UNIVERSITÀ DEGLI STUDI DI TRIESTE

CORSO DI LAUREA MAGISTRALE IN FISICA DELLA MATERIA

Laboratorio di Fisica Computazionale

Unit VI
Again on calssicsl numerical integration: the Gauss-Legendre quadrature.
Gauusian distribution and Central Limit Theorem

DAVIDE BERNOCCO
Matricola s285296

PROGRAMMING LANGUAGE USED:
Python

Contents

1 1D classical integration: "Gauss-Legendre quadrature"

This chapter shall conclude the introductory overview on the classical numerical integration methods. We are talking about the *deterministic* "Gauss - Legendre quadrature" algorithm, which exploits the approximation:

$$\int_a^b f(x)dx = \int W(x)F(x)dx \approx \sum_{i=1}^{N} w_i F(x_i) \tag{1}$$

With an increasing degree of precision, for $W(x) = 1$ and $[a, b] = [-1, 1]$, the points x_i can be chosen as the N roots of the N-th Legendre polynomial, and the associated weights w_i will be defined after them through a certain algebraic relation. The interval extreme point a and b can then be arbitrary moved having just the foresight to adjust the relevant points and their weights by a proper normalization.

The testing ground will be the well known integral $I = \int_0^1 e^x dx = e - 1$. With just a $N = 2$ order algorithm we already get a satisfactory result, with an actual error of:

$$\Delta_{N=2} = |I - I_{num}| = 0.0004$$

1.1 Comparison with other deterministic algorithms

We already studied the dependence on N of the numerical error for both the *trapezoidal* and *Simpson* methods. What about the *Gauss Legendre quadrature*? As usual,we can get a rough sense of it plotting together the *actual errors*:

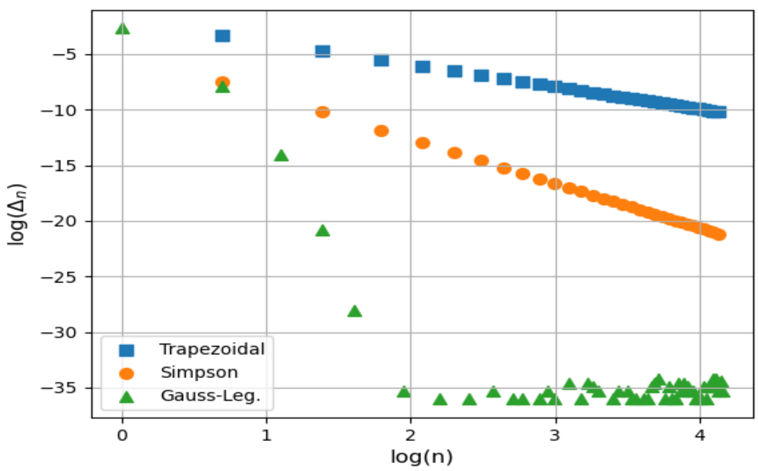

Fig 1.1 Actual errors for trapezoidal, Simpson and Gauss-Legendre algorithms, log-log scale

As one could have suspected from the previous example, the accuracy for the quadrature algorithm is way larger if compared to the others.

However, it seems our qualitative estimations about this powerful method cannot be brought further than $N = 7$: indeed, at that stage, the magnitude of the considered variable Δ_N becomes so small that it gets comparable with the machine precision for *float64* type number ($\ln 10^{-16} \approx 10^{-35}$), making the presence of numerical roundoff errors dominant!

2 (Pseudo)Random numbers with gaussian distribution: the "central limit theorem"

In this second section we are going to exploit the *central limit theorem (CLT)* in order to generate a pseudo-random variable with gaussian distribution.
What the theorem states is[1]:

Let $x_1, x_2, ..., x_N$ be a sequence of *independent and identically distributed* random variables having a distribution with *expected value* given by μ and *finite variance* given by σ^2. If we are interested in the sample average

$\bar{X}_N = \frac{x_1+x_2+...+x_N}{N}$

by the law of large numbers, for N large enough, the distribution of $\bar{X}_N$ gets arbitrarily close to the normal distribution with mean $\mu = \langle x \rangle$ and variance $\sigma^2/N = (\langle x^2 \rangle - \langle x \rangle^2)/N$.

Moreover, given $z_N = \frac{\bar{X}_N - \mu}{\sigma}$, it can be proven that must hold $\langle z_N^4 \rangle \approx 3\langle z_N^2 \rangle^2$.

We are now going to analyze three examples with three different starting distributions.

2.1 From random numbers with uniform distribution

Let's start from the case of random variables with uniform distribution

$$p(x) = \begin{cases} 1/2 & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

We see it satisfy the hypothesis of the CLT since both the expected value and the variance exist and are finite, making the theorem actually apllicable:

$$\mu_{unif} = \langle x \rangle = 0$$
$$\sigma^2_{unif} = 1/3$$

We thus expect a gaussian distribution with $\mu = 0$ and $\sigma = \sigma_{unif}/\sqrt{N}$ can be obtained as the limit distribution of the averages of uniformly distributed variables.
Here follows what we found in a simulation with $n = 10000$ samples of length $N = 500$ each:

- Numerical average of averages ≈ 0.0002 $(\mu = 0)$
- Numerical standard deviation of averages ≈ 0.0259 $(\sigma/\sqrt{N} = 0.0258)$
- $\langle z_N^4 \rangle \approx 3.0263$ $3\langle z_N^2 \rangle^2 \approx 3.0000$

This accordance with the expected result can be seen graphically as well IN *Fig. 2.1*:

2.2 From random numbers with exponential distribution

We are now going to retrace the same consideration, taking as a reference averages built from the exponential distribution:

$$p(x) = \begin{cases} e^{-x} & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In this case the expected value and variance do exist too and correspond to:

$$\mu_{exp} = \langle x \rangle = 1$$
$$\sigma^2_{exp} = 1$$

[1] "Central limit theorem", Wikipedia

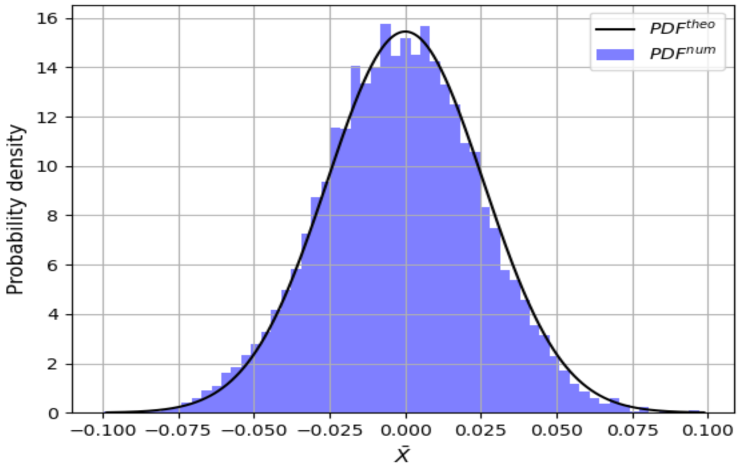

Fig. 2.1 Distribution of the averages from uniformly distributed variables

We expect a gaussian distribution to emerge as the limit distribution of the averages. What we obtain with the same input parameters as before, seems to confirm the hypothesis:

- Numerical average of averages ≈ 0.9997 $\quad (\mu = 1)$
- Numerical standard deviation of averages ≈ 0.0447 $\quad (\sigma/\sqrt{N} = 0.0447)$
- $\langle z_N^4 \rangle \approx 2.9735 \quad 3\langle z_N^2 \rangle^2 \approx 3.0000$

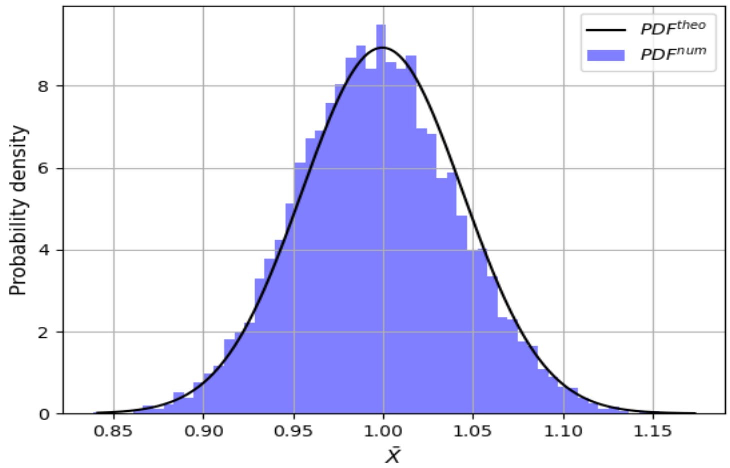

Fig 2.2 Distribution of the averages from exponentially distributed variables

2.3 From random numbers with Cauchy-Lorentz distribution

As final example we will consider the sample mean of random variables with the following distribution, specifically with $\gamma = 1$ and $x_0 = 0$:

$$p(x) = \frac{\gamma}{\pi} \frac{1}{(x - x_0) + \gamma^2} \qquad \forall x \in \mathbb{R} \tag{4}$$

Unlike the two previous cases, here we cannot even talk about expected value and variance, since they are not defined! As a consequence, the CLT as initially stated cannot be applied.
Indeed, if $x_1, x_2, ..., x_N$ are IID samples from the standard Cauchy distribution, then their sample mean $\bar{X} = \frac{1}{n}\sum_i x_i$ will also be standard Cauchy distributed and so does not converge to a mean.
More generally, if $x_1, x_2, ..., x_N$ are independent and Cauchy distributed with location parameters $x_{0,1}, x_{0,2}, ..., x_{0,N}$ and scales $\gamma_1, \gamma_2, ..., \gamma_N$, and $a_1, a_2, ..., a_N$ are real numbers, then $\sum_i a_i x_i$ is Cauchy distributed with location $\sum_i a_i x_{0,i}$ and scale $\sum_i |a_i|\gamma_i$ [2]. In our case, $\gamma_i = 1$, $x_{0,i} = 0$, $a_i = 1/N$ for all i and so we expect for the sample average distribution a final scale of 1 and location 0.

Since there is no point in finding a mean and a variance for this distribution, it is at this main characteristics of location and scale we are looking for. Numerically they can be respectively estimated through different quantities; we chose here the *mean* and the *half interquartile range*.
For a simulation with the usual input parameters, we get satisfying results:

- Median $\approx 0.0045 \quad (x_0 = 0)$
- Half the InterQuartile Range $\approx 0.9997 \quad (\gamma = 1)$

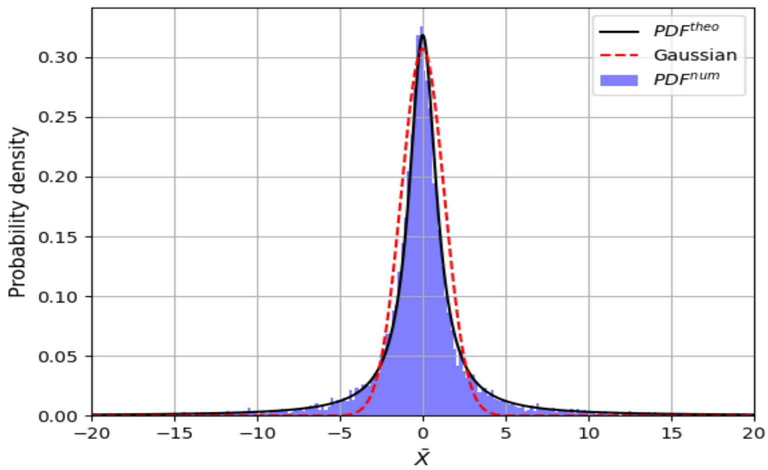

Fig 2.3 Distribution of the averages from Cauchy-Lorentz distributed variables. A gaussian curve is plotted as a comparison (red dashed)

[2]"Cauchy distribution", Wikipedia