Alexandra Cain 7 May 2019

GAUSSIAN FIT TUTORIAL UTILIZING LEGA-C DATA

Abstract:
This tutorial will demonstrate how to produce a Gaussian fit of data using Python. The data we specifically will focus on relates to the [OIII] emission line of star-forming galaxies. The data will be presented on graphs for a visual portrayal of the spectrum and specifically the [OIII] emission line. A model the of this emission line is generated using a Gaussian1D fit. The Gaussian fit produces values that allow us to calculate and understand more about the galaxy as whole.

1. Introduction:
In our lifetime, we won't be able to visit any other galaxies. Everything that we learn from them must be a result of the instruments used to image them. The data used in this tutorial is from the Large Early Galaxy Astrophysics Census (LEGA-C; van der Wel et al. 2016). LEGA-C covers redshifts from 0.6 to 1, where the redshift, z, is a definition of a given epoch in the universe. Understanding redshift is dependent on understanding the Doppler effect. Essentially, the Doppler effect says that the wavelength of waves in relation to the observer is modified by movement. If the object has velocity towards the observer, waves originating at that object will be compressed, shortening the wavelength and shifting the waves towards the blue end of the spectrum. If an object has a velocity away from the observer, waves originating at that object will be stretched, elongating the wavelength and shifting it toward the red end of the spectrum. Because space itself is expanding, galaxies that are further away from us are expanding at a faster rate than those nearby. The redshift of a galaxy indicates how far away the galaxy we are observing is. A galaxy with a lower redshift is moving slower, indicating that it is in closer proximity. As the rest-frame wavelength values are increased by a factor of $(1+z)$, we can use observed emission lines to directly determine the redshifts of galaxies. Emission-line wavelengths are consistent because they are the direct result of specific energy transitions occurring within galaxies, specifically star-forming ones.

Star formation necessitates gas. When galaxies are forming stars from their cold gas reservoirs, ultraviolet light from hot stars ionizes the surround gas to form an HII region. The excited electrons fall back down, radiating photons of discrete wavelengths, observed as recombination lines. Consequently, galaxies with new star formation have spectra with strong emission lines (in the absence of dust). One of these emission lines is doubly ionized oxygen, known as the [OIII] emission line. The [OIII] emission line qualifies as a forbidden transition: whereas, the rate of collisional de-excitation dominates over spontaneous photon emission, on the outer edges of a gaseous galaxy, the low-density environment permits these otherwise statistically rare (though not technically forbidden) transitions. The [OIII] doublet lines are found at a rest wavelength of approximately 5007 Å and 4959 Å, with a natural broadening around these wavelengths. These [OIII] lines have little collision or proximity broadening, allowing them to be modeled by a Gaussian.

2. What is Gaussian Function?
Gaussian functions are widely used functions in many scientific disciplines. They are normal distributions used to model data. They can include three to six parameters to model data. For the purposes of this tutorial, we will be using the most simplistic Gaussian, with 3 terms. The

parameters for a three term Gaussian include the amplitude, the center and the standard deviation. The equation is:

$$f(x) = Ae^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

The amplitude is represented by A, the center is represented by μ, and the standard deviation is represented by σ. The x values, which is wavelength in this tutorial, is represented by x. The y values, which corresponds to flux in this tutorial, is represented by *f(x)*.

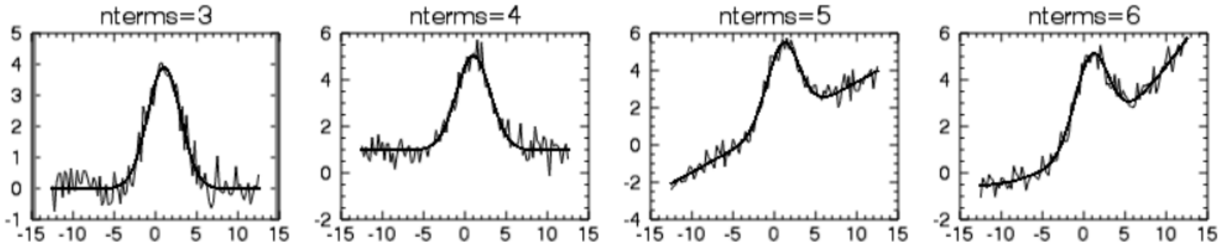

Figure 1 shows the relative shapes of Gaussian curves with three to six parameters (nterms).

The equation above correlates to the farthest left Gaussian line shape of three parameters. This picture and more information regarding Gaussian fits with more parameters can be found at https://www.harrisgeospatial.com/docs/GAUSSFIT.html

3. Getting Started:

In this tutorial, we will learn how to read in a spectrum for a LEGA-C galaxy at *z*=0.6874, shift the spectrum to the rest frame, and fit the [OIII] with wavelengths of 5007 Å and 4959 Å emission lines with two Gaussians. The first step to getting started is assuring that python is appropriately installed with the necessary packages to run this program. Open terminal. It should read the computer and the user. Typing *python* into the line and pressing enter should begin to run the python program. If the terminal does not recognize the command, either python is not installed or not connected to the terminal. Once you have python installed correctly, you need to confirm that you have the appropriate modules within python. Within the running python program, type *help("modules")* . Python will list all available modules. For this tutorial, you will need *numpy*, *astropy*, and *matplotlib*. To exit the python program, type *exit()*. It is time to start your python program! Open your python program by using a text editor. There are different text editors to use: Emacs, Nano, and Vi/Vim. This tutorial will use *emacs*. To open a new python file, just type your text editor name.py. Example: *emacs gaussian.py* . This will open the text editor of your (currently empty) file. It's time to start adding things to our program file. Alternately, you can use a Jupyter notebook. A simple tutorial introducing Jupyter can be found at https://www.dataquest.io/blog/jupyter-notebook-tutorial/.

However you plan to write your code, begin by telling the program file what modules to import and how they will be referenced.

```
In [1]: import numpy as np
        from astropy.io import fits, ascii
        import matplotlib
        import matplotlib.pyplot as plot
        from astropy.modeling import models, fitting
```

If you already have the data set you want to fit prepared, skip ahead to section 8.

4. Picking a Galaxy:

In order to be able to model an [OIII] emission line with a Gaussian, the [OIII] spectral feature must be visible. This is dependent not only upon whether the galaxy is emitting the [OIII] transition, but also if the redshift permits the transition to be observable given the rest frame of wavelengths and redshift of the source. The VLT instrument is sensitive to wavelengths observed from 5800.3 to 9499.3 Å, with the redshifts of the LEGA-C galaxies spanning approximately z=0.6 to 1. In order to convert between observed wavelength and rest wavelength, we use the following formula:

$$z + 1 = \frac{\lambda_{obs}}{\lambda_{rest}}$$

The [OIII] line is therefore observable if the redshift of the galaxy falls between $z = 0.158$-0.897, and [OIII] will only be covered for LEGA-C galaxies at z<0.9. Fortunately, there is a simpler and more effective way to pick a galaxy. Since the LEGA-C data has already been heavily analyzed, the catalogue includes many pieces of information relating to each galaxy, including the flux of the [OIII] line that we are attempting to (re-)analyze. So, let's cheat a little and just go into the catalogue and identify the galaxies that have relatively high [OIII] emission lines. Also using the catalogue, we can read off the measured redshift for the galaxy we will be analyzing.

```
In [2]: cat=('\\Users\Alex Cain\Documents')
        data = ascii.read(cat+'\DR2_legac_team.cat')
        z_spec=data['z_spec']
        galaxyid=data['id']
        model_flux_Oiii_5007 =data['model_flux_Oiii_5007']
        OIII_selection = [(model_flux_Oiii_5007 > 1400)]
        galaxies= galaxyid[OIII_selection]
        zvalue=z_spec[OIII_selection]
        print(galaxies, zvalue)
```

The first line of code establishes the path to the file. This will be different for everyone, depending on where your catalogue is located on your computer. The second line use (ascii) to read in the data from the catalogue. Next, the arrays of the (z) value, the galaxy (id), and the model flux of [OIII] are established for evaluation and reference. A new array is generated consisting of values of the model flux [OIII] that fit our parameter of a minimum flux of 1400. This value can be modified depending on how many galaxies you want to evaluate. A lower number will include weaker [OIII] lines, while a higher number will include less values but with stronger fluxes. Next, arrays for the (id) and (z) value are restricted by these [OIII] values. The arrays are then printed. From this code, you will receive the id numbers of galaxies who have model [OIII] flux greater than 1400, along with their corresponding z values. For this tutorial, we are going to use the second one on the list, galaxy id #127984 with a spectroscopic z of 0.6874. Note that not all of the LEGA-C data is public yet.

id	z_spec
127180	0.5916
127984	0.6874
132979	0.7282
129027	0.7479
108271	0.6372
109010	0.7247
111932	0.6262
113714	0.6681
118009	0.7584
73754	0.7474
...	...
95970	0.728396
102108	0.609207
181614	0.616214
183621	0.678759
185711	0.635067
220107	0.738409
244653	0.7085
257051	0.70214
31349	0.711148
32080	0.710511
78550	0.670635
Length = 37 rows	Length = 37 rows

5. Importing the Data:

```
In [3]: z_spec= 0.6874
        z= 1+z_spec

        xdata=np.loadtxt('\\Users\\Alex Cain\\Documents\\legac_spec_wave.dat')
        xval=xdata/z
```

The *z* value can then be hardcoded (written in manually). Additionally, you will need to import the observed wavelength values of the LEGA-C data. The observed wavelengths are constant because it is purely the wavelengths observed by the instrument. The observed wavelength is then converted to the rest wavelength using the redshift. Due to different *z* values, the rest wavelengths will be different for all galaxies. Now you have the rest wavelength of your galaxy.

```
In [4]: ycontdata=fits.open('\\Users\Alex Cain\Documents\legac_M1_v0.5_spec1d_127984_cont.fits')
        ycontval= ycontdata[0].data
        yspecdata=fits.open('\\Users\Alex Cain\Documents\legac_M1_v0.5_spec1d_127984.fits')
        yspecval= yspecdata[0].data
        yemdata=fits.open('\\Users\Alex Cain\Documents\legac_M1_v0.5_spec1d_127984_em.fits')
        yemval= yemdata[0].data
```

The galaxy spectra include the observed composite spectrum of the stars and gas, which is fit with a combination of the continuum (absorption-line) and emission-line spectra. While all galaxies have an underlying absorption-line spectrum from the integrated light of all of the stars in the galaxy, the most prominent features of star-forming galaxies are instead the stronger nebular emission lines originating mostly from ionized gas. The model spectrum for any object is therefore the linear combination of the best-fit continuum-only plus emission-line templates. These models set the standard values for which the observed spectrum values will be compared to. Specifically, we will first subtract the absorption-line spectrum to then consider the emission-line only flux.

6. Making a Graph:

Let's make a basic graph of the full spectra of the galaxy. This will give you a quick understanding of the galaxy you are exploring and confirm that the previous methods of picking a galaxy were effective. The x axis will be the rest wavelengths and the y axis will be our emission and continuum flux values for the first and second line, respectively. Graphing the continuum second will have it present on top of the emission line only spectrum. Next, we label the x and y axes and show our plot.

```
In [5]: plot.plot(xval,yspecval)
        plot.plot(xval,ycontval)
        plot.xlabel('Wavelength')
        plot.ylabel('Flux')
        plot.show()
```

This plot shows the strong emission lines relating to [OIII], at approximately 5007 Å and 4959 Å. This portion of the tutorial is a good place to check your understanding for both data manipulation and interpretation. You can graph the data of other galaxies,

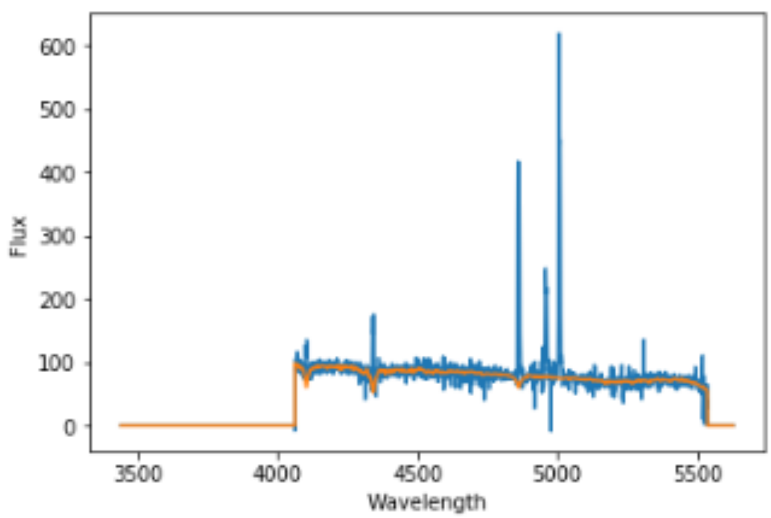

Figure 2 shows the galaxy in the rest wavelengths, with the continuum (orange) and the spectrum (blue).

7. Define Parameters and Graph:

Now that we are confident in our galaxy selection, it is time to proceed with the [OIII] emission analysis. The emission line that interests us occurs at wavelengths of approximately 5007 Å and 4959 Å. The index values (*iv*) within the rest wavelength that correspond to 40 Å below and above the target wavelength are identified. Now, the x axis consists of only the values within that index. The y values need to be restricted by the index values as well such that the arrays are the same length.

```
In [6]: wl1 = 5007
        wl2 = 4959
        iv = [i for i, x in enumerate(xval) if x>(wl2-40) and x<(wl1+40)]
        x=xval[iv]
        ycont=ycontval[iv]
        yspec=yspecval[iv]
        yem=yemval[iv]
        plot.plot(x,yspec,label='spectrum')
        plot.plot(x,ycont,label='continuum')
        plot.xlabel('Wavelength')
        plot.ylabel('Flux')
        plot.show
```

Two lines are plotted for the spectrum and continuum. The axes are labeled, and the plot shown in Figure 3. Once again, check your understanding of the data manipulation and interpretation. Questions you may want to ask yourself include: what do ($wl1$) and ($wl2$) represent, and how are they used in regard to data manipulation? You can also pick another galaxy and go through the process again.

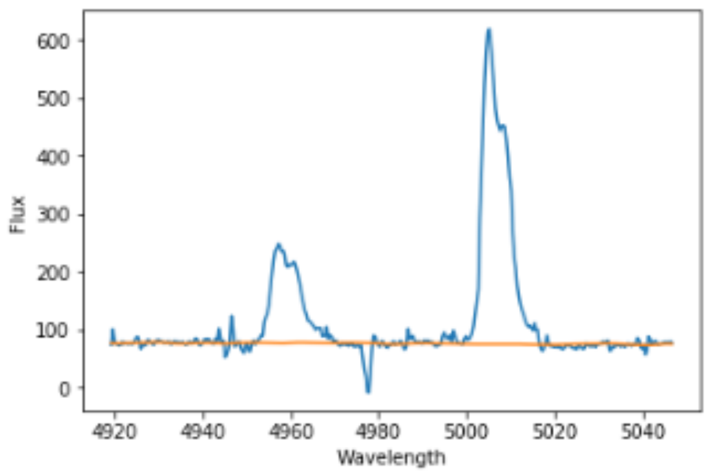

Figure 3 shows the [OIII] emission lines (blue) and the continuum (orange) for a reduced range.

8. Generate Fit and Graph:

Now that you have the observed data corresponding to the [OIII] emission lines ready to go, it is time to fit the Gaussian function to the data. Please note that the emission line is with respect to the continuum. Therefore, the y values in this analysis equal the y values of the spectrum minus the y values of the continuum.

The fitter uses *LevMarLSQFitter* to minimize the least squares fit of your function to arrive at the appropriate values. For the model, establish the model as Gaussian1D and attach some parameters to guide the fitting program. Parameters should be general approximations to your expected values for each variable. There are two ways to accomplish this. We can look at the Figure 3 to identify an approximate amplitude and hardcode the value in, or we can use ($max(y)$) determine the maximum value of the emission flux. The mean wavelength is set at 5007 Å. The standard deviation is set to 1 which is a generally standard initial prediction.

```
In [7]: y=yspec -ycont
fitter= fitting.LevMarLSQFitter()
model1= models.Gaussian1D(amplitude=max(y), mean=5007, stddev= 1, fixed= {'mean': True})
gaussian1=fitter(model1,x,y)
print(gaussian1)
amp2=gaussian1.amplitude/3
model2= models.Gaussian1D(amplitude= amp2 , mean= 4959, stddev=1, fixed= {'mean': True, 'amplitude': True})
gaussian2=fitter(model2,x,y)
print(gaussian2)
plot.plot(x,yspec,label='spectrum')
plot.plot(x,ycont,label='continuum')
plot.plot(x,gaussian1(x), label='Gaussian 1')
plot.plot(x,gaussian2(x), label='Gaussian 2')
plot.legend(loc='upper left')
plot.xlabel('Wavelength')
plot.ylabel('Flux')
plot.show()
```

Printing the models presents the values of the parameters. The mean was fixed at 5007 Å. The amplitude of flux is found to be about 494 with arbitrary units and a standard deviation around the mean of 2.79 Å. Remembers, the [OIII] emission line has strict 1:3 ratio regarding the amplitudes observed at the two different wavelengths. Using this information, the Gaussian fit for the second wavelength at 4959 Å is fixed at the amplitude of the first Gaussian divided by three. The mean is fixed as well, and the fit is generated to give us a standard deviation of 3.17 Å The two results from the Gaussian fits are shown below

```
Model: Gaussian1D
Inputs: ('x',)
Outputs: ('y',)
Model set size: 1
Parameters:
        amplitude       mean        stddev
    ----------------- ------ -----------------
    477.3053347743253 5007.0 2.940444494329284
```

```
Model: Gaussian1D
Inputs: ('x',)
Outputs: ('y',)
Model set size: 1
Parameters:
         amplitude        mean        stddev
    ------------------ ------ -----------------
    159.10177825810842 4959.0 3.169667217258401
```

Figure 4 shows the continuum, the spectrum, and the two Gaussian models for the galaxy.

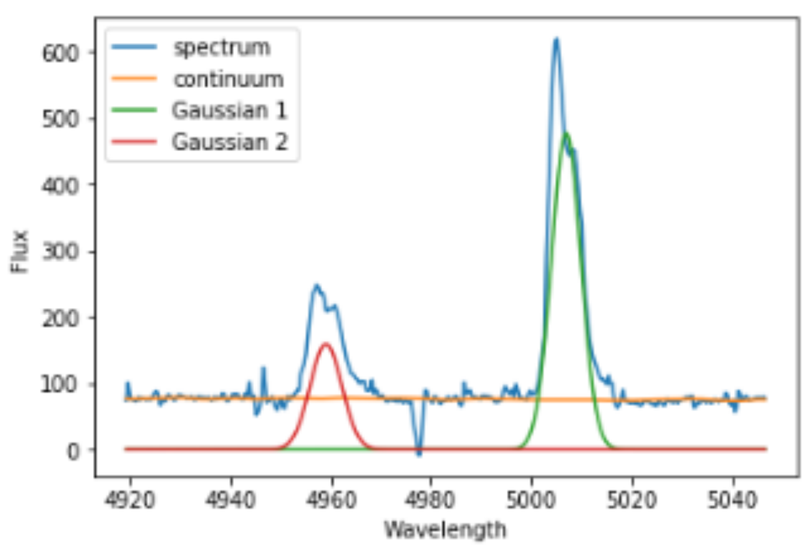

Figure 4 shows the continuum (orange), the emission (blue) and the Gaussian fits for both the 5007 Å (green) and 4959 Å (red) wavelengths.

This tutorial deals primarily with the fitting of a simple Gaussian. The 5007 Å wavelength Gaussian was fit first and the amplitude from this fit was directly used to produce the fixed amplitude of the 4959 Å wavelength Gaussian fit. There are more complex ways to relate your Gaussian fits that more accurately characterize the [OIII] doublet. If you'd like to learn more about ways to relate your Gaussian fits, http://docs.astropy.org/en/stable/modeling/ is a great place to start.

9. Understanding Your Produced Values:

Remember, this data has already been analyzed. This gives us the opportunity to compare our result with the result of an expert. The emission line includes the Gaussian fit of the spectrum, while our calculated Gaussian fit of the spectrum is also included

```
In [8]: plot.plot(x, gaussian1(x)+gaussian2(x), label='Gaussian')
        plot.plot(x,yem,label='emission')
        plot.xlabel('Wavelength')
        plot.ylabel('Flux')
        plot.legend(loc='upper left')
```

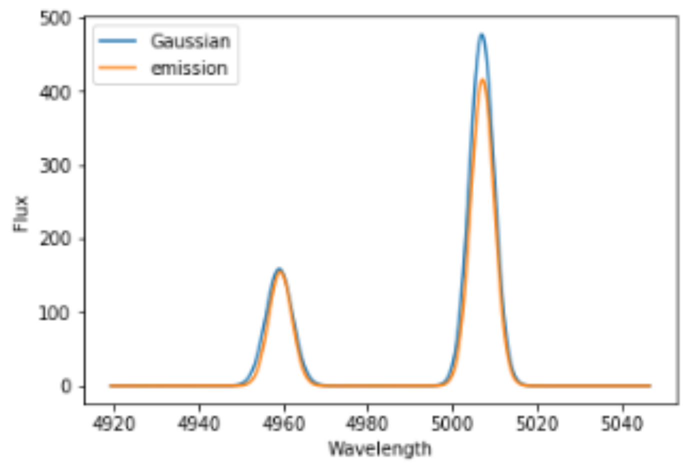

Figure 5 shows our Gaussian fit (blue) and the emission line (orange) attributed to this galaxy by an expert

Our Gaussian fits are slightly different than the experts. This is likely due to the simplistic approach we took to relating our Gaussian models. Still, we have two Gaussians of the correct wavelengths and ratio!

So, what can we do with this information? We can identify the amplitude and the full width half maximum. The full width half maximum is the width of the model at half the maximum height. In a Gaussian model:

$$FWHM = 2\sqrt{2ln2}c$$

The full width half maximum helps us define resolution within the spectrum, while the amplitude tells us the strength of the emission line. This can inform us about the galaxy, such as the rate of star formation, the electron density, the temperature, and the amount of ultraviolet light begin emitted.

Gaussian fits are a staple in astrophysics. So take the time to understand how to fit them and what information they can offer you!