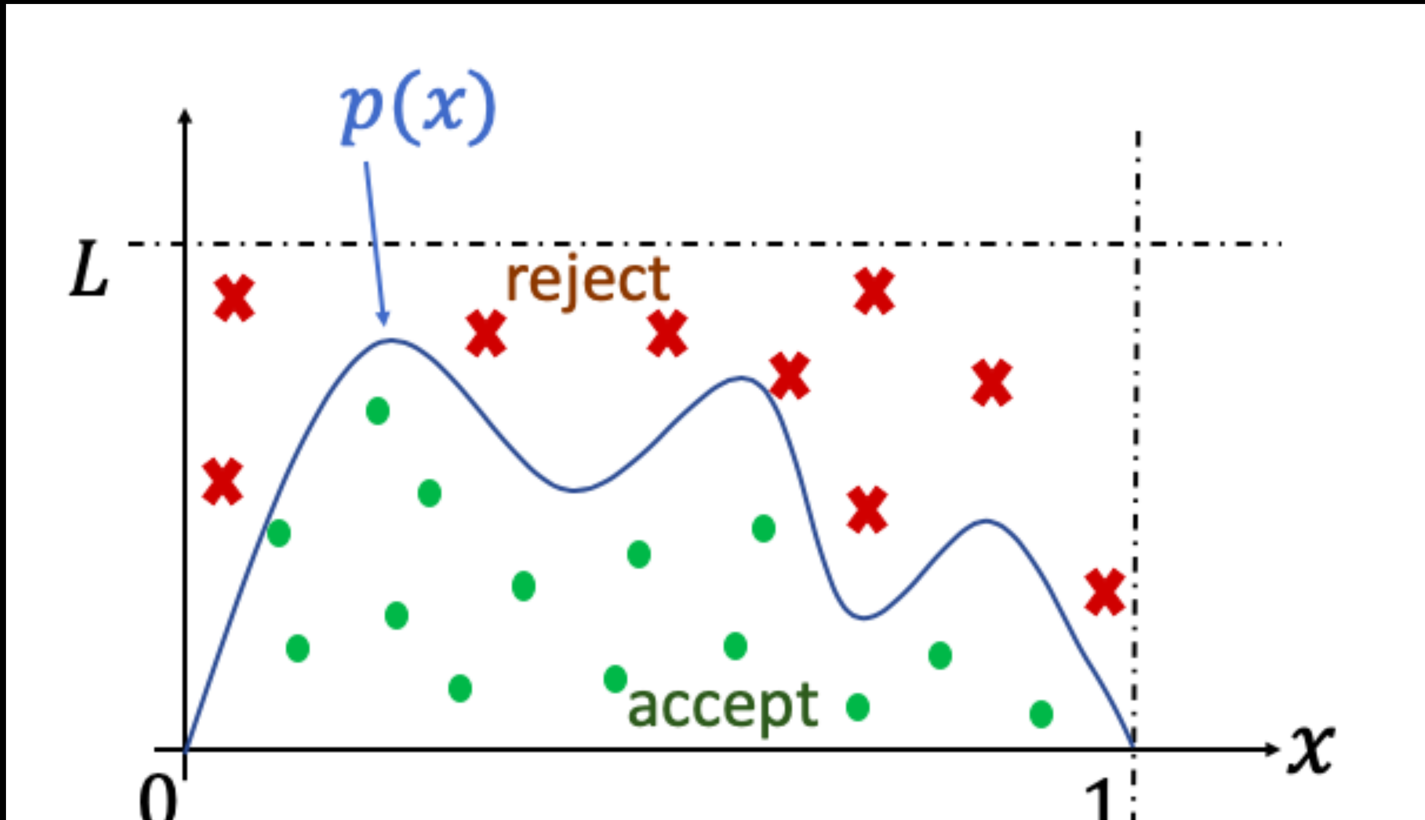
- Now consider a random variable u uniformly distributed in [0,1]

$$x = P^{-1}(u) \Longrightarrow p(x)dx = du$$

- This requires knowing analytically the inverse of P

Acceptance-rejection method

- Consider x a random variable uniformly distributed between [0,1]
- Draw a first value of the x (x_1)
- Draw a second value of x (x_2)
- Accept x_1 if $p(x_1) \geq x_2 L$
- the resulting distribution of accepted points follow p(x)



- Let's return to the cross section evaluation

$$\sigma = \int dx_1 dx_2 \sum_{\text{subp}} f_{a_1/p}(x_1)\, f_{a_2/\bar{p}}(x_2)$$
$$\frac{1}{2\hat{s}(2\pi)^{3n-4}} \int d\Phi_n(x_1 P_A + x_2 P_B;\ p_1 \dots p_n)\Theta(\text{cuts})\overline{\sum}|\mathcal{M}|^2(a_1 a_2 \to b_1 \dots b_n)$$

that requires a suitable choice of the integration variables

- Initially we map the integration region into a 3n-2 hypercube

$$dx_1 dx_2 d\Phi_n = J \prod_{i=1}^{3n-2} dr_i$$

- It is easy to reconstruct the momenta and implement the cuts
- This procedure generate weighted events with weight

$$\mathbf{w} = \sum_{\{\mathbf{r_i}\}} \frac{\mathbf{J}}{\mathbf{2\hat{s}(2\pi)^{3n-4}}} \sum_{\mathbf{subprocesses}} \mathbf{f(x_1)f(x_2)}\overline{\sum}|\mathcal{M}|^{\mathbf{2}}\ \mathbf{\Theta}(\mathrm{cuts})\ ,$$

- Now it is possible to generate distributions
- Unweighted events can also be obtained

✷ Example: $\mathrm{e^+e^-} \to \mathbf{2}$ particles in the final state

$$d\Phi_2 = \frac{1}{4}\frac{|\vec{p}_1^{cm}|}{\sqrt{s}}\, d\cos\theta_1 d\phi_1 = \frac{1}{4}\frac{|\vec{p}_1^{cm}|}{\sqrt{s}} \times 4\pi \times dr_1 dr_2$$

with $\cos\theta_1 = -1 + 2r_1$ and $\phi_1 = 2\pi r_2$. More, I can construct the (massless) momentum with this

$$p_1 = \frac{\sqrt{s}}{2}\left(1,\ \sin\theta_1\cos\phi_1,\ \sin\theta_1\sin\phi_1,\ \cos\theta_1\right)$$

$$p_2 = \frac{\sqrt{s}}{2}\left(1,\ -\sin\theta_1\cos\phi_1,\ -\sin\theta_1\sin\phi_1,\ -\cos\theta_1\right)$$

References

- S. Weinzierl, hep-ph/0006269
- F James Rep. Prog. Phys. 43 (1980) 1145
- Barger & Phillips, chapter 11
- G. Peter Lepage, arXiv:2009.05112
- O. Mattelaer https://cp3.irmp.ucl.ac.be/projects/madgraph/attachment/wiki/Madrid2015_2%20MC/15_09_07_Madrid_MC.pdf