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 – Three-dimensional Fermi gas at low temperature

Submission due on: 24 January 2022

Fermions are particles that have half-integer spin and obey Fermi–Dirac statistics. Examples include electrons, protons and neutrons. This model can be accurately applied to various kinds of systems which contain a large number of fermions. Some key examples are the behaviour of charge carriers in metal, ultracold Fermi seas generated by quantum optics techniques, neutrons in a neutron star, and electrons in a white dwarf.

In this project, we will study the statistics of three-dimensional Fermi gas at low temperature. Consider a three-dimensional box system which contains a large number of ^{40}K atoms with mass $m = 40\,m_0$, where m_0 is the atomic mass unit. At low enough temperatures and a particle density of $n = 1$, the energy ϵ of the atoms follows the Fermi-Dirac distribution

$$f(\epsilon) = A \frac{\sqrt{\epsilon}}{\exp\left(\beta(\epsilon - \mu)\right) + 1} \tag{1}$$

where $\beta = 1/(k_B T)$ denotes the inverse energy scale related to the temperature T, μ is the chemical potential of the system, and A is a pre-factor related to the properties of the system. Let's consider $T = 3\,\mathrm{K}$ and $\mu = 30\,k_B T$. Under these conditions, we have $A = 3.5 \times 10^{31}$.

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) At low temperatures, the fermions show quantum behavior, which means that the mean energy per particle follows

$$\mathbb{E}\left[\epsilon\right] \simeq \frac{3}{5} n \mu. \tag{2}$$

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

    d) Calculate $\mathbb{E}\left[\epsilon\right]/k_B T$. Compare it with what you learnt from exercise sheet 5 and comment.

    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 $\langle \epsilon \rangle$ of the experiments are distributed according to a Gaussian distribution. In the following sub-tasks, we will look at the convergence behaviour of $\langle \epsilon \rangle$. 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) Take $N_{exp} \gg 1$ and show the validity of the central limit theorem for $\langle \epsilon \rangle$.

    c) Compute the variance of $\langle \epsilon \rangle$ 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 ϵ 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 $ln\ L(x_1, x_2, ..., x_N \,|\, \mu)$ at a given μ. Then, use the maximum likelihood method to estimate μ and its variance.

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

    c) The parameter μ can also be estimated by computing the sample mean $\langle \epsilon \rangle$ of Eq. (1) with the MC integral. Perform this estimation of μ 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: Measurement with systematic errors*

In an actual measurement, the obtained data may come with systematic errors. Let's simulate this effect by generating $N = 10^4$ data points which follow Eq. (1). We assume that there is an experimental error in determining the energy values, which follows a Gaussian smearing with standard deviation $\sigma = 2k_BT$. Now, if we repeat the three estimation techniques in Task 4, how is the outcome influenced by this fluctuation? Which of the estimation techniques is more fragile and which one is more robust? Perform the simulation 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.