COMPUTATIONAL PHYSICS

Monte Carlo methods

Fundamentals and Integration

Importance sampling

Acceptance-rejection

Singularities and Multi-dimensional integrals

What and what for ?

- Invented in the context of the atomic bomb development and published by Metropolis&Ulam
- Is a class of computational algorithms.
- Relies on repeated *random* sampling to solve *deterministic* problems.
- Involves simulation from *probability distributions*.
- Is used in wide range of fields e.g. physics, finance, data analysis.
- In most practical applications it is a precious help to solve *optimization* and *integration* problems.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 247 SEPTEMBER 1949 *Volume 44*

THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM
Los Alamos Laboratory

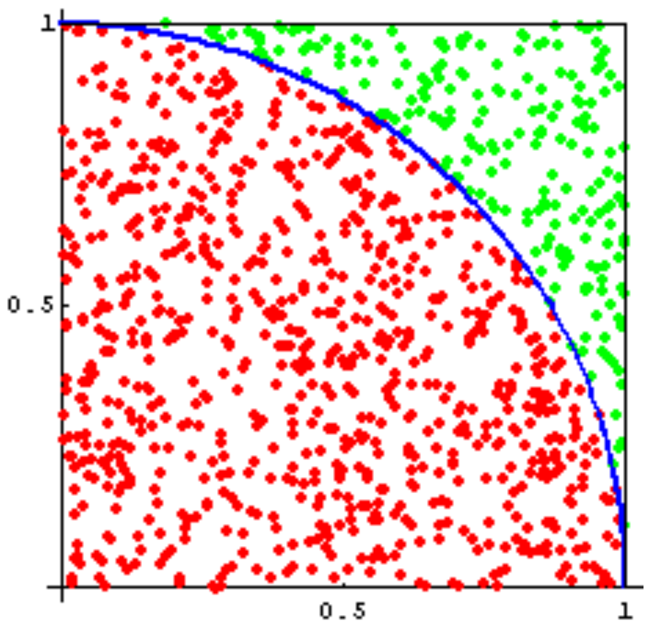
We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

Random processes and simulations

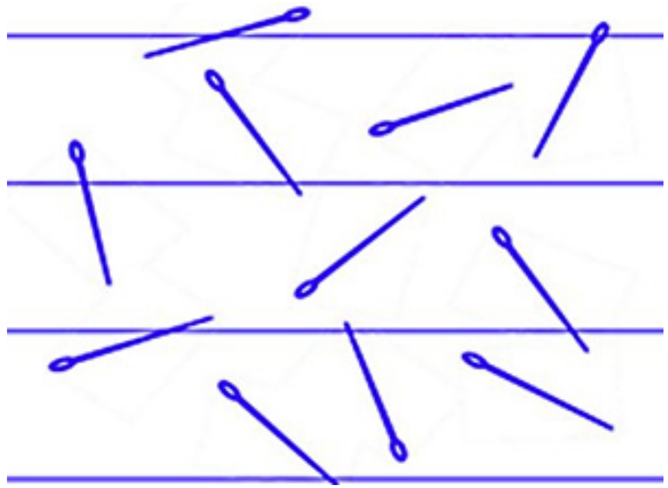
- Random draws following a given probability are a powerful tool to estimate the evolution of physical system or even simple mathematical expressions e.g. integrals !

Two "classical examples"

- Estimation of Pi



- Buffon needle experiment

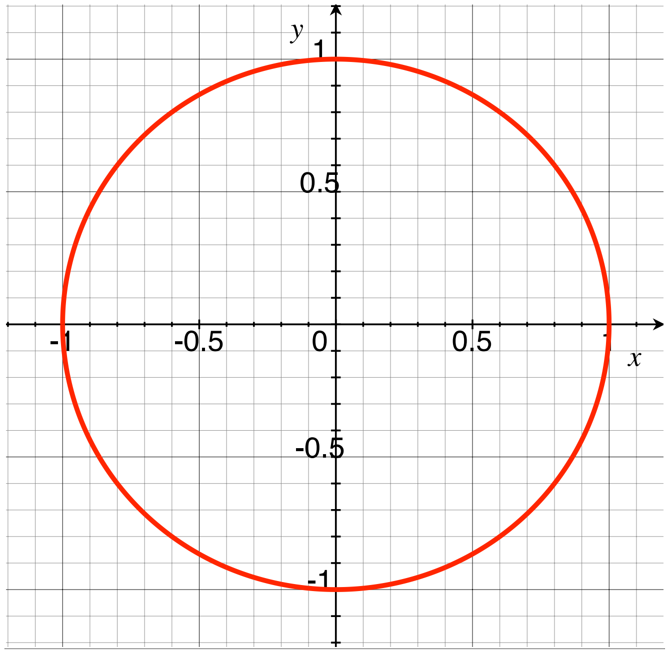


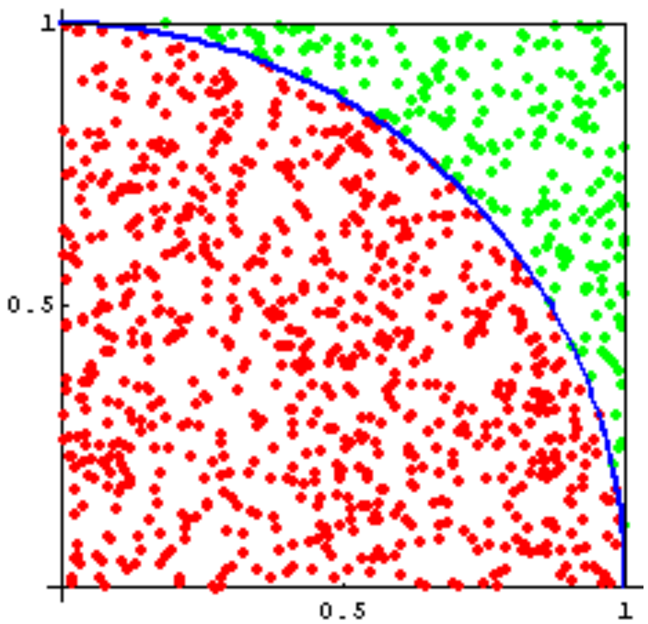
Estimating Pi…

- *Basic principle*:

  → Shoot random pairs (x,y) inside a [-1,1]x[-1,1] domain

  → Simple logic dictates that, for large number of "draws", the ratio of hits inside the unit circle to the total draws is roughly $\pi \times 1^2 / (2 \times 2) = \pi/4$!



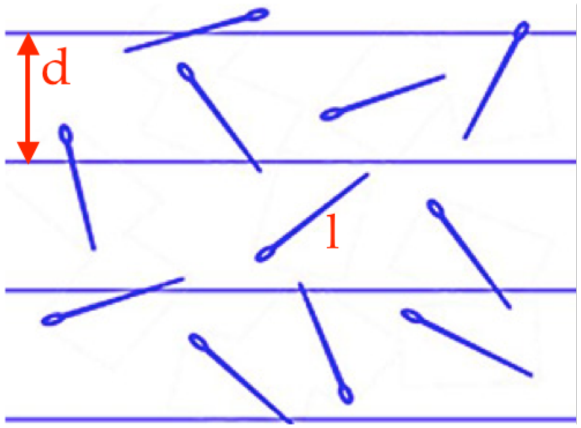


$$\frac{\pi}{4} \approx \frac{\text{Hits inside}}{\text{Total draws}}$$

We just estimated π !!!

Buffon Needle

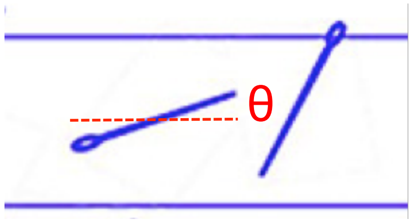
- *"Suppose that you drop a short needle on a ruled paper – what is then the probability that the needle comes to lie in a position where it crosses one of the lines ?"* – Comte de Buffon 1777



- Interestingly enough…tossing many small needles (l<<d) the probability also involves the number π !
- Indeed, the probability = $\dfrac{2l}{\pi d}$ …

Some hint on how to get there......

- If the center of the needle is closer to one of the ruled lines by less that $l \times \sin(\theta)$ then we have a line crossing !



Recalling a bit of statistics...

- Consider a data sample consisting of N sample points of variable X i.e. $(X_1,X_2,\ldots,X_N)$
- One defines the expected value of the sample X as

$$E(X) = \langle X \rangle = \frac{1}{N}\sum_{i=1}^{N} X_i$$

- The variance* of the sample X is defined as

$$Var(X) \equiv \sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \langle X \rangle\right)^2 = \langle X^2 \rangle - \langle X \rangle^2$$

*$\sigma^2_x = (N-1)/N\ \sigma^2$

- The standard deviation of the sample X is defined as

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \langle X \rangle\right)^2} = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$$

Probability density functions (PDFs)

- If p(X) gives the probability that an event X has to occur e.g. to get value X_i, then one knows that the following holds:

$$\int_{-\infty}^{+\infty} p(X)dX = 1$$

- The mean of the sample X with a PDF given by p(X) is defined as

$$< X >= \int_{-\infty}^{+\infty} p(X)X\,dX \qquad \text{or} \qquad < X >= \sum_{i=1}^{N} p(X_i)X_i$$

$$< f(X) >= \int_{-\infty}^{+\infty} p(X)f(X)\,dX \qquad \text{or} \qquad < f(X) >= \sum_{i=1}^{N} p(X_i)f(X_i)$$

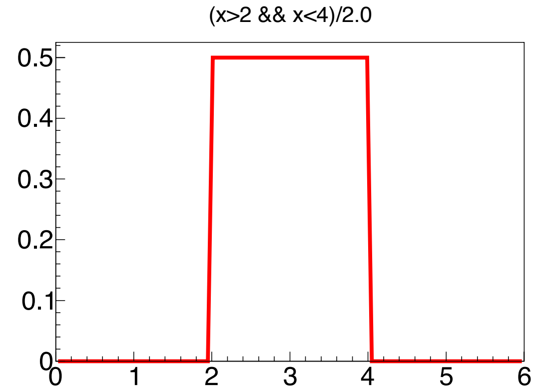
- The standard deviation of the sample X is defined as before

$$\sigma_X = \sqrt{< X^2 > - < X >^2} \qquad \sigma_{f(X)} = \sqrt{< f(X)^2 > - < f(X) >^2}$$

Notable PDFs...

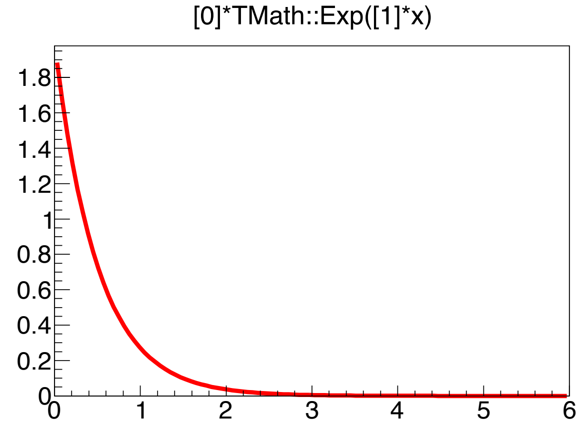
- Uniform distribution: X[a,b]

$$p(X) = \frac{1}{b-a} \qquad a < X < b$$



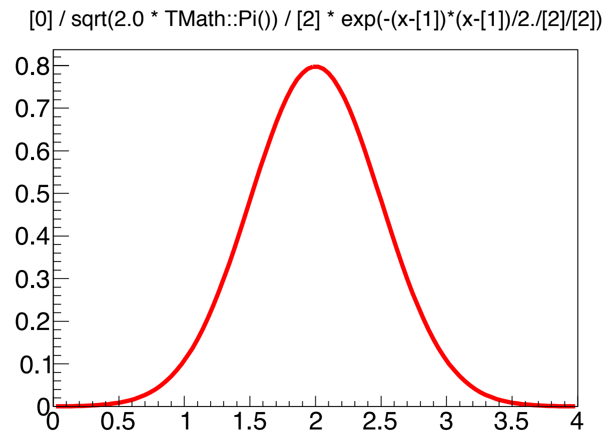
- Exponential distribution: X[0,+∞]

$$p(X) = \lambda e^{-\lambda X}$$



- Normal distribution: X[-∞,+∞]

$$p(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$



About random numbers...

- Easier said than done…a computer does what it is told to do unless it is broken... → sequenced numbers are ***ALWAYS*** predictable...
- The "***catch***": if one can create a absurdely long "*pseudo-random*" sequence, even if it is *periodic* it can be quite handy !
- One of the most simple yet effective (*if properly tuned*) pseudo-random generators algorithms is based on *Linear Congruential generators*

$$\boxed{Y_i = (aY_{i-1} + c)\%m}$$ →

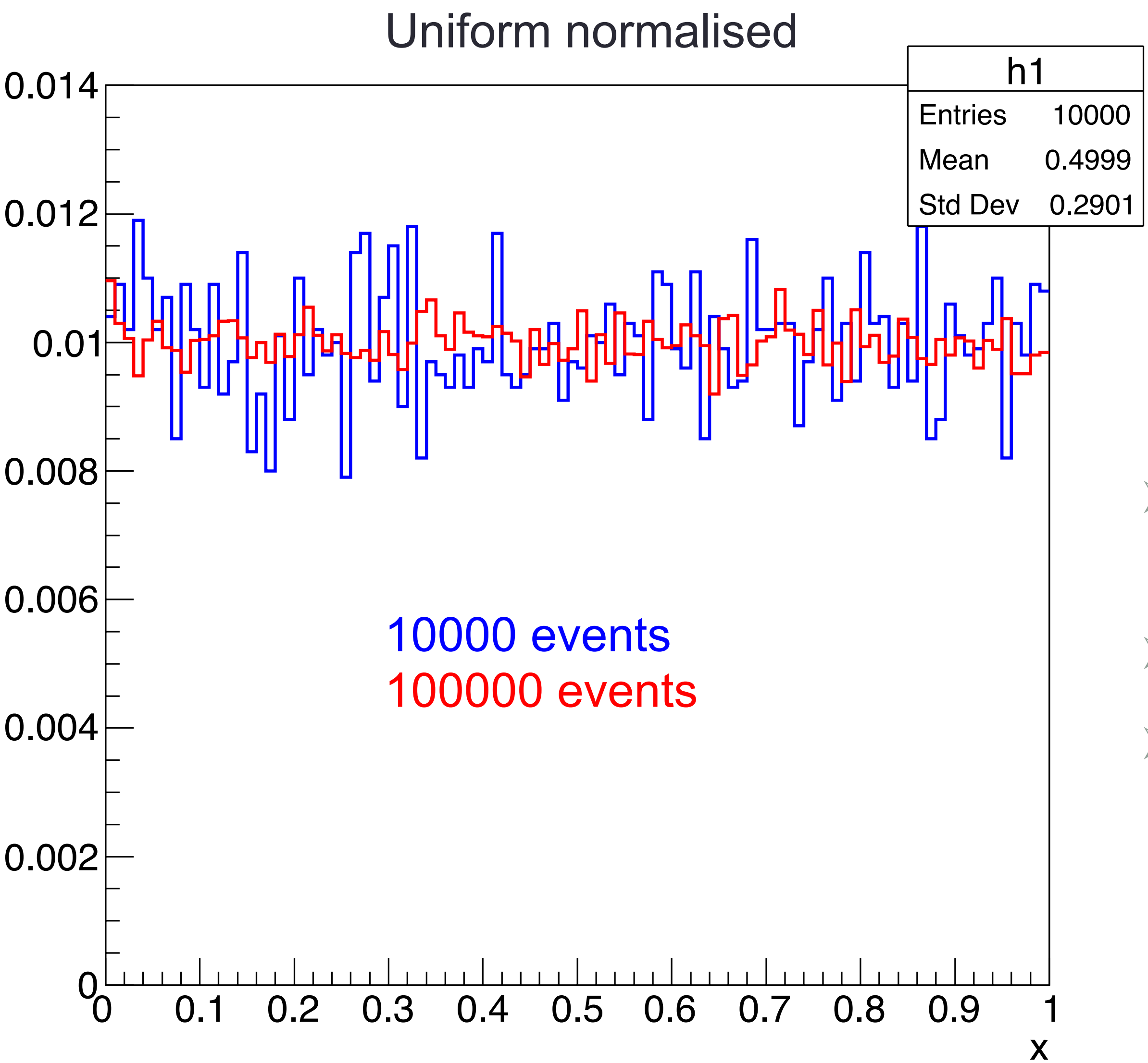
At most we get *period-m* but we can get worse (lower) if ***a*** and ***c*** are *not optimal*

Example: a=5,c=7,m=4 → Y = 2,1,0,3,2,…

✓ It is possible to obtain a good uniform random sequence with properly chosen parameters e.g. GCC uses $m=2^{31}$, a=1103515245, c=12345

✓ *No correlations* between random numbers, *fast* algorithm, *huge* period.

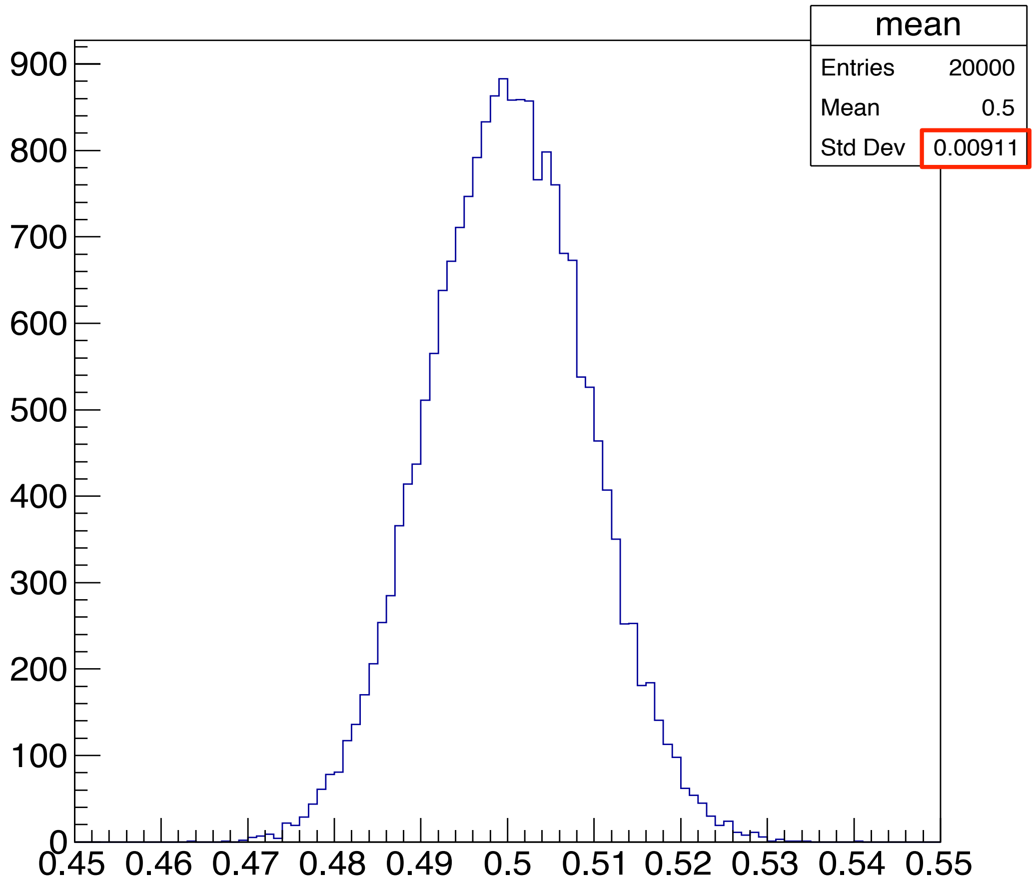
What do uniform distributions look like ?



$$\sigma = 0.29 \sim 1/\sqrt{12}$$

- Uniform(0,1) courtesy of ROOT.
- Stdev $\sigma = 1/\sqrt{12}$
- Same 100 bins, with higher number of events → lower variance in normalised histogram.

Revisiting the Central Limit Theorem



→ 100 bins from 20,000 samples of mean estimates from 1000 sampled Uniform(0,1)

→ $$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{12}}\frac{1}{\sqrt{\mathbf{1000}}} \cong \boxed{0.00913}$$

→ One says that the ***error on estimating the mean of the distribution is*** $\frac{\sigma}{\sqrt{n}}$

→ What if one wished to analyse the distribution of the normalised bin occupancy i.e. $(N_{bin}-<N_{bin}>)/<N_{bin}>$?!

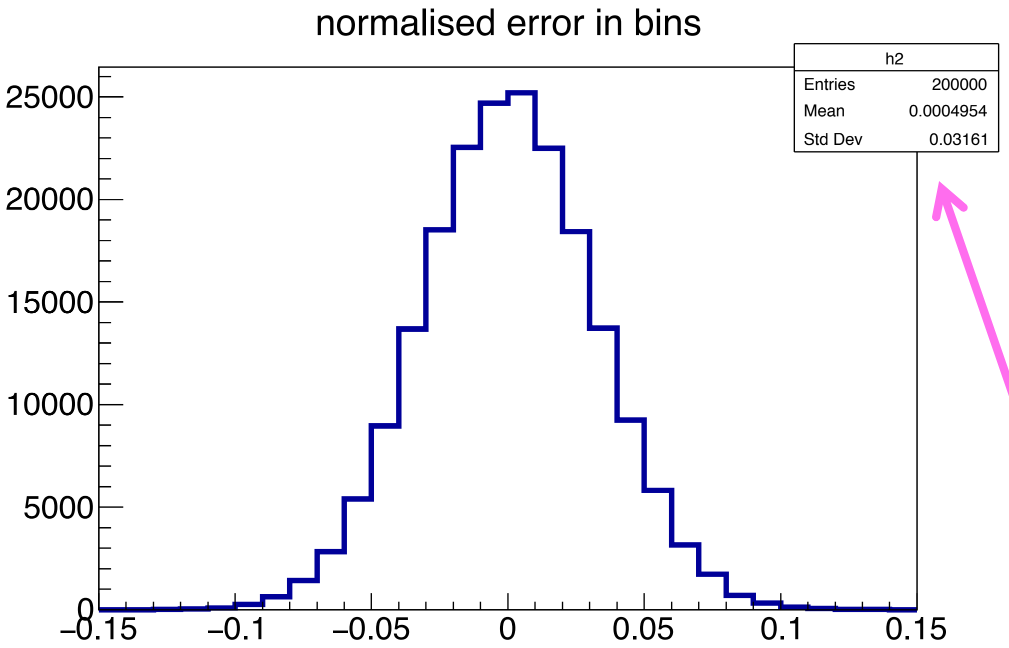
Revisiting the Central Limit Theorem

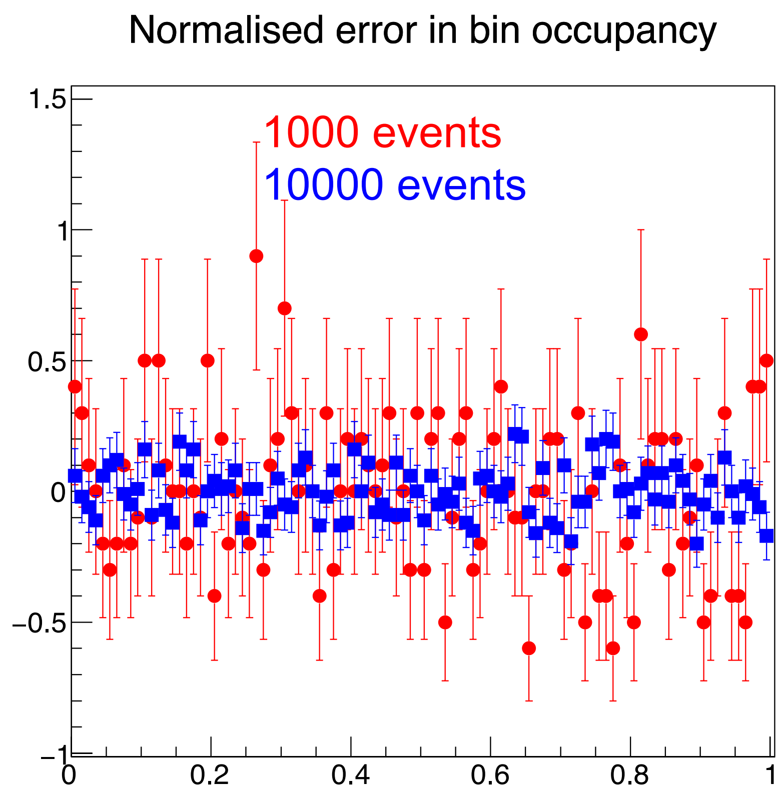
→ 100,000 events of Uniform(0,1)

→ 100 bins in histogram

→ Obtain each bin occupancy and normalise it's deviation from mean (here 1000)

→ Make 2000 experiments to get enough statistics...





$$\sigma = \frac{1}{\sqrt{N_{events} / N_{bins}}}$$

Generating randoms "*beyond*" uniform

- ***Goal***: generate random numbers ***X*** that do NOT follow a Uniform(0,1) distribution.

- ***Strategy***: use a change of variable ***X(Y)*** that conserves the cumulative distribution function (CDF) and with ***Y*** following Uniform(0,1) i.e.

$$\int_0^y dy' = \int_a^x p(x')\,dx' \quad \Leftrightarrow \quad y = \int_a^x p(x')\,dx'$$

$$\int_0^y dy' = 1$$

$$\int_a^b p(x')\,dx' = 1$$

- If the variable X follows a distribution function statistics p(x) that is *integrable analytically* and easily *invertible*, then one can easily obtain the functional x→x(y).

  - ✓ For each ***y*** in Uniform(0,1) we get automatically the corresponding ***x*** following p(x)

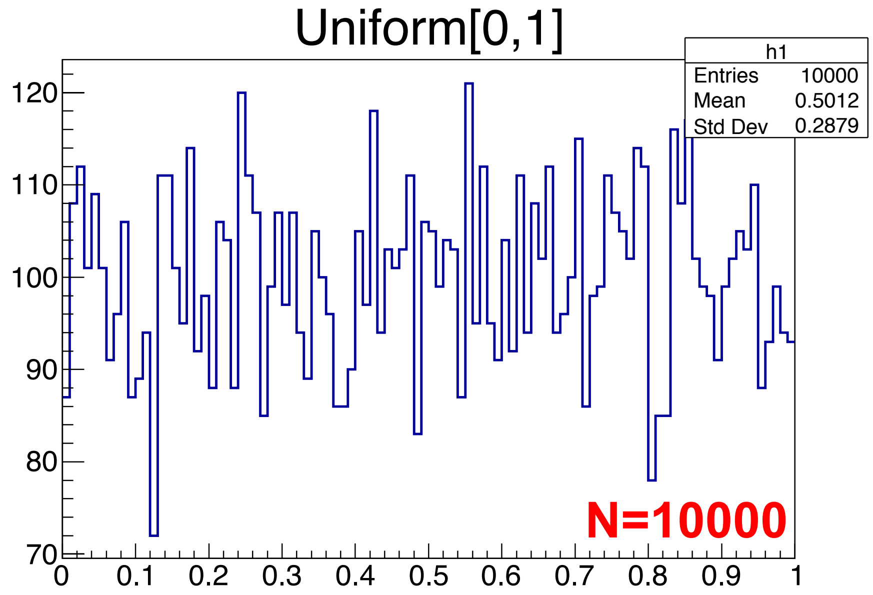
Uniform [0,1] → Uniform [a,b]

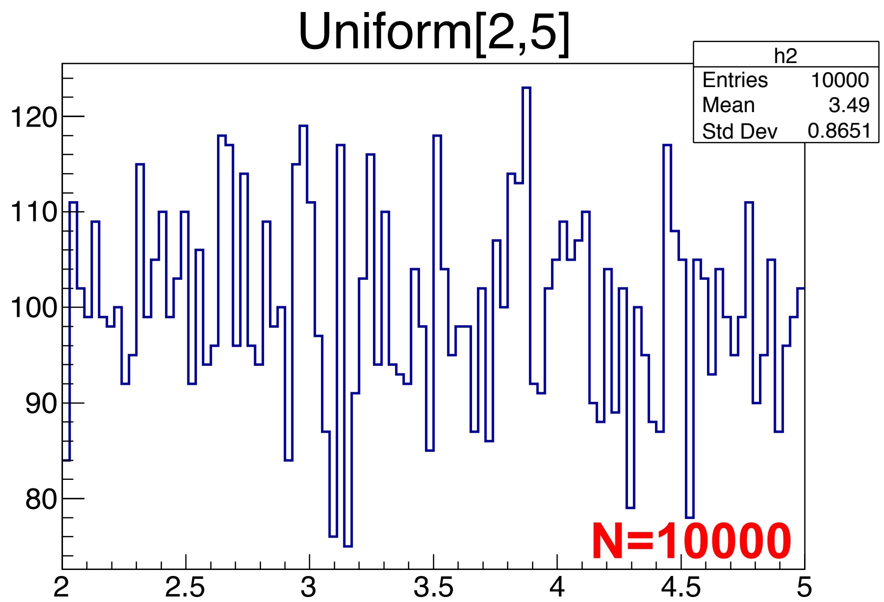
- ***Goal***: transform a Uniform(0,1) into a Uniform(a,b).
- For Uniform(a,b) $p(x)=\dfrac{1}{b-a}$
- Substitution:

$$y=\int_a^x \frac{1}{b-a}dx'=\frac{x-a}{b-a}$$

$$x-a=(b-a)y$$

$$\boxed{x=\underbrace{a}_{\text{shift}}+\underbrace{(b-a)}_{\text{rescale}}y}$$





Uniform [0,1] → Exponential [0,+∞[

- ***Goal***: transform a Uniform(0,1) into a exponential (λe^-λ).

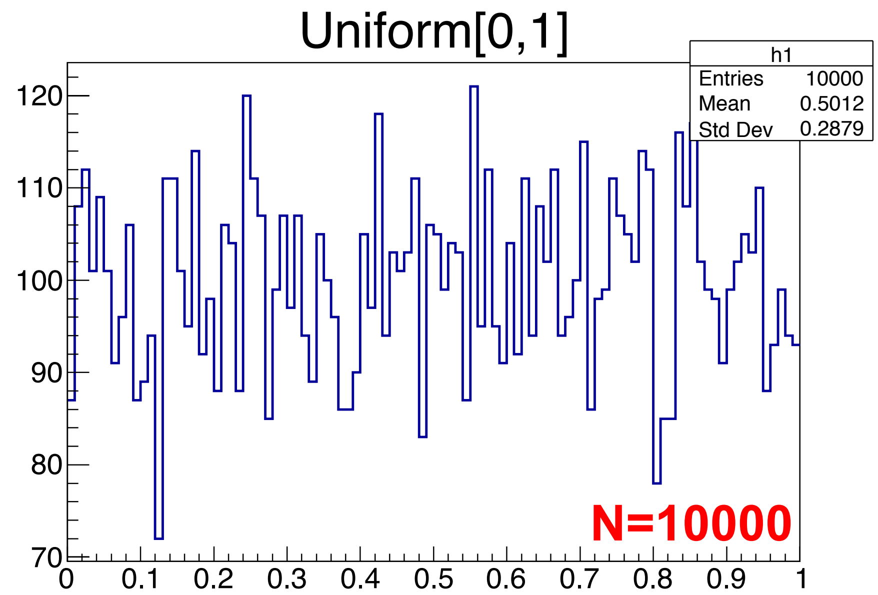
- Target PDF : $p(x) = \lambda e^{-\lambda x}$

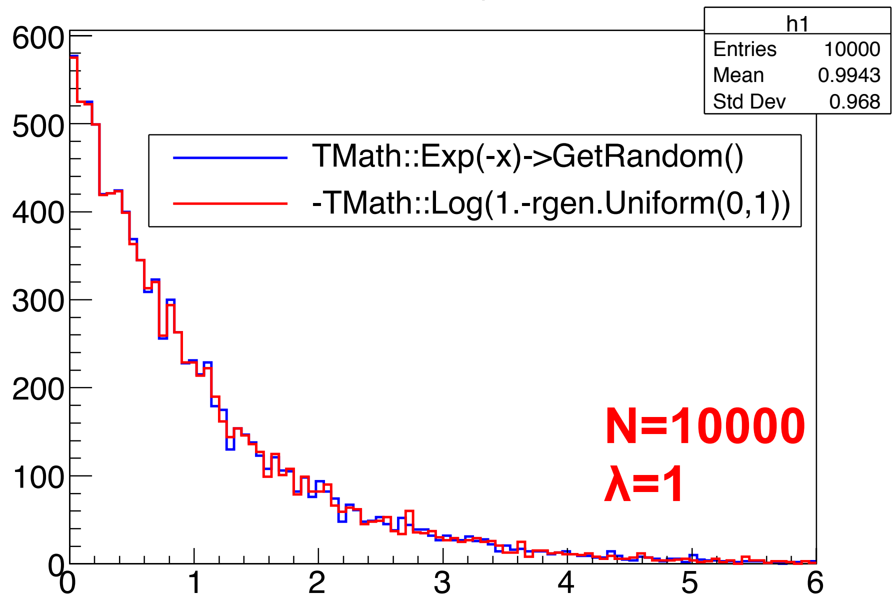
- Substitution:

$$y = \int_0^x \lambda e^{-\lambda x'} dx' = 1 - e^{-\lambda x}$$

$$\Leftrightarrow e^{-\lambda x} = 1 - y$$

$$\boxed{x = -\frac{\ln(1-y)}{\lambda}}$$





Let's see now where

- *Distribution functions*
- *Probabilities*
- *Statistics*

come handy in Monte Carlo methods...

Monte Carlo Integration

- ***Basic idea***: can one calculate integrals of a function f(x) with **x** in **[a,b]** by using *randomnly sampled values of f(x)* ?!
- If so, it would be a ***great simplification*** over traditional methods e.g. Simpson.
- This is where statistics comes in... some useful maths results first...

- For f(x) integrable in [a,b] → $F = \int_a^b f(x)\,dx$

  and Mean Value Theorem for Integrals defines $<f> = \frac{1}{b-a}\int_a^b f(x)\,dx$

- On the other hand, statistics says $<f(X)> = \int_{-\infty}^{+\infty} p(X) f(X)\,dX$ and with p(X) a ***uniform distribution***

$$p(X) = \frac{1}{b-a} \Longrightarrow <f> = \frac{1}{b-a}\int_a^b f(x)\,dx$$

...<f> from sampled f(x)...

Monte Carlo Integration (cont.)

- If $\quad < f >= \dfrac{1}{b-a}\int_a^b f(x)\,dx$ then $\int_a^b f(x)\,dx = (b-a) < f >$ and one also has $< f >= \sum_{i=1}^{N} f(x_i)/N$ with ***x_i uniform sampled set of [a,b]***
- Therefore

$$F = \int_a^b f(x)\,dx = \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i)$$

Value estimate

- To estimate the error in the estimated F, get the standard deviation !

$$\sigma_F = (b-a)\sigma_{<f>} = (b-a)\frac{\sigma_f}{\sqrt{N}}$$

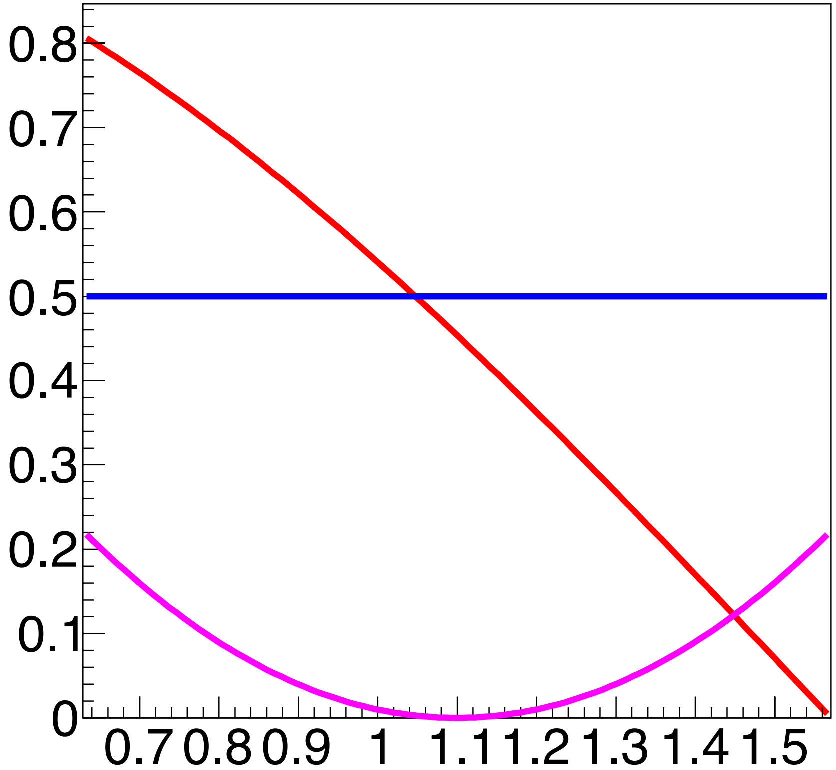
CLT result...

and since $\quad \sigma_f = \sqrt{< f^2 > - < f >^2}$

$$\sigma_F = \frac{(b-a)}{\sqrt{N}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}[f(x_i)]^2 - \left[\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right]^2}$$

Error estimate

Monte Carlo Integration - Example



Three test cases ($x \in [0.2\pi, 0.5\pi]$):

$f(x) = \cos(x)$ **I = 0.412215**

$f(x) = 0.5$ **I = 0.471239**

$f(x) = (x - 0.35\pi)^2$ **I = 0.069764**

With 1000 sample U(0.2π,0.5π), one gets the approximations

$$\int_{0.2\pi}^{0.5\pi} \cos(x)\,dx \sim 0.413828 \pm 0.007099$$

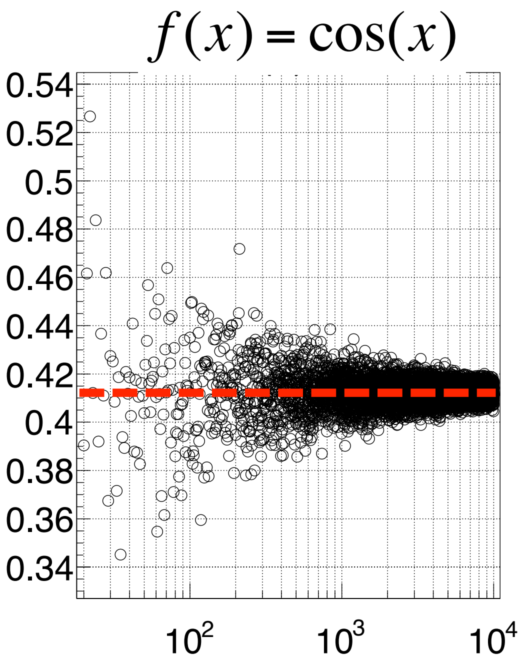
$$\int_{0.2\pi}^{0.5\pi} 0.5\,dx \sim 0.471239 \pm 0.0$$

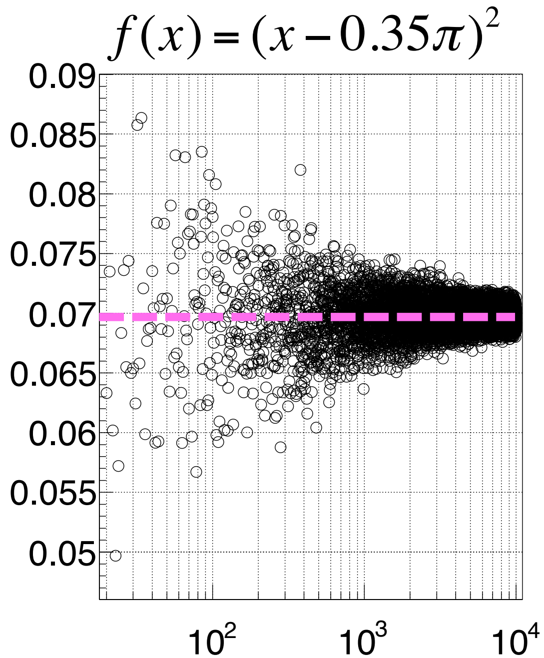
$$\int_{0.2\pi}^{0.5\pi} (x - 0.35\pi)^2\,dx \sim 0.070060 \pm 0.001982$$

Immediate notes:

- A ***1.7%*** vs ***2.8%*** error is an implicit evidence of how *wasting random samples* where *f(x)* is negligible hurts....*(wait a couple of slides more)*...

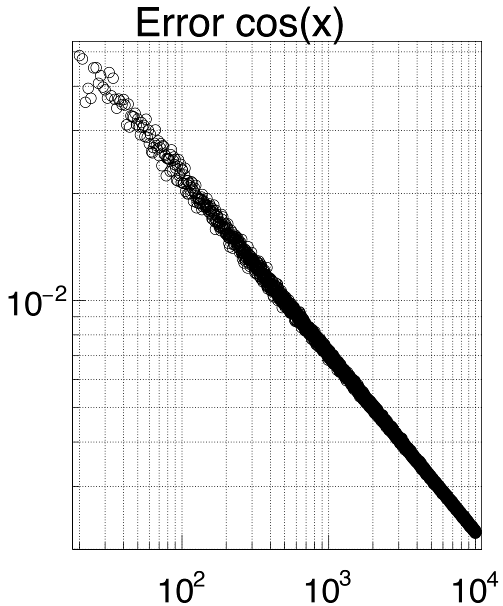
Monte Carlo Integration – Example (cont.)

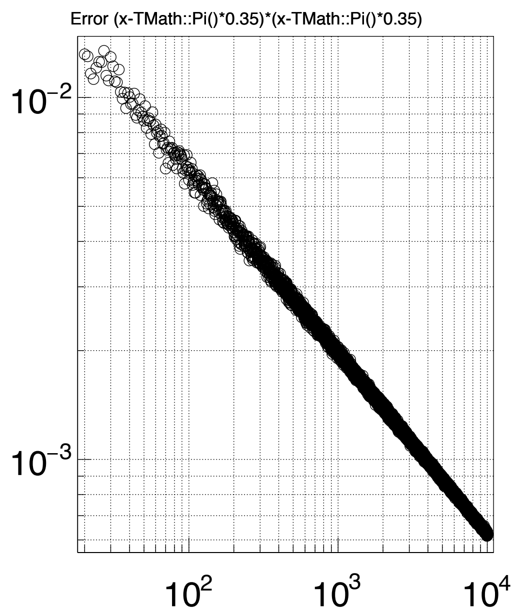




Simple experiment:

- Take 20 to 10000 elements sample from U(0.2π,0.5π).
- Calculate the MC integral and the error using each sample.





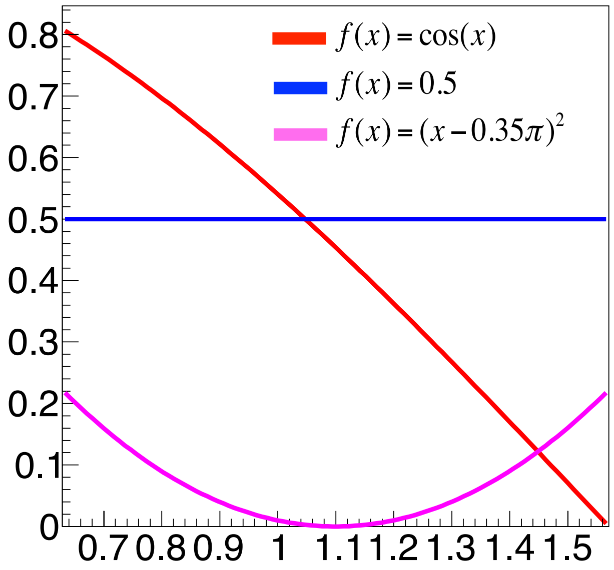
Simple outcome:

- Difficult to get a clearly converged value with reasonable N.
- Error scales with $1/\sqrt{N}$ as anticipated...

Let’s see now two notable tecniques of ***Reduction Variance***...

- *Importance sampling*
- *Acceptance-rejection*

Importance sampling – basic idea



- It is indeed clear that one is wasting random numbers in U(0.2π,0.5π) when "*throwing blindly*" if the integrand function is not *reasonably constant* in the interval.

→ Where f(x) is small the contribution to the integral is also small !!!

✧ ***Innovation***: define a new PDF function p(x) that, scaled, is "close" to f(x) and then use f(x)/p(x) as integrand instead !.....sort of….let's see...

$$F=\int_a^b f(x)\,dx=\int_a^b \boxed{\frac{f(x)}{p(x)}}\,p(x)\,dx$$

$$\int_a^b p(x)\,dx=1$$

$$F=\left\langle\frac{f}{p}\right\rangle_{p(x)}=\int_a^b\left(\frac{f(x)}{p(x)}\right)p(x)\,dx$$

Importance sampling - implementation

- If one defines a new PDF function p(x) that, scaled, is "close" to f(x) and then use f(x)/p(x) as integrand instead, all samples are "worth it".
- Introducing the PDF function p(x), "***map***" to Uniform(0,1) and job done !

$$F = \left\langle \frac{f}{p} \right\rangle_{p(x)} = \int_a^b \left(\frac{f(x)}{p(x)} \right) p(x)\,dx \xrightarrow{dy = p(x)dx} F = \int_0^1 \left(\frac{f(x(y))}{p(x(y))} \right) dy$$

- Therefore, one finally obtains

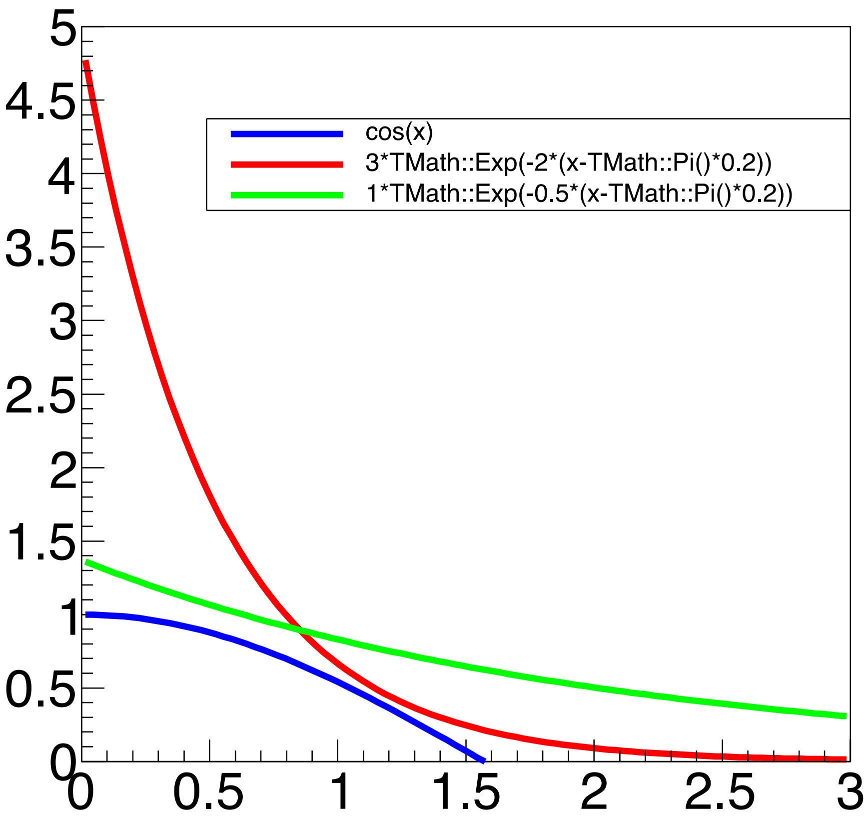
$$F = \int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{p(x)} p(x)\,dx = \int_0^1 \frac{f(x(y))}{p(x(y))} dy = \left\langle \frac{f}{p} \right\rangle_y = \frac{1}{N} \sum_{i=1}^{N} \frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}$$

- Ideal p(x) would be f(x) itself but one would need value of $\int_a^b f(x)\,dx$ to normalise it as a PDF....(-;

Importance sampling – cos(x) example

$$F = \int_a^b f(x)\,dx = \frac{1}{N}\sum_{i=1}^{N} \frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}$$

✧ An exponential function is a handy “invertible” and integrable function over the interval $x \in \left[0.2\pi, 0.5\pi\right]$



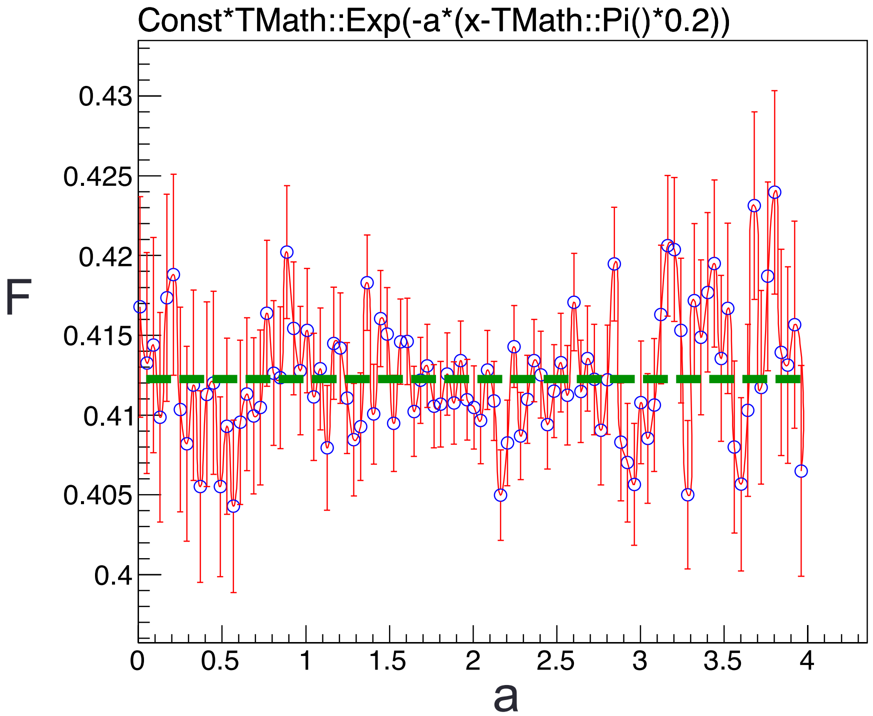
✧ If one is to give less relevance to abcissa where cos(x) is small, can you determine ***which of the curves to the left*** should be used to get the lowest integral error & best integral estimate ?!

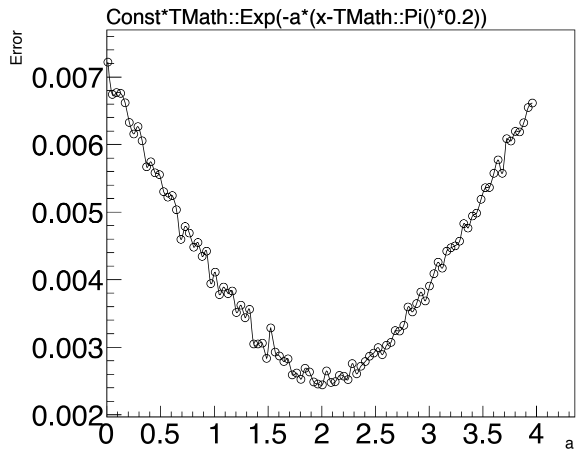
$$\sigma_F = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}\right]^2 - \left[\frac{1}{N}\sum_{i=1}^{N}\frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}\right]^2}$$

Importance sampling – cos(x) example

$$F = \int_a^b f(x)\,dx = \frac{1}{N}\sum_{i=1}^{N}\frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}$$

$$\sigma_F = \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}\right]^2 - \left[\frac{1}{N}\sum_{i=1}^{N}\frac{f\left[x(y_i)\right]}{p\left[x(y_i)\right]}\right]^2}$$





- Too “flat” exponential ~Uniform sampling.
- Too “sloped” means f(x)/p(x) can become too high (p(x)<<f(x)) → error increases…

Acception-Rejection – basic idea

- Imagine ROOT doesn't exist and one needs to get a ***sample*** following the statistics of ***any*** given ***distribution function*** p(x) ?!

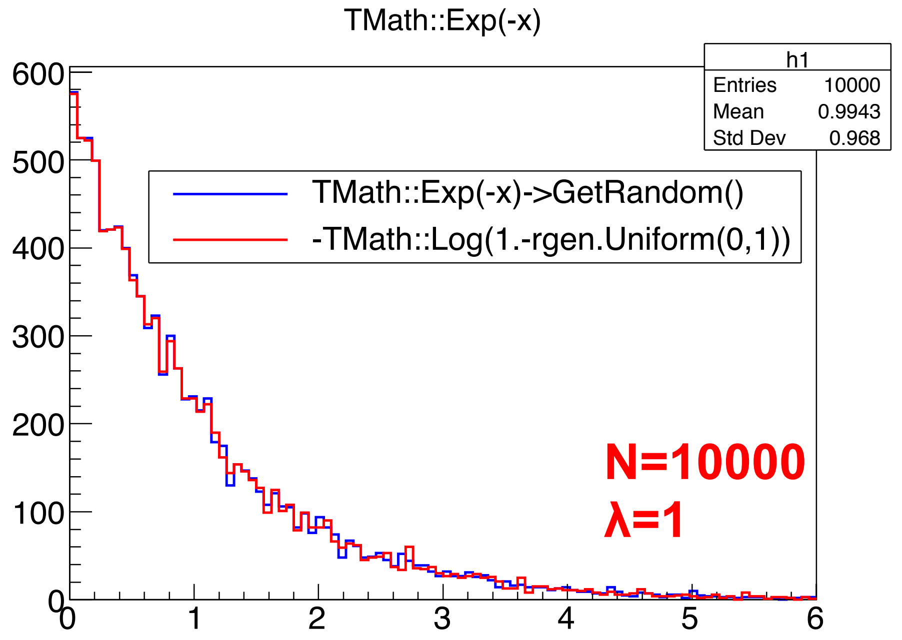
- If f(x) is easily *integrable* and *invertible* e.g. exponential, one saw before that:

- Target PDF : $p(x) = \lambda e^{-\lambda x}$

$$y = \int_0^x \lambda e^{-\lambda x'} dx' = 1 - e^{-\lambda x}$$

- Substitution: $\boxed{x = -\frac{\ln(1-y)}{\lambda}}$

with ***y*** following Uniform(0,1).



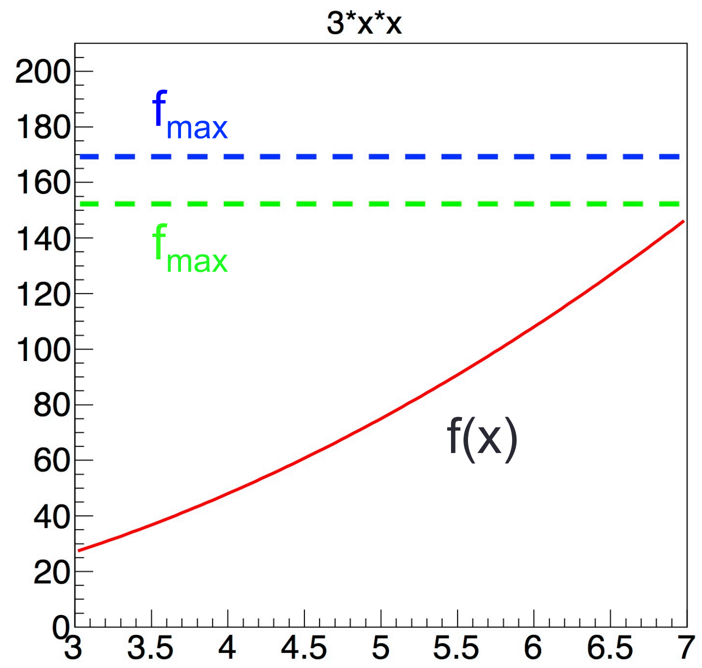
✧ But what if we ***CAN'T*** easily get the Cummulative Dist. Function and/or *invert* it to get x(y) ???

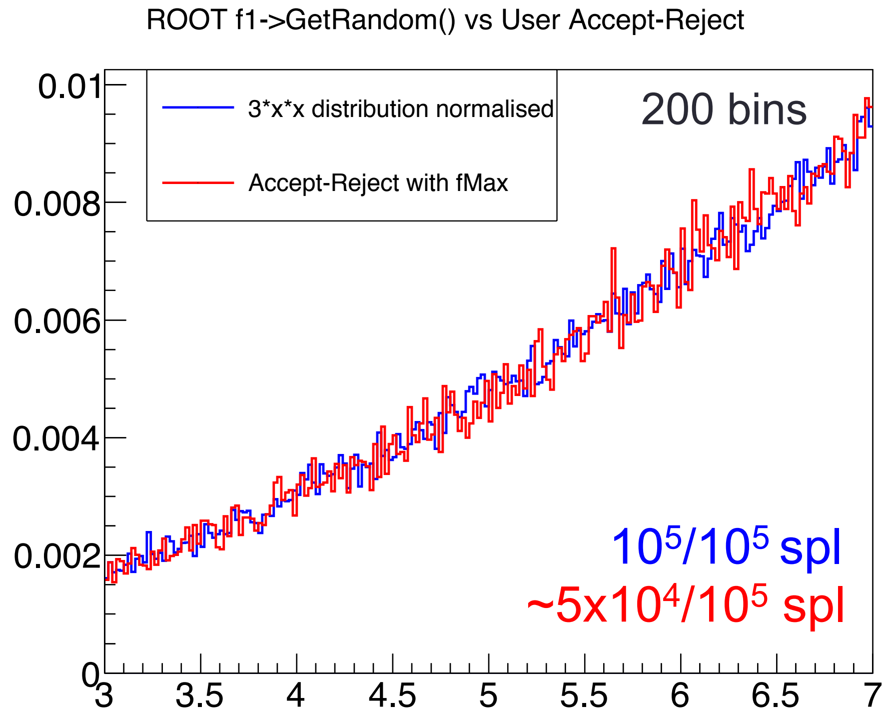
✧ ***Acceptance-Rejection*** to the rescue !

Acception-Rejection – implementation

- ***Aim***: generate a sample of values x_i in the interval [a,b] according to distribution function f(x).

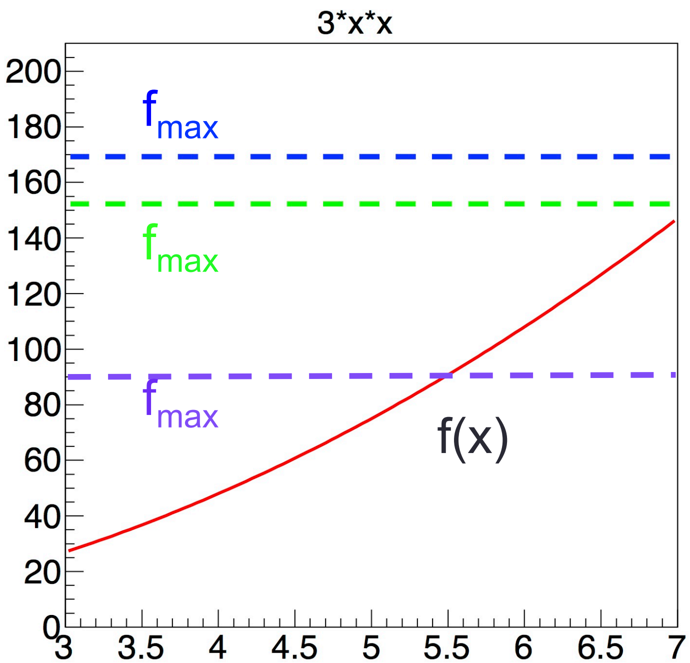
① Assign a "*capping*" value f_{max} ***larger than*** f(x) in [a,b]
② Generate a uniform random (x_R) in [a,b].
③ Compute $f(x_R)$ and $u_f=f(x_R)/f_{max}$
④ Generate a Uniform(0,1) random u_R.
⑤ If $u_R \le u_f$ then ***ACCEPT*** the value x_R.

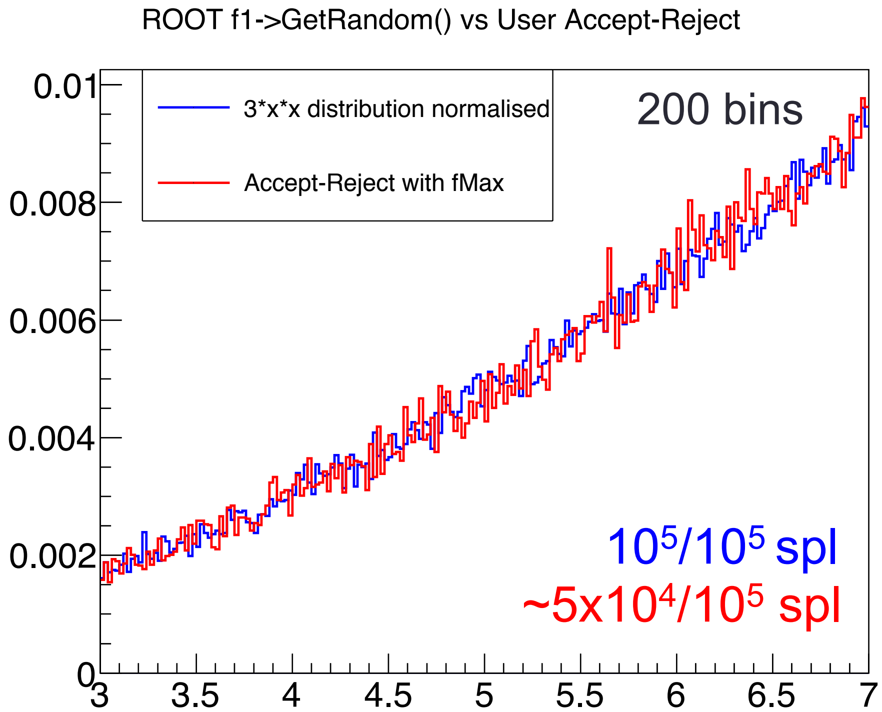




Acception-Rejection – implementation

- ✓ Can you guess what the histogram would look like if one used a min(f(x)) < fmax < Max(f(x))





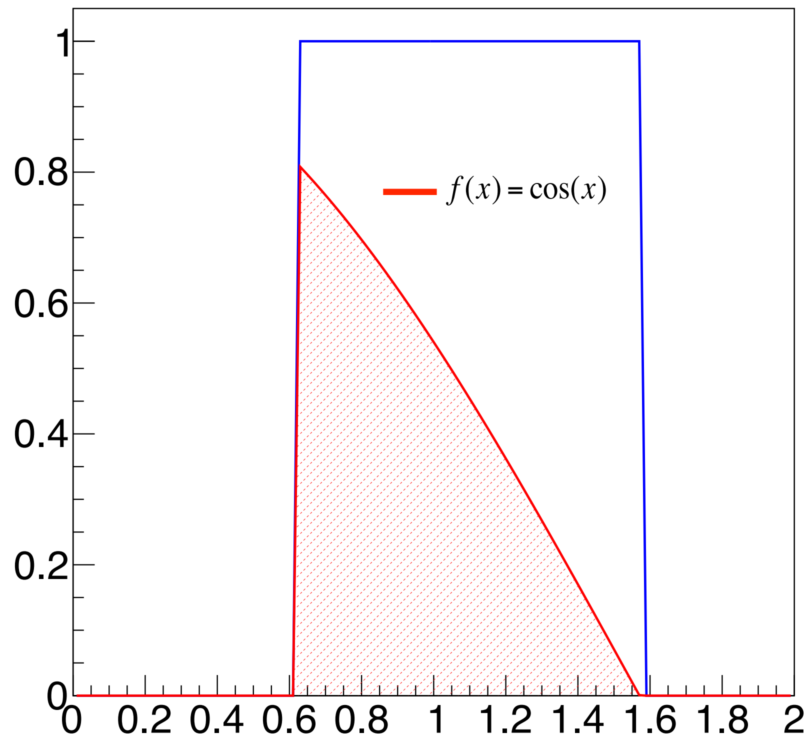
Acception-Rejection – Integration

- ***Idea***: use acceptance-rejection type method to value (accept) only 2D points that define the "area" defined by the integral ! (assume f(x)>0)

① Make envelope [a,b] x f_{max} with area **$A=(b-a)f_{max}$**.
② Generate a uniform random (**x_R**) in [a,b] and a uniform random (**f_R**) in [0,f_{max}].
③ Count number of events (***N_R***) with **$f_R \leq f(x_R)$**.

④ Integral $$I=(b-a)f_{\max}\frac{N_R}{N}$$

⑤ Error $$\sigma_I=\frac{(b-a)f_{\max}}{N}\sqrt{N_R\left(1-\frac{N_R}{N}\right)}$$



$$\int_{0.2\pi}^{0.5\pi}\cos(x)\,dx \sim 0.4316558 \pm 0.0148492$$

(458/1000 samples)

Using now fMax=max(f(x))

$$\int_{0.2\pi}^{0.5\pi}\cos(x)\,dx \sim 0.419364 \pm 0.0119954$$

(533/1000 samples)

Acception-Rejection – one step further

- ***Innovation***: like importance sampling, we mustn't waste samples → Auxiliary function to "cap" the acc-rej !!!

① Define ***q(x)*** in [a,b], easily *integrable* / *invertible* and with $q(x) \geq f(x)$.

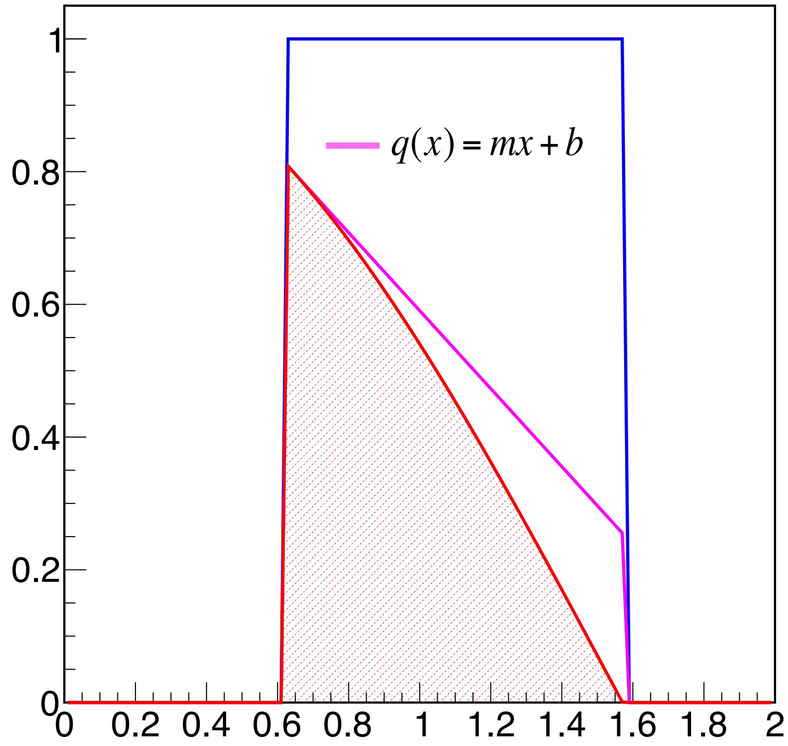
② Normalise q(x) to get a PDF ***p(x)*** i.e. **p(x)=Cq(x)**.

③ Get uniform random (**x_R**) in [a,b] and random (**f_R**) from p(x).

④ Count events (***N_R***) with **$f_R \leq f(x_R)$**.

$$I = Area\_f_{Aux} \frac{N_R}{N}$$

$$\sigma_I = \frac{Area\_f_{Aux}}{N} \sqrt{N_R \left(1 - \frac{N_R}{N}\right)}$$



$$\int_{0.2\pi}^{0.5\pi} \cos(x)\,dx \sim 0.419364 \pm 0.0119954$$

(533/1000 samples)

Using importance sampling

$$\int_{0.2\pi}^{0.5\pi} \cos(x)\,dx \sim 0.414679 \pm 0.0059977$$

(827/1000 samples)

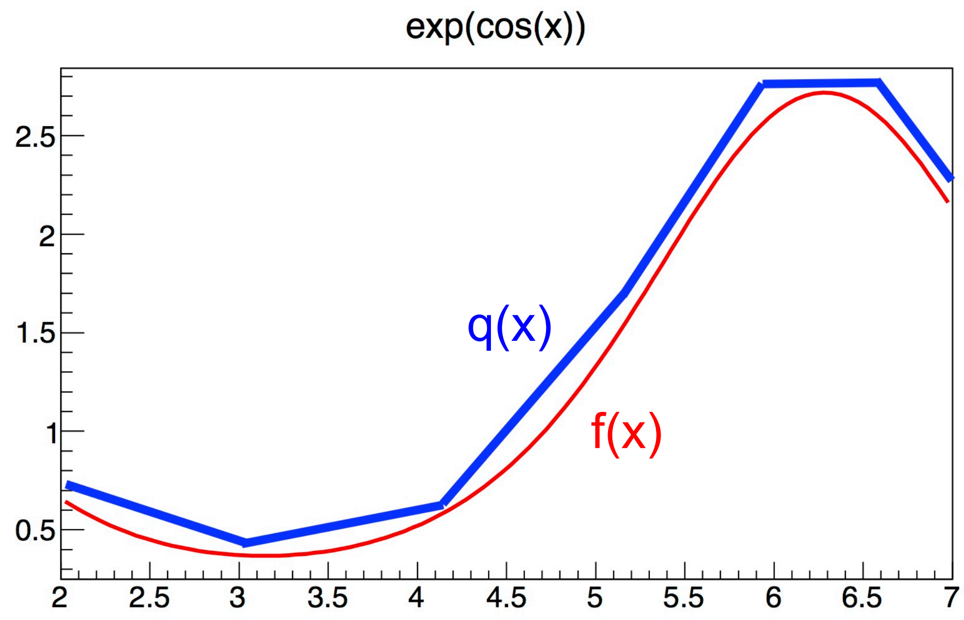
Acception-Rejection – final thoughts

- ✓ If the auxiliary function q(x) to "cap" the function f(x) to be integrated is not "*close enough*" to f(x), the fraction of random numbers accepted might not be "*high enough*" → *larger uncertainty* in $\int_a^b f(x)\,dx$

- ✓ The *efficiency* of acception-rejection does not rely on having q(x) defined by a single analytical function ! It can be a "*step-pyramid*" piecewise shape...

→ The smaller sized are the steps the higher is the overall efficiency

→ ...but added computations mean longer time to calculate the integral...



Integrals with singularities

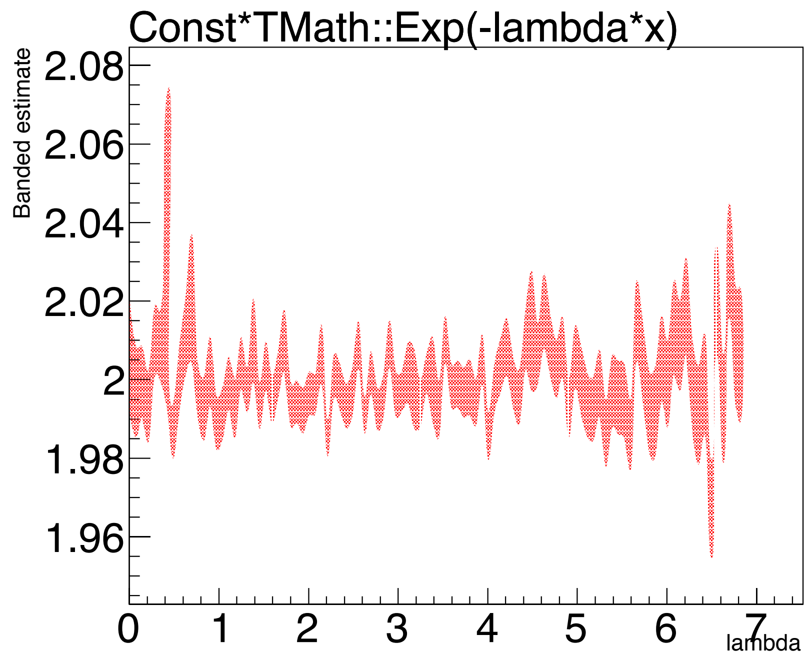
- ✓ One knows all to well that some integrals can have singularities in the integrand or have unbounded integration range → ***Improper integrals***

$$\int_0^{+\infty} e^{-x}\,dx \quad \textit{(I)} \qquad \text{or} \qquad \int_0^1 \frac{1}{\sqrt{x}}\,dx \quad \textit{(II)}$$

- ✓ For improper integrals of second kind ***(II)***, one can always get support from ***importance sampling*** !

$$F = \int_0^1 \frac{1}{\sqrt{x}}\,dx = \int_0^1 \underbrace{\left[\frac{e^{\lambda x}}{a\sqrt{x}}\right]}_{\frac{f(x)}{p(x)}} \underbrace{\left(ae^{-\lambda x}\right)dx}_{dy}$$

$$1 = \int_0^1 ae^{-\lambda x}\,dx \Leftrightarrow a = \frac{\lambda}{1-e^{-\lambda}}$$



Integrals with singularities (cont.)

✓ If it is the integration domain that is unbounded, nothing like a good variable substitution (that you know already from Calculus)

(I) $$\int_0^{+\infty} e^{-x}\,dx \underset{t=\frac{1}{x+1}}{\Leftrightarrow} \int_1^0 -e^{-\left(\frac{1}{t}-1\right)}\frac{1}{t^2}\,dt = e\int_0^1 \frac{e^{-\frac{1}{t}}}{t^2}\,dt$$

INTEGRAL =1

Integral=1.00312 ; **Error**=0.0158558 ***N=1000 (no importance sampl.)***

Integral=0.998234 ; **Error**=0.00158355 ***N=100,000 (no importance sampl.)***

Multi-dimensional Integrals

- ✓ Multi-dimensional spaces are common in physics e.g. Kinetic theory with 3d-space, 3d-velocity time dependent distribution function of particles...

$$\mathrm{I}=\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \cdots \int_{a_n}^{b_n} dx_n f(x_1, x_2, \cdots, x_n)$$

- ✓ **Conventional approach** (*Brute-force*)

$$x_i = a_i + u_i^R (b_i - a_i) \quad \text{with} \quad u_i^R \in [0,1] \longleftarrow$$

n-independent Uniform(0,1) random values

$$I = \frac{\prod_{j=1}^{n} (b_j - a_j)}{N} \sum_{i=1}^{N} f(x_1, x_2, \cdots, x_n)\Big|_i$$

N - the total number of samples
From the multi-dimensional function

Multi-dimensional Integrals - final remarks

- ✓ Much like 1D integrals, one is not restricted to uniform sampling...

- ✓ **Importance smapling**

$$I=\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \cdots \int_{a_n}^{b_n} dx_n \frac{f(x_1,x_2,\cdots,x_n)}{p(x_1,x_2,\cdots,x_n)} p(x_1,x_2,\cdots,x_n) = \left\langle \frac{f(x_1,x_2,\cdots,x_n)}{p(x_1,x_2,\cdots,x_n)} \right\rangle_p$$

Final remarks

- ✓ Above ~5 dimensions, MC methods become more competitive since error always scales as $1/N^{1/2}$ whereas "analytical" trapezoidal methods gradually deteriorate i.e. error ~ $1/N^{2/k}$ where k is the dimensionality.

- ✓ Contrary to Trapezoidal, Simpson,…, MC methods rely on random samples points with the possibility to easily fine-tune the sample distribution → ***Flexibility***