Statistical Methods for Experimental Physics (Part-I)
Assessment 1

Nicholas Wardle, Aidan Crilly

October 2025

Answer all parts of at least question 1.1. The additional problem (marked with a $*$) is *not for credit* but you can attempt them as practice in solving these types of problems.

Your answers must be submitted in the form of a pdf converted from a .ipynb notebook. You should run the notebook so that all outputs are produced (please be careful not to print information for every sample generated for example) and then you can convert this to a pdf directly inside Juypter. **You must name your submission** Name_CID_A1.pdf. If you prefer to include handwritten answers, you must scan these and include them in the same pdf.

Note, if you cannot convert directly to a pdf, you can first download as a .html file and then convert this to a pdf by printing to pdf in a web browser. Make sure your code is visible in the pdf, as this can cut off in the conversion process (you can orient the page to landscape to help with this).

1 Assessed Problems

1.1 Neutron Time of Flight

The nuclear fusion reaction between a pair of deuterium ions produces a 2.45 MeV neutron - n - via,

$$D + D \to n + \text{He} \tag{1}$$

When these DD reactions happen within a thermal plasma, the fusion neutrons are emitted with a (small) Gaussian spread in energies from Doppler broadening. Nuclear fusion experiments use the energy spread of these neutrons to measure ion temperature. The time it takes for these neutrons to travel a known distance can be used to estimate their energy. This is known as the time of flight (TOF) method:

$$E_n = \frac{1}{2} m_n v_n^2 \ , \tag{2}$$

$$v_n = \frac{d}{t} \tag{3}$$

where d is the detector distance, t is the arrival time, v_n is the speed of the neutron, E_n is the energy of the neutron and m_n is the neutron mass.

We will assume the detector has a Gaussian response function. We define $\phi(X; \mu, \sigma)$ as the Gaussian distribution for variable X with mean μ and standard deviation σ. The detector response function and neutron energy distribution are then given by the following:

$$R_{\text{detector}} = \phi(t; 0, \sigma_t) \ ,$$

$$\frac{dN}{dE} = \phi(E_n(t); E_0, \sigma_E) \ ,$$

where $E_0 = 2.45$ MeV for neutrons from the DD reactions.

NOTE: We haven't covered likelihood functions or maximum likelihood estimation (MLE) in lectures yet, so you will need to either wait until next week to complete parts (ii)-(iv) or read Week 4 of the lecture notes in advance.

i [**3 marks**] The saddlepoint approximation finds a Gaussian approximation of a strongly peaked function of the form $\exp\left[-f(x)\right]$ by expanding about the exponent function about its maxima:

$$g(x) \approx \exp\left[-f(x_0)\right] \exp\left[-\frac{1}{2} f''(x_0)(x - x_0)^2\right]$$

i.e the distribution $g(x)$ is then approximated by a Gaussian with mean x_0 found as the value of x for which $f'(x) = 0$, and variance $\delta^2 = 1/f''(x_0)$.

(a) Use this method to find an expression for the TOF distribution of deuterium fusion neutrons at the detector, $g = \dfrac{dN}{dt}$, using $f = -\ln \dfrac{dN}{dE}$.

(b) The measured TOF data is a convolution of the true TOF distribution with the detector response function, R_{detector}. Using the result of part (a), find the Gaussian approximation to the *measured* TOF signal as a function of time. **Hint:** The convolution of two Gaussians is another Gaussian with a variance equal to the sum of the two variances.

ii [**2 marks**] The residual sum of squares (RSS) function for a set of measured data points y_i is given by,

$$\text{RSS} = \sum_i (y_i - f(x_i))^2,$$

for a model $f(x_i)$ evaluated at the known values x_i. Write down the RSS for the TOF data using the Gaussian approximation derived in part (i). Which parameters are degenerate in this model?

iii [**2 marks**] The value of parameters at which the RSS is minimised provides an estimate of those parameters from data. Using the data provided in `nTOF_data_15m_detector.csv`, implement in Python a minimisation procedure to estimate the energy spread of the neutron spectrum (σ_E) from the TOF data at a fixed detector response width ($\sigma_t = 10$ ns) from the RSS function you derived in part (ii).

You should assume that the detector is positioned at a distance of 15m, and note that the signal data has been normalised such that you do not need an amplitude parameter. **NOTE:** You will need to lookup the neutron mass and remember to convert units of eV to Joules for this question.

iv [**3 marks**] Plot the minimum RSS estimate of the energy spread (σ_E) as a function of assumed detector response width (σ_t). What does this reveal about the model?

2 Additional Problems

2.1 Temperature Estimator from the Maxwell-Boltzmann Distribution*

The Maxwell-Boltzmann distribution describes the distribution of the speed, $v \in [0, +\infty)$ in m/s, of atoms in a gas at temperature T in kelvin (k). Its probability density is given by,

$$f(v;T) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 e^{-\dfrac{mv^2}{2kT}},$$

where m is the atomic mass and k is a constant known as the Boltzmann constant. The speed v of any atom will be a random variable $v \sim f(v;T)$.

i Show that the first algebraic moment of the Maxwell-Boltzmann distribution μ_1 (the expectation of v), is given by,

$$\mu_1 = E[v] = 2\sqrt{\frac{2kT}{m\pi}}.$$

where you may use the identity, $\int_0^{+\infty} x^3 e^{-bx^2} dx = \frac{1}{2b^2}$. .

ii An experimental apparatus is designed to measure the momenta of molecules of gas (and hence their speeds v). Suppose the experiment observes a large number n of speeds $\{v_1, v_2, ..., v_n\}$. These observations are independent and identically distributed random variables distributed as $v_i \sim f(v;T)$.

Using the *method of moments*, show that an estimator for the temperature of the gas $\hat{T}$ based on the sample mean, is given by,

$$\hat{T} = \frac{\pi m \bar{v}^2}{8k},$$

where $\bar{v}$ is the sample mean of the data set. .

iii In the data file provided `data_gas_speeds.csv`, the measured speeds of Caesium gas, from a data sample of measurements using the experimental apparatus, is given in the column "`speeds`". Read in the file (hint: use `pandas` to read the `.csv`) and calculate the estimator $\hat{T}$ for the data sample.

You may use the following constants

$$m = 132.905 \text{ amu}$$
$$k = 1.38 \times 10^{-23}$$

where 1 amu $= 1.66 \times 10^{-27}$ is the atomic mass unit.

iv Plot a *density* histogram (in matplotlib, you must specify `density=True` in the plotting of the histogram) of the speeds from the data sample and draw the Maxwell-Boltzman probability density function on top with the Temperature parameter $T = \hat{T}$ from your previous answer. On the same plot, draw the probability density function of the Maxwell-Boltzmann distribution for $T = 300$ and $T = 500$ kelvin too.

You may code the probability density yourself, or you may use the `scipy.stats.maxwell` pdf (see `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.maxwell.html`). For the latter, you need to set the 'scale' keyword argument appropriately as in the code below.

```python
import scipy.stats as st

# the function below calculates the maxwell-boltzmann
# pdf at a value x=100, for a specific value of the parameter T=300
T = 300
x = 100
scale = 1./((m/(k*T))**0.5)
pdf_x = st.maxwell.pdf(100,loc=0,scale=scale)
```

Remember to include axis labels on your figures!

v Estimate the bias and variance of your estimator $\hat{T}$ as a function of T. To do this, you will need to generate Monte Carlo data samples from the Maxwell-Boltzmann distribution for a specific value of T, and calculate $\hat{T}$ for each sample. The sample of values of $\hat{T}$ for each value of T can be used to estimate the sample bias and variance of $\hat{T}$.

Note that,

- Each Monte Carlo data sample must have the *same* number of events as in the original data sample
- Remember to make enough Monte Carlo data samples for a good estimate of the bias and variance (I recommend at least 1000 samples).
- The sample bias for our purposes here can be defined as $\text{bias}(\hat{T}) = \bar{\hat{T}} - T$ (i.e use the sample mean instead of the expectation value that we saw in lectures).
- The variance of $\hat{T}$ should be the *sample variance* using the Monte Carlo data samples.
- Investigate a sensible number of Temperatures in the range 100-500 kelvin.

To generate the samples, you may code up your own random generator or, much easier, use the `scipy.stats` in-built functionality as in the code below.

```python
# Generate MC data sample of speeds at temperature $T$ of a given size.
scale = 1./((m/(k*T))**0.5)
monte_carlo_sample = st.maxwell.rvs(loc=0,scale=scale,size=number_in_mc_sample)
```

Plot your *bias and sample standard deviation* of $\hat{T}$ from the Monte Carlo samples, where the latter is defined as $\sqrt{V(\hat{T})}$, of $\hat{T}$ as a function of T. Remember to include axis labels in your figures and a legend to make it clear what the plot is.

vi The bias of $\hat{T}$ can be calculated using the definitions given in lectures. The bias of $\hat{T}$ for a sample of size n is given by,

$$\text{bias}(\hat{T}) = \frac{\pi m \nu_2}{8kn},$$

where ν_2 is the variance of the Maxwell-Boltzmann distribution. Plot the bias of $\hat{T}$ from your solution, and the bias you estimated from part v as a function of T on the same axis to compare the results. You can use the fact that the variance of the Maxwell-Boltzmann distribution is given by $\nu_2 = \frac{kT}{\pi m}(3\pi - 8)$.