5 Final assignment: an X-ray source safari

*This assignment is marked out of 70. All data are either included in this document, or available via blackboard. Data analysis should be performed using a **Jupyter/IPython Notebook**, and submitted via blackboard. Any mathematical derivations should preferably be typeset in LaTeX and converted to a PDF. If it is logical to include such a derivation in your notebook, you can do so using **Markdown**, which is supported by the notebook.*

Deadline: Wednesday 4 January, 23:59. Late submissions will have their actual marks reduced by 20% (up to 1 day late) or 40% (up to 2 days late - submissions later than this will not be graded!).

This is the fifth of five problem sets and is worth 35% of the final grade.

5.1 Introduction

In this final assignment, you will be tested on your ability to interpret data using the methods of maximum likelihood estimation, significance and hypothesis testing which you have learned during the last few weeks of the course. The assignment is based on an observation of a nearby galaxy by a hypothetical X-ray telescope. From that observation, you have extracted the spectral data (a list of measured photon energies, in units of keV) from the ten brightest sources in the observed field-of-view. You have been sent (via email) a unique data file, which contains the data for your ten sources.

The file consists of two columns: the first gives the source number (from 1 to 10) and the second gives the photon energy. Thus, each photon is labelled with the corresponding source and the data from all sources is concatenated into a single long list for convenience. To obtain the spectra for each source, the data from the entire file can be read in and the source number used to separate the photon energies for each source using an appropriate method in Python (e.g. see the Michelson data described in the week 1 tutorial for one approach to doing this).

5.2 What you need to do

The aim of this assignment is to analyse the data from the ten sources by using appropriate techniques to fit models to the photon spectra, determine and report the best-fitting parameters (MLEs) and their 1-dimensional confidence intervals (using *both* the covariance matrix and brute-force approach where possible) and check the significance of any features (e.g. emission lines) which you see in the spectra. Based on your results, and using goodness-of-fit and/or hypothesis tests where appropriate, you should try to identify the different sources in your data, guided by the information given below (which should also be used to guide your model fitting). Note that the sources themselves cover a range of fluxes: some have only a couple of hundred photons while the brightest few sources have $\sim 10^4$ photons. You will need to choose your approaches to the data accordingly.

Throughout the assignment you should work as if you are presenting your results to other scientists: explain what you are doing clearly **and be sure to state all your assumptions!** Display your data and model fits using clear plots with appropriate labelling and choice of axes. If appropriate, use plots to support your arguments based on the formal tests, e.g. about the reality of features in the data, or to explain why one model fits better than another (is there systematic structure in the residuals, which goes away when fitting the better model?).

Note that although this assignment is based on a hypothetical astronomical data set, you do not need any astrophysical background to understand and interpret the data: all necessary information for interpreting the data and fitting models is given below, and the situation has been simplified so that background knowledge of X-ray astronomy will not help you. Our aim here is to test your ability to fit models to binned data, to fit features in the spectrum with appropriate models and interpret their significance, and to interpret your results using what you have learned in the course over the last few weeks.

5.3 A bestiary of X-ray sources

Your data set contains the brightest ten sources from your observation, which limits the range of possible X-ray sources to the more luminous objects that could be present in your target galaxy. Here are the possibilities you should consider in your analysis. Where appropriate we describe the spectral shapes expected for each source type, and give 'typical' ranges of parameter values that might be expected. Note that $N(E)$ is the continuum flux in units of photons/keV at energy E keV. N_0 is a normalisation defined at an energy $E_0 = 1$ keV (although the exact definition does not matter since the normalisation can be fitted as a free parameter).

- *Ultra-luminous X-ray sources (ULXs):* the physical nature of these sources is uncertain: they may be stellar-mass black holes accreting at very high rates, or more massive 'intermediate mass' black holes, accreting at more moderate rates. A common and apparently unique feature is that many ULXs show spectra with some type of cut-off or break to a steeper slope at high energies. You can model this break either with a broken power-law model:

$$N(E) = \begin{cases} N_0(E/E_0)^{-\Gamma_1} & \text{for } E \leq E_{\rm bk} \\ N_0(E_{\rm bk}/E_0)^{-\Gamma_1}(E/E_{\rm bk})^{-\Gamma_2} & \text{for } E > E_{\rm bk} \end{cases} \tag{1}$$

  or with an exponentially cut-off power-law:

$$N(E) = N_0(E/E_0)^{-\Gamma}\exp(-E/E_{\rm cut}) \tag{2}$$

  and find which model is most appropriate (if any). The break/cutoff energies are typically in the range 6-9 keV with indices below and above the break of Γ =1.6–2 and 2.6–3.0 respectively.

- *Young supernova remnants (SNRs):* the X-ray emission from SNRs is typically characterised by steep power-law emission ($N(E) \propto E^{-\Gamma}$) due to synchrotron radiation, with index Γ =2.3–2.8, and prominent emission lines from the hot plasma, primarily from Si XIII at 1.83 keV and S XV at 2.4 keV. Ionised Fe emission may also be visible at 6.7 keV, if there is sufficient signal-to-noise at those energies.

- *Ultrasoft X-ray binaries (XRBs):* the emission from these luminous black hole X-ray binary systems may also be visible. This thermal emission from the accretion disk may be modelled simply with a blackbody spectrum:

$$N(E) = \frac{N_0(E/E_0)^2}{\exp(E/k_{\rm B}T) - 1} \tag{3}$$

  where $k_{\rm B}T$ is the equivalent temperature of the disk blackbody spectrum in units of keV, and should be typically in the range 0.8-1.4 keV.

- *Contaminating background sources - Active Galactic Nuclei (AGN):* your data contains sources associated with the target galaxy, but such samples are also likely to be contaminated by 'background' sources which may be large distances behind the target galaxy. The most common type of contaminating background sources are AGN - supermassive black holes in distant galaxies which are undergoing strong accretion and emitting significantly in X-rays as a result. AGN are interesting in their own right, and it is useful to identify them from their X-ray spectrum if at all possible. We can consider two types that might be identifiable:
  Blazars: show relatively steep (Γ =2–2.5) featureless power-law spectra.
  Seyferts: show typically flatter power-law spectra (Γ =1.7–2.2) than blazars and an important difference is that they may show detectable emission from Fe at 6.4 keV (assuming the object is at a moderate redshift so the line appears at or close to the rest-frame energy).

- *The unexpected:* Besides these objects just described, as a scientist you should also be prepared for the unexpected! If you see anything interesting in your spectra which you can't easily identify with the characteristics described for these sources, investigate it and report it!

5.4 Other important information

Although we assume here a perfect 'instrumental response', i.e. the X-ray detector records (on average) numbers of photons proportionate to their actual fluxes, you can assume that any narrow line features present in the data are observed at the instrument resolution, not their true width. Therefore, you may assume that if you see (or think you see) such features, they may be fitted with a Gaussian profile (note that this is the same as a normal distribution shape), with a peak energy $E_{\rm line}$ keV and width $\sigma = 0.06\sqrt{E_{\rm line}}$ keV.