The Consortium for Enabling Technologies and Innovation

Virtual Summer Meeting for Young Researchers

INTRODUCTORY BAYESIAN APPROACH TO GAMMA SPECTRA ANALYSIS FOR ISOTOPIC IDENTIFICATION

Ryan Lester, The University Of Texas At Austin
Cole Thompson, The University Of Texas At Austin

ETI

National Nuclear Security Administration

Disclaimer: This work was done as part of an AI for Engineering workshop and was not funded by ETI.

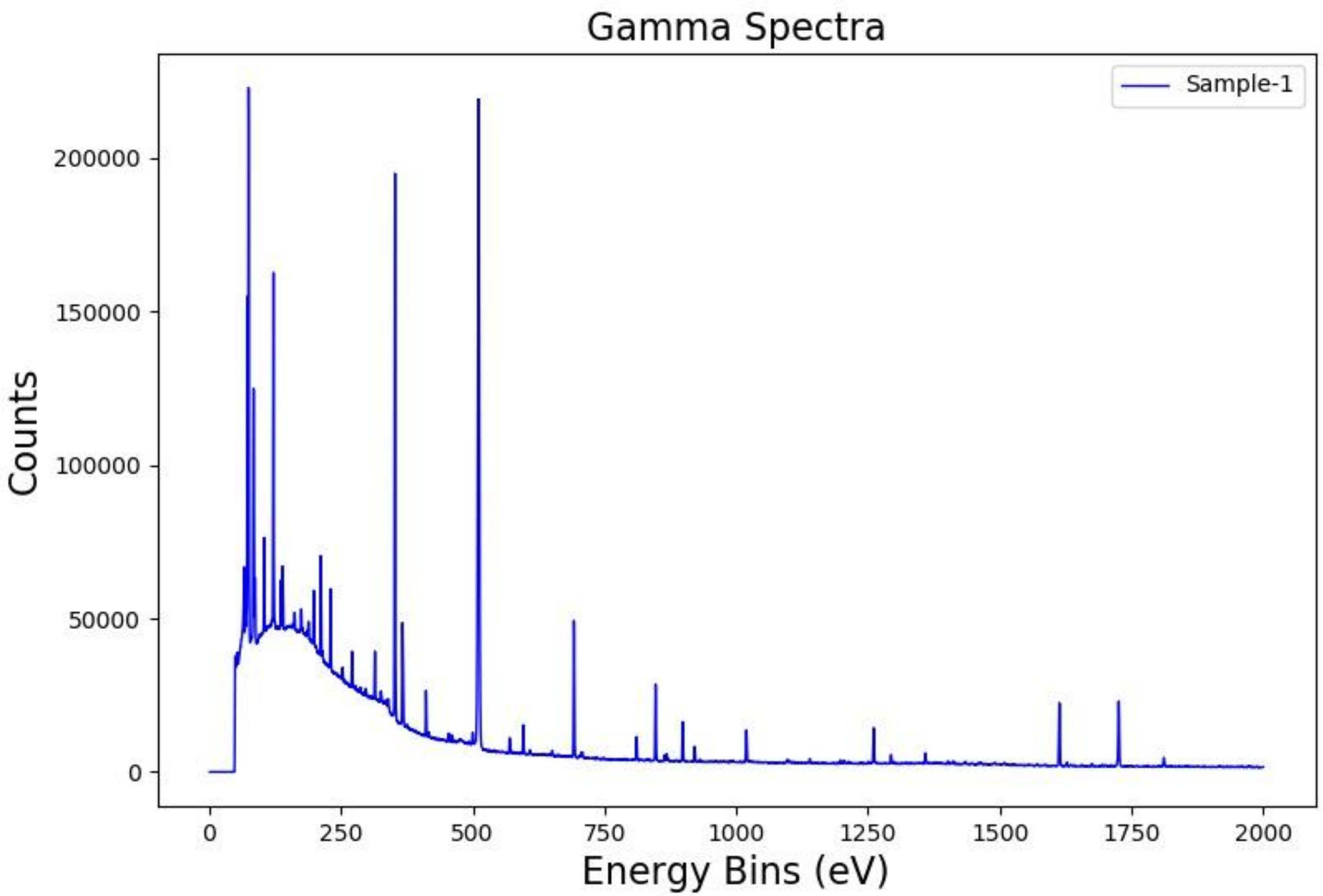
What is PGAA?

- Concept
- Challenges

Step 1:

Establish Background

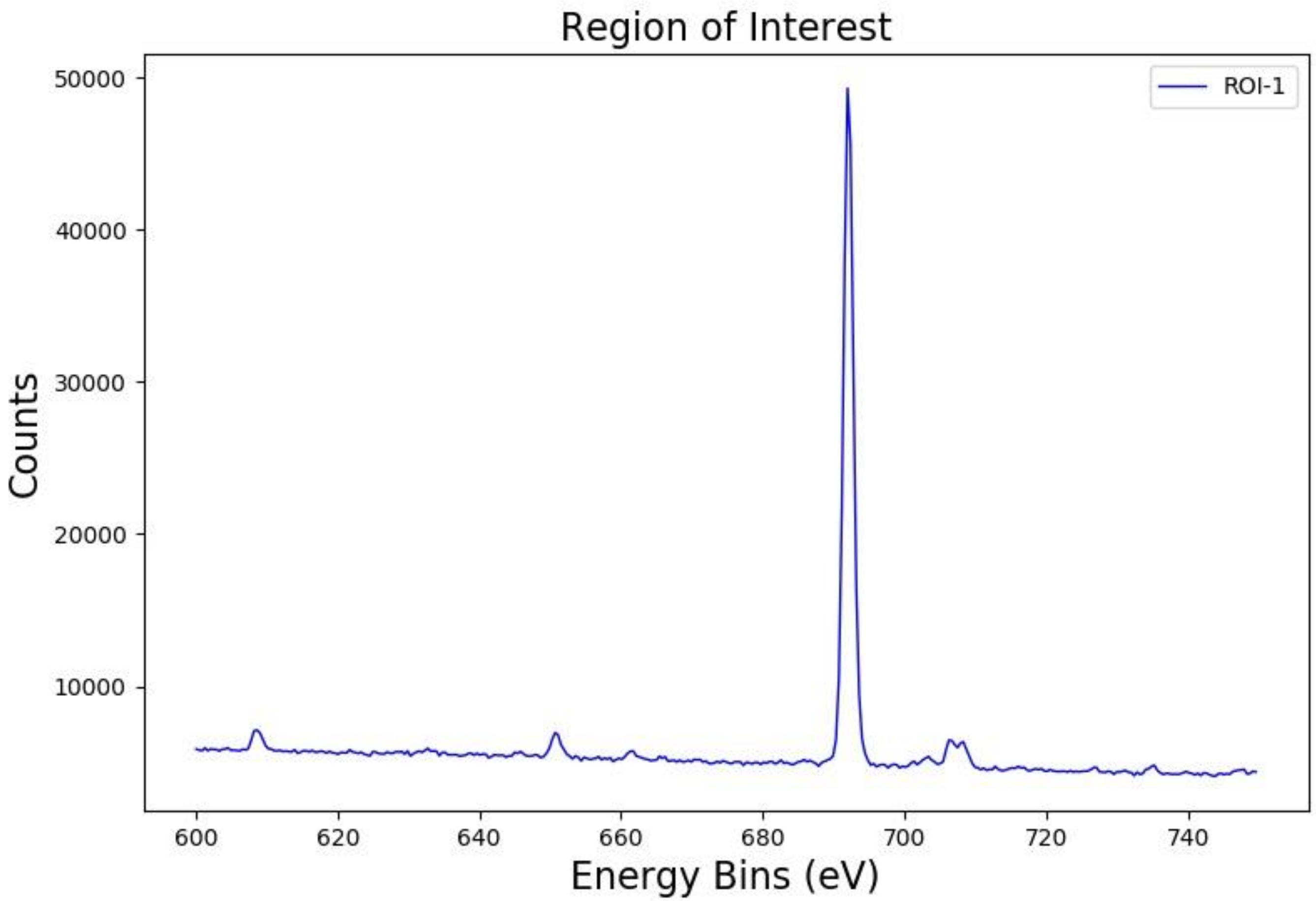
- Loading and parsing dataset



Step 2:

Establish ROI

- Defining an energy window around a Region of Interest (ROI)
- Plot showing



Step 2a:

Materials

PGAA Database

- Material selection
- Materials present
- List of most likely cross-section and material

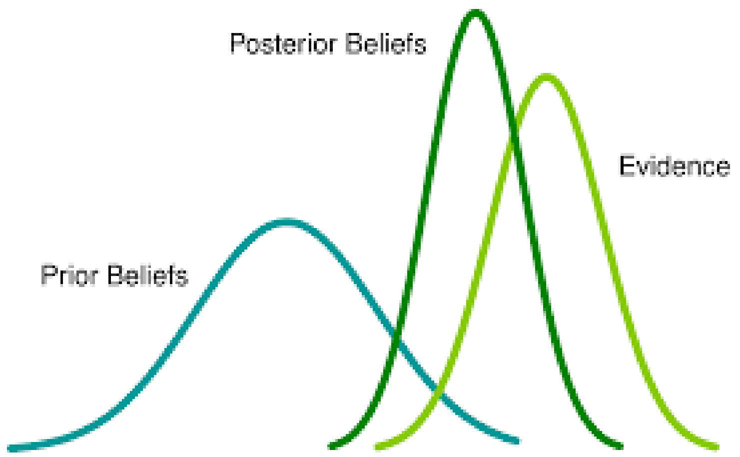
	A	isotope	Z	T	Gamma Energy	Gamma Energy Value	Sigma	Sigma Value	Uncertainty	Knaught	Knaught value	Knaught unc
0	1	2-H	1	p	2223.24835(9)	2223.248	0.3326(7)	3.326000e-01	7.000000e-04	1.0000(21)	1.000000e+00	2.100000e-03
1	2	3-H	1	p	6250.243(3)	6250.243	0.000519(7)	5.190000e-04	7.000000e-06	0.001560(21)	1.560000e-03	2.100000e-05
2	3	4-He	2	p	20520.5	20520.460	4.2E-11(12)	4.200000e-11	1.200000e-11	3.2E-11(9)	3.180000e-11	9.090000e-12
3	6	7-Li	3	p	477.595(3)	477.595	0.00153(8)	1.530000e-03	8.000000e-05	0.00067(4)	6.680000e-04	3.490000e-05
4	6	7-Li	3	p	6768.81(4)	6768.810	0.00151(9)	1.510000e-03	9.000000e-05	0.00066(4)	6.590000e-04	3.930000e-05

	A	isotope	Z	T	Gamma Energy	Gamma Energy Value	Sigma	Sigma Value	Uncertainty	Knaught	Knaught value	Knaught unc
31611	238	239-U	92	p	4118.54(5)	4118.54	0.00148(15)	0.00148	0.00015	1.88E-05(19)	1.880000e-05	1.910000e-06
31612	238	239-U	92	p	4612.40(5)	4612.40	0.0031(3)	0.00310	0.00030	3.9E-05(4)	3.950000e-05	3.820000e-06
31613	238	239-U	92	p	4660.62(5)	4660.62	0.0034(3)	0.00340	0.00030	4.3E-05(4)	4.330000e-05	3.820000e-06
31614	238	239-U	92	p	4672.59(5)	4672.59	0.00116(13)	0.00116	0.00013	1.48E-05(17)	1.480000e-05	1.660000e-06
31615	238	239-U	92	p	4806.38(5)	4806.38	7E-05(7)	0.00007	0.00007	9E-07(9)	8.910000e-07	8.910000e-07

Step 3: Bayesian Statistics

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$Posterior = \frac{Likelihood * Prior}{Marginal}$$



Step 4: Define Priors

- A prior tells the code that something will always lie within a certain range or at a specific number
- Potential automation– have code “learn” priors

```
######    Means    ######
    if N == 1:

        pass

    else:

        for i in range(N-1):

            if means[i] > means[i+1]:

                return -np.inf

    for i in range(N):

        if means[i] > E[-1] or means[i] < E[0]:

            return -np.inf
```

Step 5: Markov Chain Monte Carlo

- Primarily used for numerical approximations of multi-dimensional integrals

- Create arbitrary starting points for “walkers” a sufficient distance from each other
  - “walkers” spend most time in regions of higher probability
  - Produce a cluster of points where the posterior probability is largest

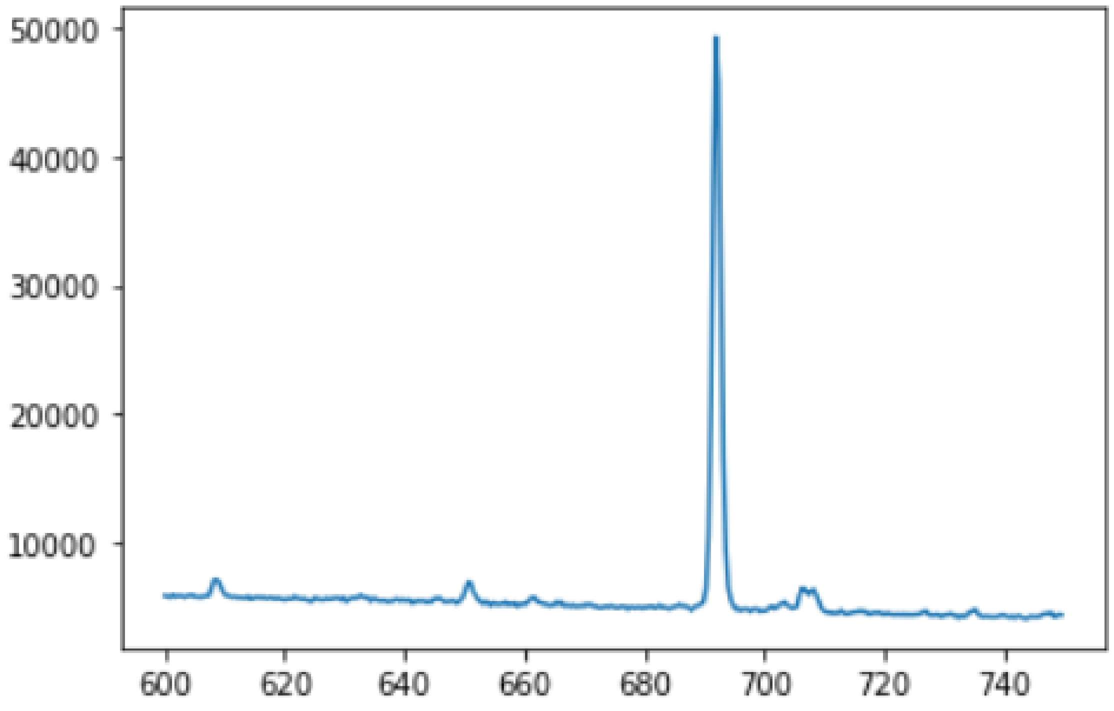
- These points are used to compute expected values efficiently

Step 5:

Initialization

- Slope
- Background bias
- Peak location (mean)
- Amplitudes
- Peak widths
- Background noise

Then setting 150 Walkers to each take 1000 steps

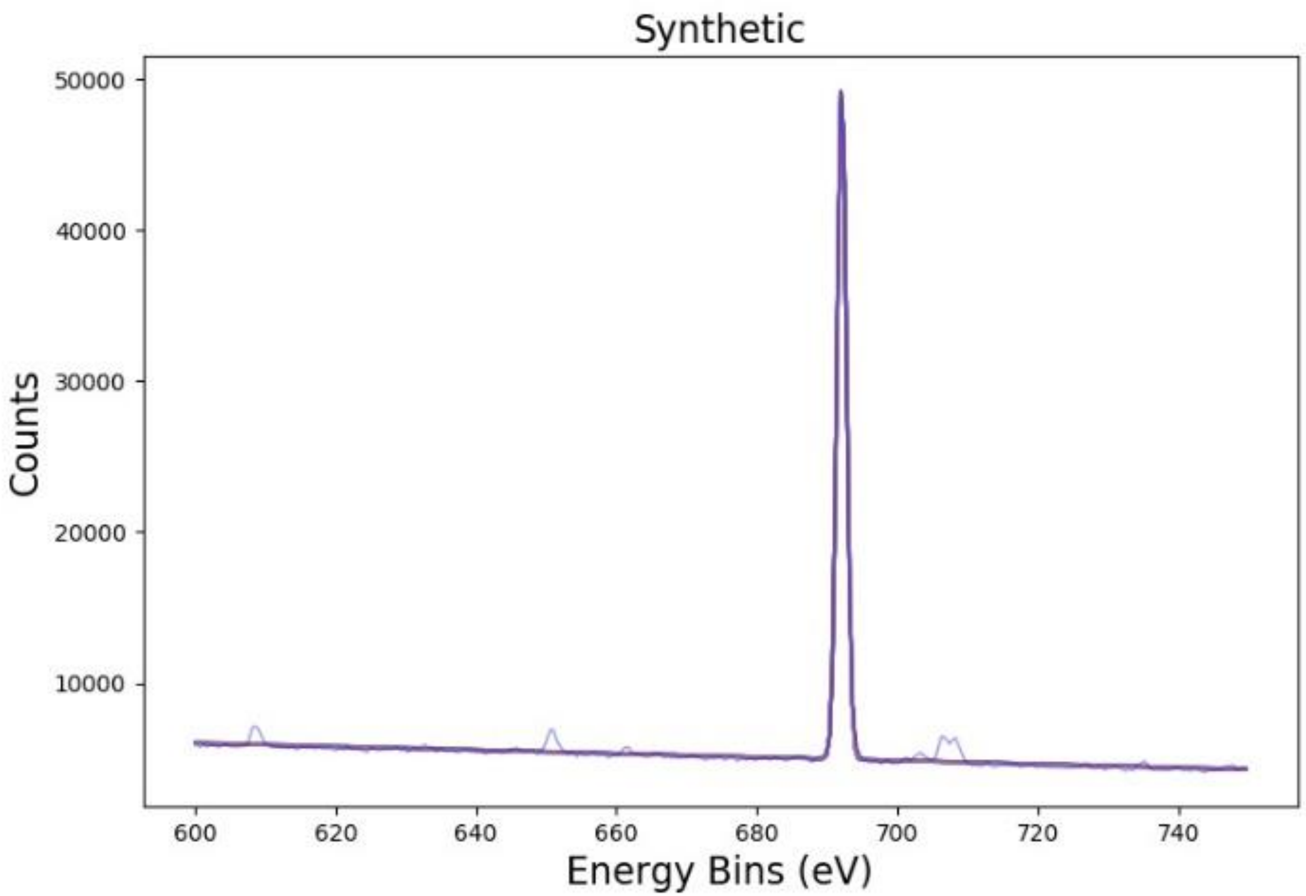


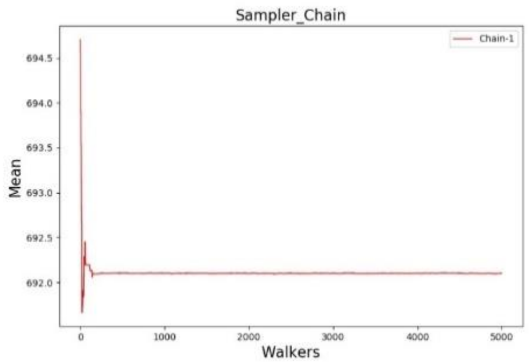

```
What value should I use for mean?1 (done to quit) 692
What value should I use for mean?2 (done to quit) done
Number of Parameters:
6
```

Step 5:

- Overlay of original plot and a “Synthetic” consisting of 10 walkers' paths

- The goal is for the walkers to fit the original path perfectly purely based on the priors



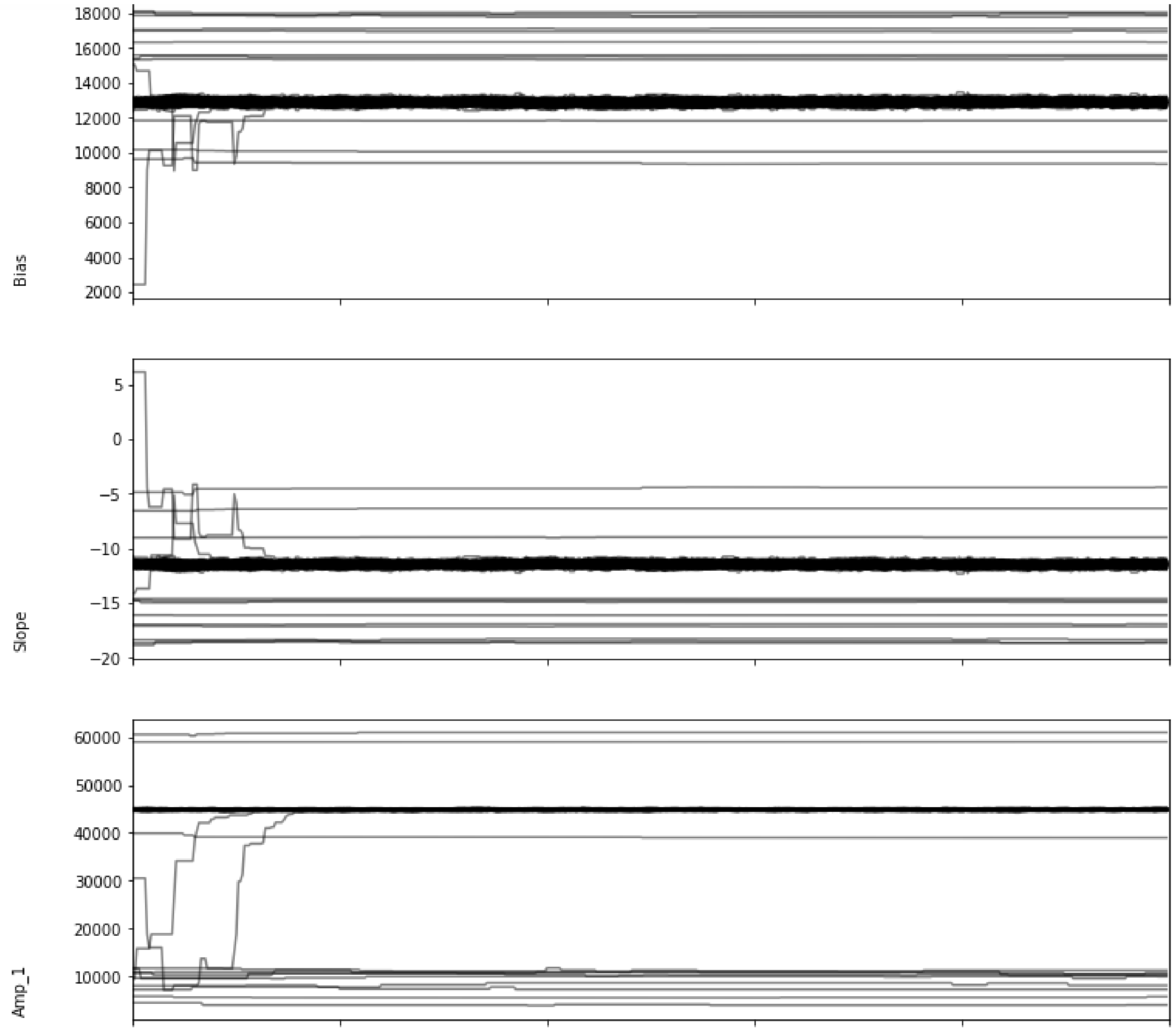


Step 6:

Sanity Check/Distribution

- Darker the lines mean higher cluster of walkers.
- All other lines represent paths a walker takes.

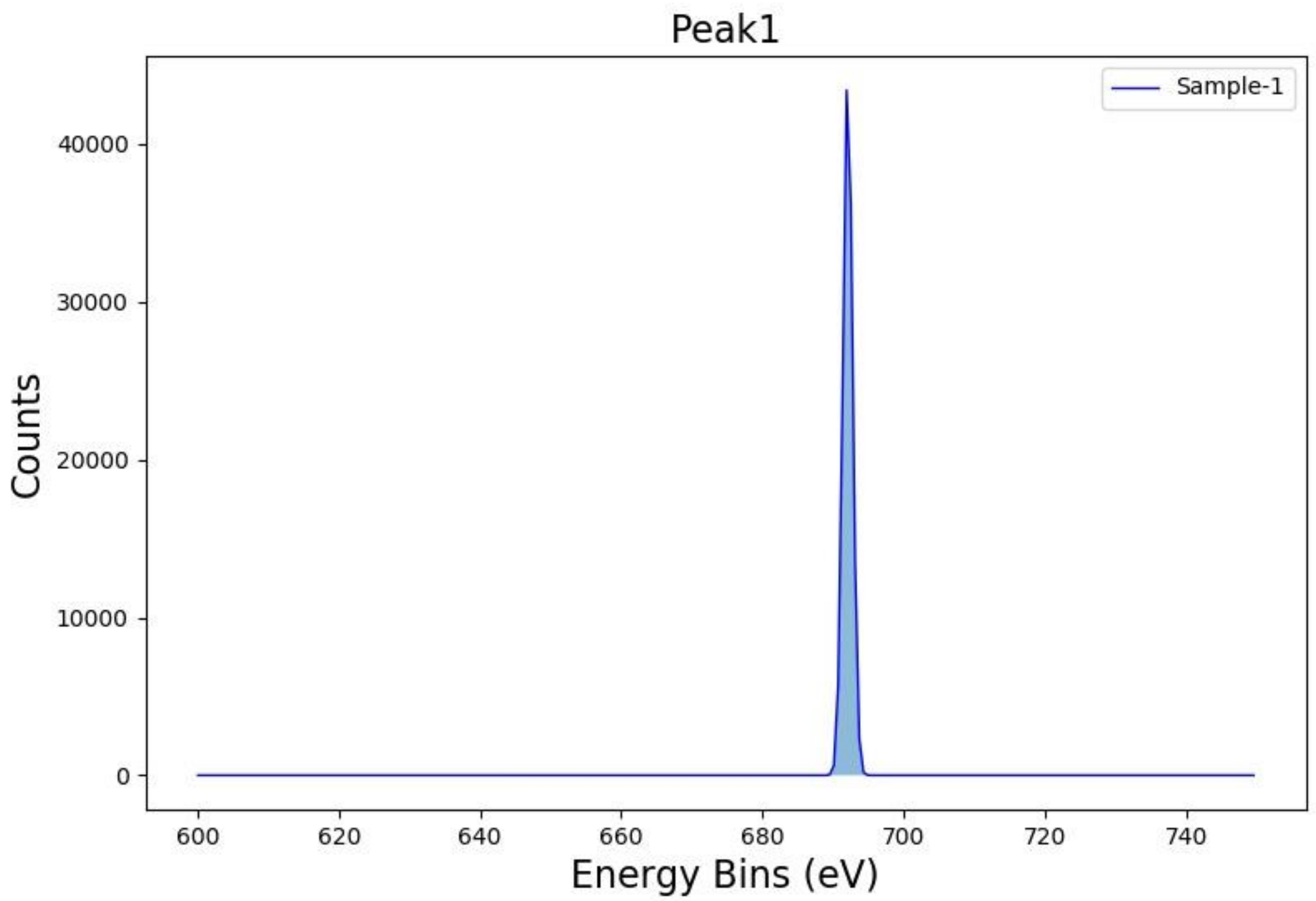
- There should be a visible trend that the walkers are trying to merge to a singular point



Step 7:

Plotting

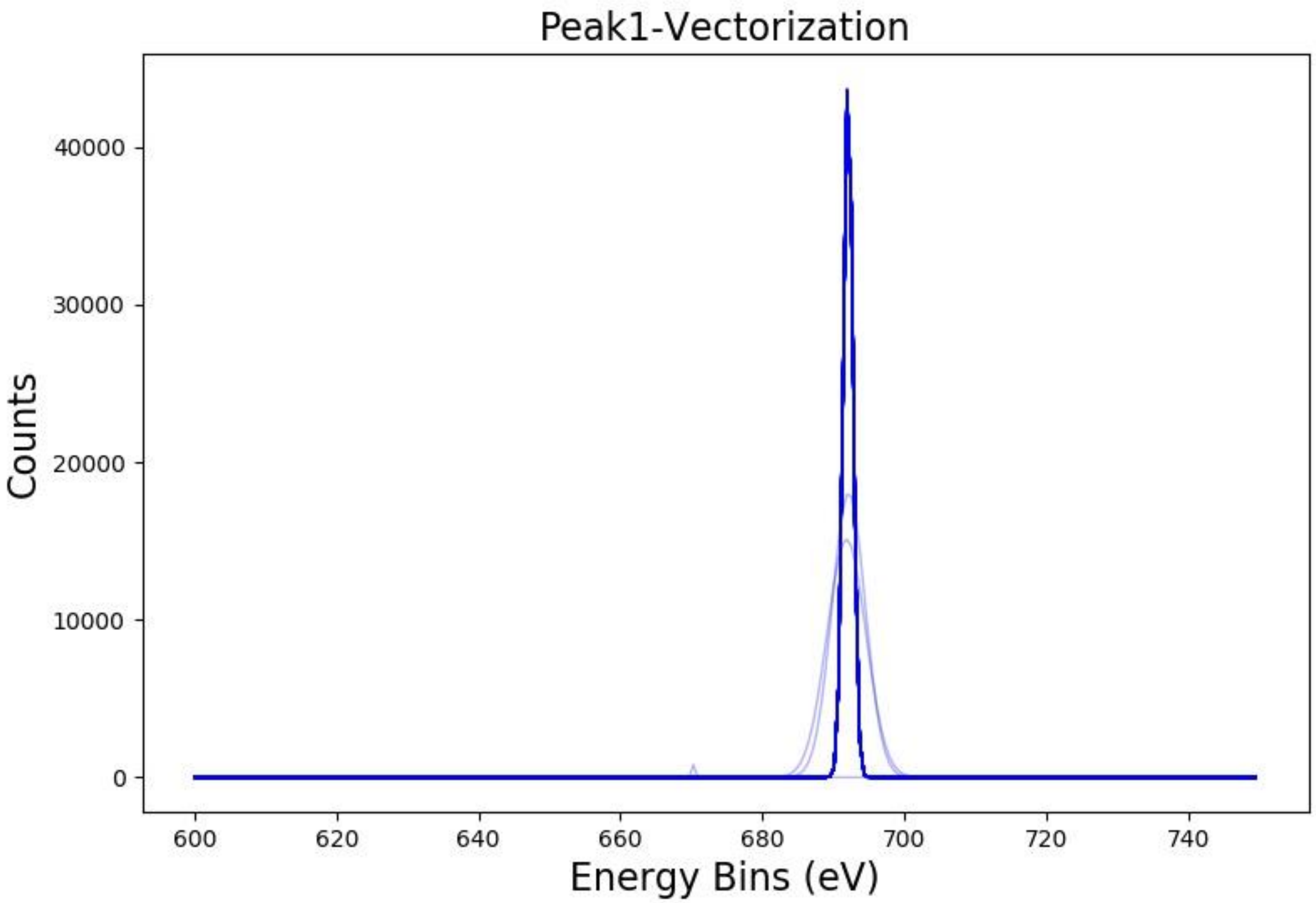
- This is now a fit of the peak of interest.



Step 7:

Vectorization

- This plot depicts a gradient for which walkers are fitting the peak, the darker lines meaning more walkers on that path, lighter meaning less. Fewer lines meaning less variation



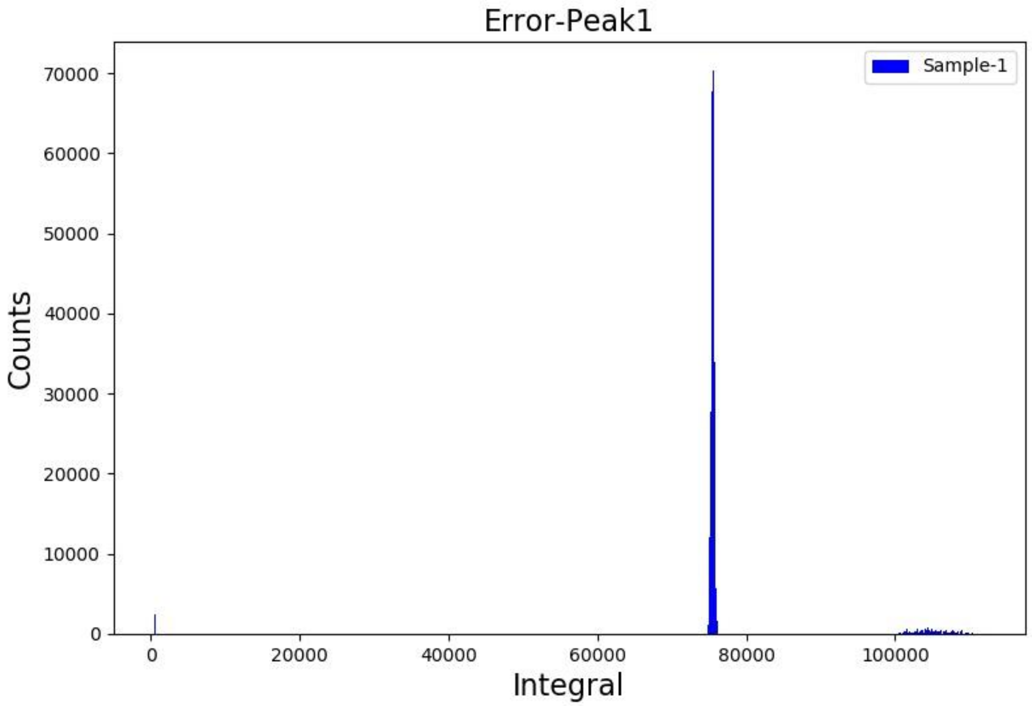
Step 7:

Error

- Y-axis, number of samples saying this is the right integral.

- X-axis, area under the curve.

- Then computing the error from our bounds we get great results.




```
print('error in P1 lower:',((((fp-sp)+fp)-fp)/fp)*100)
print('error in P1 upper:',((((ep-fp)+fp)-fp)/fp)*100)
print()
# print('error in P2 lower:',((((fp1-sp1)+fp1)-fp1)/fp1)*100)
# print('error in P2 upper:',((((ep1-fp1)+fp1)-fp1)/fp1)*100)
```

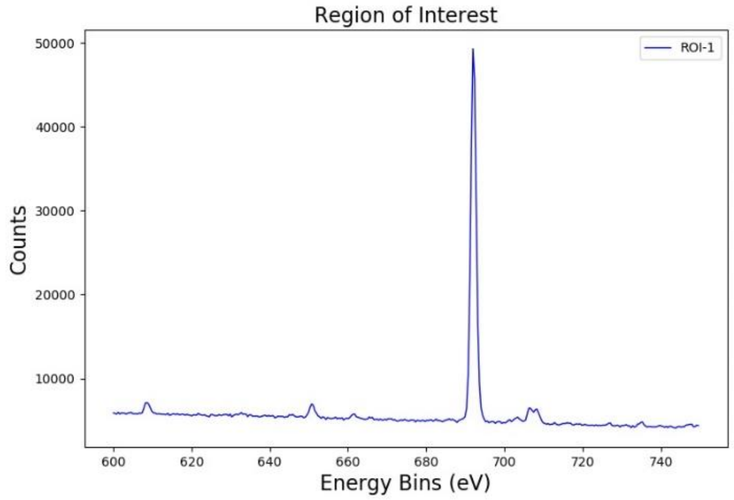
```
error in P1 lower: 0.2869768750046384
error in P1 upper: 0.30498320321024125
```

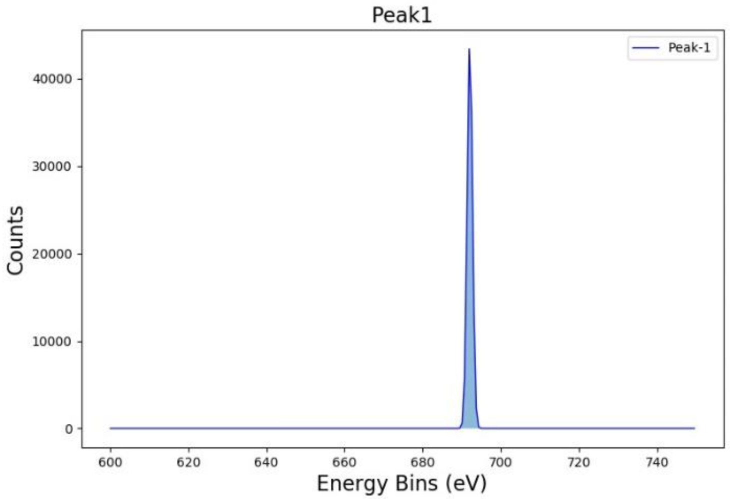
Applications of Bayesian Approach

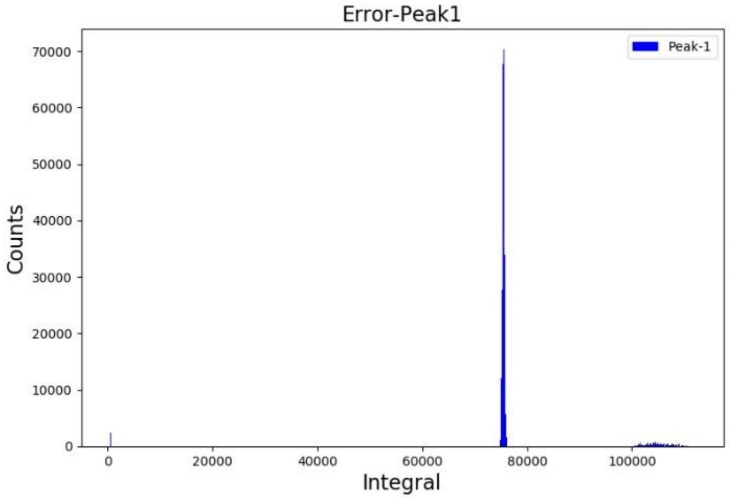
- Primarily used for numerical approximations of multi-dimensional integrals
  - Single, multi, overlapping peaks
  - Custom solutions possible

- Limitations and sources of error

Single Peak

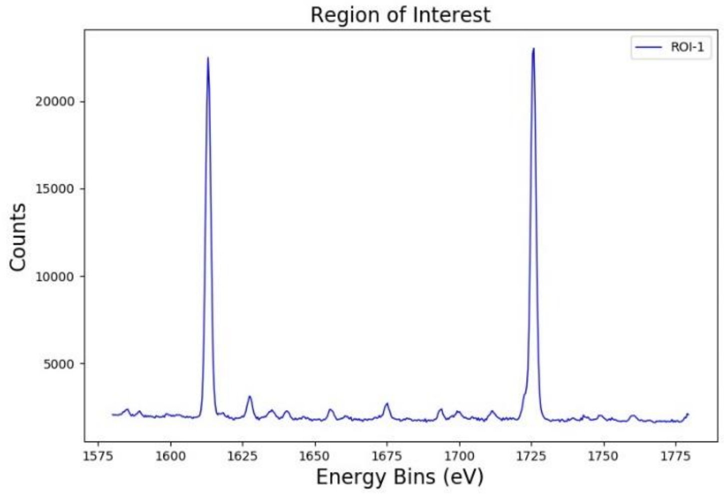


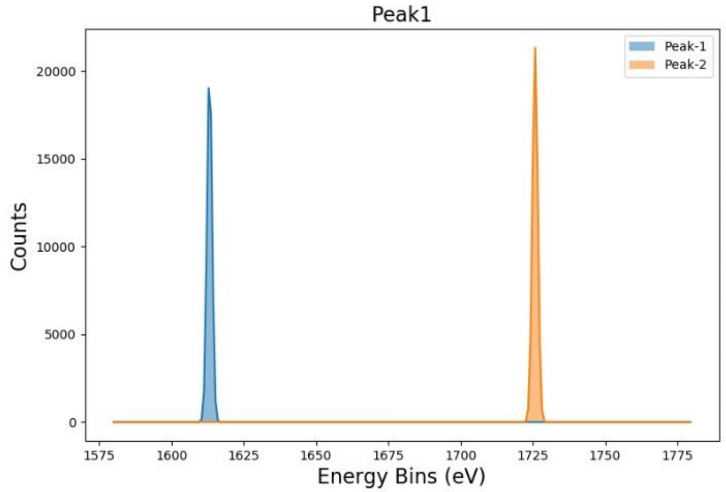


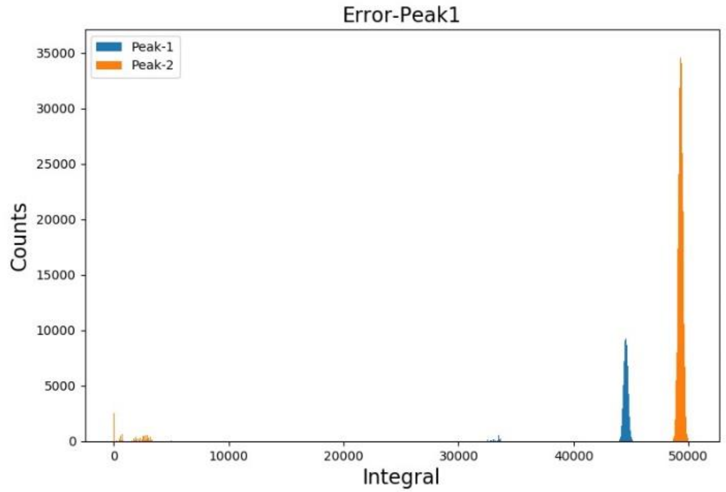



```
error in P1 lower: 0.28697
error in P1 upper: 0.30498
```

Multi Peak



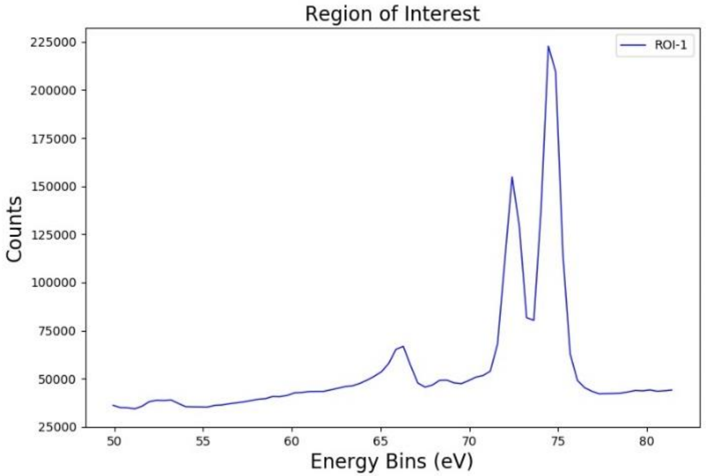


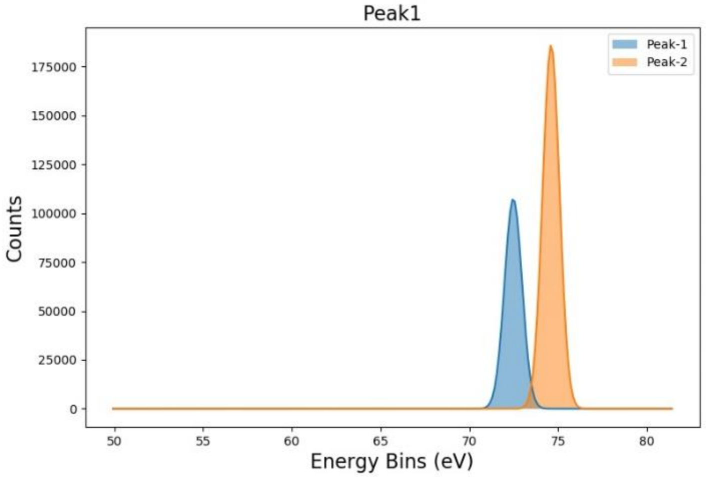


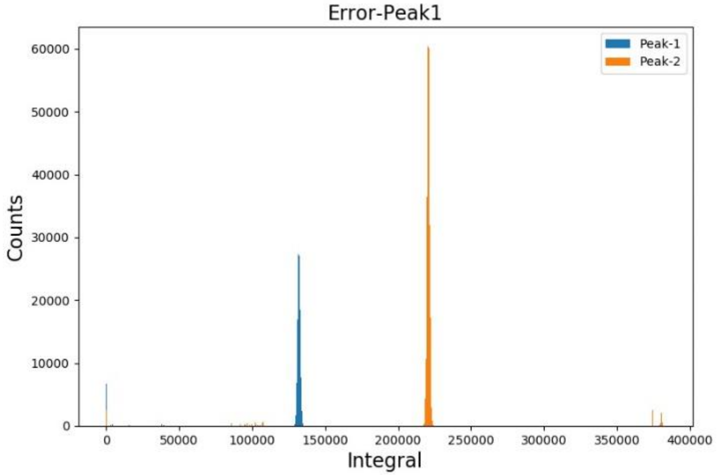

```
error in P1 lower: 0.43563
error in P1 upper: 0.42188

error in P2 lower: 0.40227
error in P2 upper: 0.40272
```

Overlapping Peaks








```
error in P1 lower: 0.71501
error in P1 upper: 0.8930

error in P2 lower: 0.5486
error in P2 upper: 0.42021
```

Notes & Next Steps

- What can we do to help predict the shape of ugly peaks?
- Data Sampling: provide hundreds of previous spectra to help code “learn”
- Define priors in a generalized way

Conclusions

- Decrease in error
- Assists reading of spectra
- Flexible approach
- Beneficial to PGAA field

Thank you! Questions?

- Ryan Lester
  - ryanlester10@sbcglobal.net

- Cole Thompson
  - cole.thompson02@utexas.edu