Statistical Methods in Physics (14P058)

Prof. Federico Sánchez (federico.sancheznieto@unige.ch)
Dr. Hepeng Yao (hepeng.yao@unige.ch)
Dr. Knut Zoch (knut.zoch@unige.ch)

Final project I – One-dimensional Harmonic Oscillator

Submission due on: 3 February 2023

The harmonic oscillator is a widely used model in both classical and quantum physics. For low temperature atomic physics, gases are usually trapped by a harmonic potential formed by laser beams or magnetic fields.

In this project, we will study the statistics of a one-dimensional ideal gas in the presence of a harmonic trap. Consider a system which contains a large number of ^{40}K atoms with mass $m = 40\, m_0$, where m_0 is the atomic mass unit. The potential energy E_ω follows the probability density function

$$f(E_\omega) = \frac{A}{\sqrt{E_\omega}} e^{-\beta E_\omega} \tag{1}$$

where $\beta = 1/(k_B T)$ denotes the inverse energy scale related to the temperature T, and A is a pre-factor related to the properties of the system. Let's consider $T = 30\,\mathrm{K}$ and $A = \sqrt{\beta/\pi}$.

Task 1: *Basic statistical properties*

Equation (1) fulfils the properties of a probability distribution function. The first task is to assess basic statistical properties of the distribution with the help of Monte Carlo (MC) processes.

a) Demonstrate that Eq. (1) is a probability density function.

   Hint: You can answer this question either analytically or numerically.

b) Build a MC process to generate this probability density function for a large number of data points, $N \gg 1$. Fill your results in a histogram, plot them and verify their correctness by comparing them with the curve of the analytical form in Eq. (1).

   Hint: In principle, the value ϵ can take the range $[0, +\infty)$. Here, you should choose a cutoff according to the shape of Eq. (1).

c) According to the equipartition theorem, the expectation value of E_ω should follows

$$\mathbb{E}\left[E_\omega\right] = \frac{1}{2} k_B T \tag{2}$$

   Compute the sample mean $\langle E_\omega \rangle$ and compare with the expectation value.

d) Use another Monte Carlo method to compute $\langle E_\omega \rangle$ and compare the efficiency with the previous one.

e) Compute the variance, skewness and kurtosis of the distribution and comment.

Task 2: *Convergence*

If the experiment is repeated N_{exp} times, it can be seen that the sample means $\langle E_\omega \rangle$ of the experiments are distributed according to a Gaussian distribution. In the following sub-tasks, we will look at the convergence behaviour of $\langle E_\omega \rangle$. Each of the N_{exp} experiments still contains $N \gg 1$ data points generated with the MC process.

a) Take $N_{exp} = 1$ and show the law of large numbers based on the data points you generated.

b) Given one experiment composed of many measurements, compute the variance of $x = \sqrt{2E_\omega/m\omega^2}$, where ω is a constant value. Compare it with what is expected from the analytical formula and comment.

c) Take $N_{exp} \gg 1$ and show the validity of the central limit theorem for $\langle E_\omega \rangle$. Compute the variance of $\langle E_\omega \rangle$ and compare it with the value you expect.

Task 3: χ^2 *distribution*

By filling the numbers generated from the MC process into a histogram, we can study the χ^2 properties of the distribution. For now, we fix $N_{exp} = 1$ and $N \gg 1$, according to the probability density function in Eq. (1). Fill the obtained E_ω values in a histogram.

a) What is the expected p.d.f. for the number of entries per bin?

b) Show that the χ^2 of the obtained entries per bin follows a χ^2 distribution. Instead of calculating the empirical mean value per bin from the experiments, use the nominal value from the p.d.f. formula per bin as the mean value per bin.

c) How does the χ^2 distribution change with the number of bins?

Task 4: *Parameter estimation*

We will now use three different methods to perform parameter estimation and compare the results from the different methods. Let's again start with one single experiment ($N_{exp} = 1$) with $N \gg 1$ data points according to the probability density function in Eq. (1).

a) Compute the log likelihood $\ln L(E_{\omega,1}, E_{\omega,2}, ..., E_{\omega,N} \,|\, T)$ at a given T. Then, use the maximum likelihood method to estimate T and its variance.

b) Compute the goodness of fit at various T and use the least squares method to estimate T and its variance.

c) The parameter T can also be estimated by computing the sample mean $\langle E_\omega \rangle$ of Eq. (1) with the MC integral. Perform this estimation of T and its corresponding variance.

d) Compare the results from the three questions above and discuss them.

Task 5: *Hypothesis testing*

Generate $N_{exp} = 200$ experiments and combine them into pairs, that is, forming a total of 100 pairs of experiments. Now let's assume we obtained these datasets in actual experiments and we would like to – for each pair – test the hypothesis that the two sets of experimental data originate from the same distribution. Set a threshold for the hypothesis testing and perform Kolmogorov tests for all dataset pairs. Then, comment on your outcome with respect to the threshold you set.

Task 6: *Bonus question*

Generate N_{exp} number of experiments with N data in each. Compute the exact Fisher information with the equation

$$I_X(T) = \mathbb{E}\left[-\frac{\partial^2 \ln\, L(X|T)}{\partial^2 T}\bigg|_{T=T_0} \right] \tag{3}$$

Compare your results with the variance $V(\hat{T} - T)$ and discuss.

N.B. This task is not required for your report, but if you hand in a solution, it can be accredited for bonus points.