Q1. **The three doors problem.** Consider the following game-show set-up. Behind one of three closed doors, labelled A, B and C, lies a highly desirable prize. A contender is invited to select one door at random. They pick one –say door A– but do not open it yet. The game-show presenter (who *knows* where the prize is), then opens one of the remaining 2 doors –say door C–, which has no prize behind it. The contender is then given the option to modify their choice of door before opening it. To maximise their chance of winning the prize, should they do so, or stick with their original choice? Justify your answers.

Q2. **Murder rate in Morse's Oxford.** In the famous *Inspector Morse* TV series set in the 1960s, there were no fewer than 103 murders in just 33 episodes.

(a) What is the probability of having 7 murders in a single episode?

(b) What is the probability of going for two episodes without murders?

[Hint: *Neglect the possibility that murderers take holidays, and assume Poisson statistics.*]

Q3. **COVID testing.** The lateral flow tests used for community COVID testing in the UK have a sensitivity of 78% and a specificity of 99.6%. The first number means that, if 100 people who have COVID took the test, on average 78 of them would receive a positive result. The second number means that, if 1000 people who did not have COVID took the test, on average 4 of them would receive a false positive result.

Before heading into their lectures for the day, an Oxford undergraduate student takes a routine lateral flow test. The latest official data available on that day indicate that the rate of COVID in the population in Oxford is 302 per 100 000 people.

(a) What is the probability that the student has COVID if the test result is positive?

(b) What is the probability that the student does not have COVID if the test result is negative?

(c) How would your answers to questions (a) and (b) change if the incidence of COVID in Oxford was 10 times larger or smaller?

(d) How would your answers to questions (a) and (b) change if you were also told that the student went to a party a few days ago and feels a little unwell, but that there are roughly as many cases of flu as of COVID in Oxford? (A qualitative answer is sufficient.)

Q4. **Balls from an urn.** An urn contains C balls, of which W are white and $C - W$ are black. A ball is drawn at random N times, and replaced each time in the urn.

(a) What is the probability distribution for the number of times a white ball is drawn, n_W?

(b) What is the expectation value of n_W and its variance?

(c) Evaluate the expectation value and variance of n_W for the case $C = 100$, $W = 30$, $N = 10$.

Q5. **Dice throwing I.** Two dice, each with the shape of a regular tetrahedron (i.e., a polyhedron composed of 4 equilateral, identical triangles), are thrown together. The 4 sides of each die are numbered from 1 to 4.

(a) Sketch the probability distribution of the sum of their values.

(b) Now 1000 dice of the same sort are thrown. What is the probability distribution of the sum of their values?

Q6. **Poisson distribution.**

(a) A radioactive source emits 600 particles in a 4 hours interval. What is the probability distribution for the number of particles emitted in a time $t = 1$ minute? Calculate the probability of the source emitting $r = 0, 1, 2, \ldots, 10$ in 1 minute and sketch a histogram of these values.

(b) Now consider a source with an average rate $\lambda = 70$ min^{-1}. Calculate the probability of the source emitting $r = 90$ counts. [Hint: *If you encounter numerical issues because of the large value of λ, use Stirling's approximation: $n! \approx n^n e^{-n}\sqrt{2\pi n}$, compute the natural logarithm of the Poisson distribution and then exponentiate it back at the end.*]

(c) Compare your result above with the Gaussian approximation to the Poisson distribution, with the appropriate choice for μ and σ.

Q7. **Dice throwing II.** A fair, 6-sided die is tossed 100 times. What are the mean and standard deviation of the sum of the resulting values? Justify your answer.

Q8. **Known variance.** *This is a particularly important problem, make sure you understand it.*

The length ℓ of a rod is measured 5 times, yielding the following data (in meters):

$$2.32, 2.28, 2.27, 2.30, 2.32. \tag{1}$$

Each measurement has an uncertainty of $\sigma = 0.05\,\mathrm{m}$.

(a) What is the likelihood function for each measurement and for data set as a whole? Justify any assumptions you make in your derivation, and identify explicitly the data and the parameter you are trying to estimate.

(b) Find the Maximum Likelihood Estimate (MLE) of the length, ℓ_{ML}.

(c) Determine symmetric 90%, 95% and 99% confidence intervals around ℓ_{ML}.

Q9. **Unknown variance.** The momentum of a monoenergetic beam of particles is measured using an instrument whose measurement accuracy is not known, yielding the following data (in units of MeV/c):

$$8.6, 12.7, 11.6, 9.8, 10.0, 10.0, 10.1, 9.5, 10.6, 8.5. \tag{2}$$

In the following, we assume that the measurements are independent and Gaussian distributed.

(a) What is the Maximum Likelihood Estimate (MLE) of the momentum of the beam?

(b) What is the unbiased estimate of the standard deviation of each measurement?

(c) Give a 95% confidence interval for the value of the mean. How would you report your result?

(d) After the measurements are done, the physicist who set up the beam tells you that its momentum was in fact $8.0\,\mathrm{MeV}/c$. How many sigma away is this value from your MLE for the momentum? What does this mean?

Q10. **Error propagation.** Consider a function $z = f(x, y)$, where x and y are independent random variables.

(a) By expanding $f(x, y)$ about $(x, y) = (\mu_x, \mu_y)$, where μ_α represents the expectation value (mean) of α, show that $\mu_z \simeq f(\mu_x, \mu_y)$.

(b) Hence, show that the variance of z obeys

$$\sigma_z^2 \simeq \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2, \tag{3}$$

where the partial derivatives are evaluated at $(x, y) = (\mu_x, \mu_y)$.

(c) A model for the Earth's atmosphere predicts that the temperature T at a height h above Earth's surface varies as $T = T_0 - L\,h$, where T_0 is the surface temperature and L is known as the adiabatic lapse rate. On a certain day, measurements give $T_0 = 295 \pm 2\,\mathrm{K}$ and $L = (6.5 \pm 0.5) \times 10^{-3}\,\mathrm{K\,m^{-1}}$. Determine the expectation value of T and its experimental uncertainty for a weather balloon at a height $h = 2800 \pm 200\,\mathrm{m}$.

Q11. **Photon counts and source strength.** An astronomer measures the photon flux from a distant star using a very sensitive instrument that counts single photons. After a given exposure time t, the instrument has collected $\hat{r}$ photons. One can assume that the photon counts, $\hat{r}$, are distributed according to the Poisson distribution. The astronomer wishes to determine λ, the emission rate of the source.

(a) What is the likelihood function for the measurement? Identify explicitly the data and the known and unknown parameters of the problem.

(b) If the true rate is $\lambda = 10$ photons min^{-1}, what is the probability of observing $\hat{r} = 15$ photons in $t = 1$ min?

(c) Derive an expression for the Maximum Likelihood Estimate (MLE) for the rate, λ_{ML}, in terms of $\hat{r}$ and t. Evaluate λ_{ML} if $\hat{r} = 10$ and $t = 1$ min.

(d) Upon reflection, the astronomer realizes that the photon flux is the superposition of photons coming from the star plus "background" photons coming from other faint sources within the field of view of the instrument. To measure the background rate λ_b, they point the telescope away from the source, where the telescope should pick up only background photons, and measure the photon counts there. They then point to the star again, measuring $\hat{r}_t$ photons in a time t_t. Derive an expression for the MLE of the rate λ_s of photons coming from the star. Evaluate $\lambda_{s,\mathrm{ML}}$ if $\hat{r}_t = 34$, $t_t = 2$ min, and $l\lambda_b = 10$ photons min^{-1}?
[Hint: *The total number of photons $\hat{r}_t$ is Poisson distributed with rate $\lambda_t = \lambda_s + \lambda_b$, where λ_s is the rate for the star.*]

(e) Instrumental errors aside, what is the measurement uncertainty (standard deviation) on the total photon count $\hat{r}_t$? What is the resulting uncertainty on the source rate, $\sigma(\lambda_s)$?
[Hint: *You may use Equation (3), and assume that the background rate has been measured so precisely that its uncertainty is negligible.*]

(f) How confident can the astronomer be that they have detected a source at all? How long would the exposure time need to be to achieve a 5σ detection?

(g) Is it possible that the measured total count-rate $\hat{r}_t/t_t < \lambda_b$? How should the astronomer address such a possibility to ensure that the measured source rate is physical?

Q12. **Bayesian estimation of the mean of a Gaussian.** This problem takes you through the steps to derive the posterior distribution for a quantity of interest θ, in the case of a Gaussian prior and Gaussian likelihood.

Let us assume that we have made N independent measurements, $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$ of a quantity of interest θ (this could be the temperature of an object, the distance of a galaxy, the mass of a planet, etc). We assume that each of the measurements in independently Gaussian distributed with known experimental standard deviation σ. Let us denote the sample mean by $\hat{x}$, i.e.

$$\hat{x} = \frac{1}{N}\sum_{i=1}^{N} \hat{x}_i. \tag{4}$$

Before we do the experiment, our state of knowledge about the quantity of interest θ is described by a *prior*, which we will take to be a Gaussian distribution with mean μ_0 and variance σ_0^2:

$$p(\theta) = \mathcal{N}(\mu_0, \sigma_0^2). \tag{5}$$

Such a prior might come e.g. from a previous experiment we have performed. The new experiment is however much more precise, i.e. $\sigma_0 \gg \sigma$. For simplicity, we will assume $\mu_0 = 0$ (it is always possible to chose the units of θ such that this is the case.)

(a) Write down the likelihood function for the measurements and show that it can be recast in the form:

$$\mathcal{L}(\theta) = L_0 \exp\left(-\frac{1}{2}\frac{(\theta - \hat{x})^2}{\sigma^2/N}\right), \tag{6}$$

where L_0 is a constant that does not depend on θ.

(b) By using Bayes theorem, compute the posterior probability for θ after the data have been taken into account, i.e. compute $p(\theta|\hat{x})$. Show that it is given by a Gaussian wit mean

$$\mu_{\rm p} = \hat{x}\,\frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{N}} \tag{7}$$

and variance

$$\sigma_{\rm p}^2 = \left[\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right]^{-1}. \tag{8}$$

[Hint: *you may drop the normalization constant from Bayes theorem, as it does not depend on θ.*]

(c) Show that, as $N \to \infty$, the posterior distribution becomes independent of the prior.

(d) Show that, as $N \to \infty$, the mean of the posterior distribution converges to the MLE of the mean for θ. This means that for a large number of measurements, the Bayesian result matches the frequentist MLE result.