Lecture 5:Uncertainty

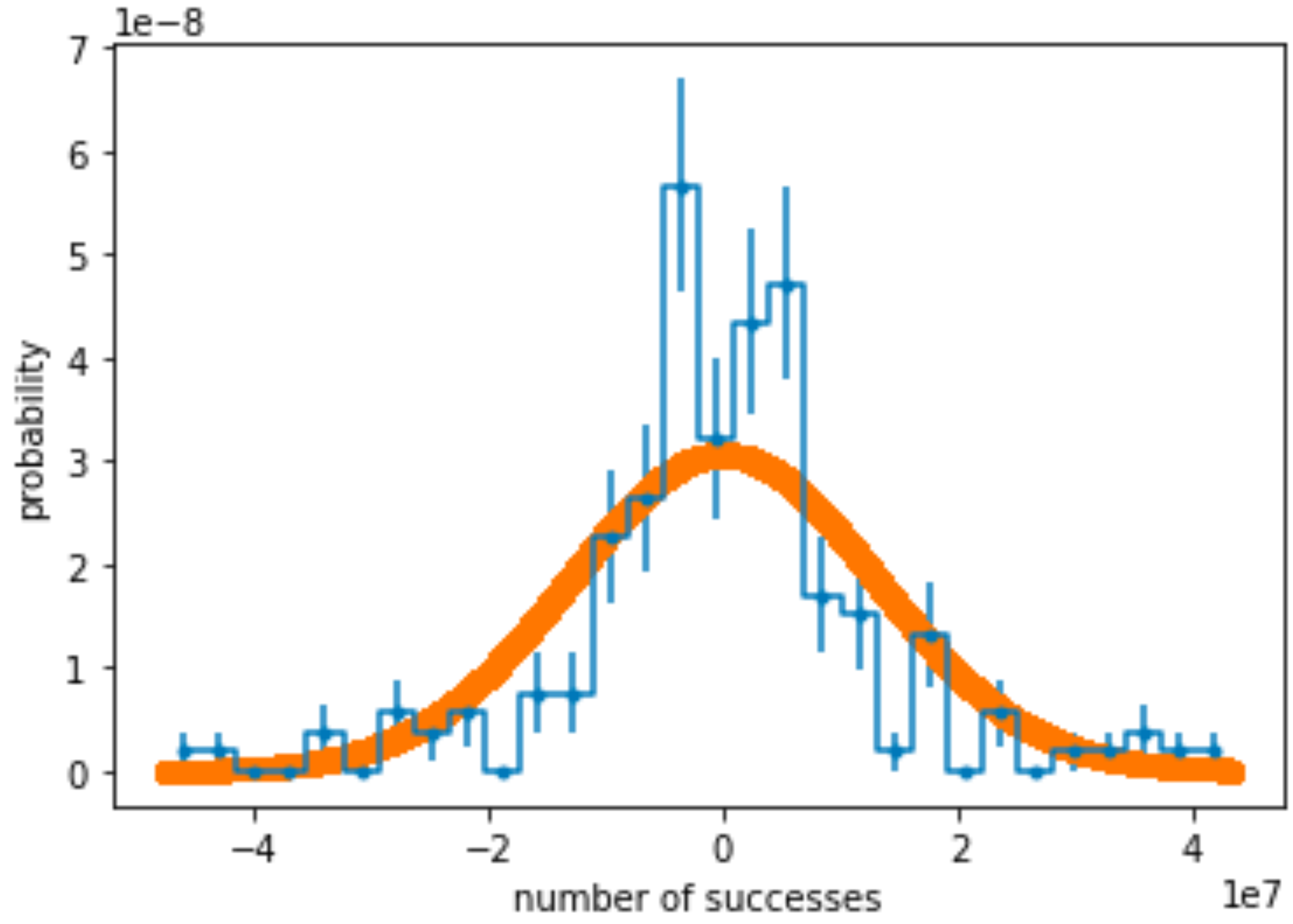
Uncertainty

What are we fitting?

- Lets say a fitted parameter has a distribution
  - We would like to quote the uncertainty on that distribution
  - Traditionally we do this by computing the standard deviation
  - Standard Deviation $\sigma = \sqrt{Var[\theta]}$
  - Standard deviation is just a number
  - It does not reflect what the actual distribution was about

How do we deal with unc?

- Our uncertainties have the same standard deviation
  - You may know there is an implicit assumption its Gaussian



Likelihood

Hikers in Scotland

A mathematician, a physicist, and an engineer are hiking through Scotland.

The engineer looks out, sees a black sheep, and exclaims, "Hey! They've got black sheep in Scotland!"

The physicist looks out and corrects the engineer, "Strictly speaking, all we know is that there's at least one black sheep in Scotland."

The mathematician looks outand corrects the physicist, " Strictly speaking, all we know is that is that at least one side of one sheep is black in Scotland."

What is Likelihood?

- Lets take an example:
- 40% of all humans have smelly pee after they eat asparagus
  - Ref: Scientific American Article.
- If you have a room with 56/100 people who have smelly pee
- What is the probability?
-

What is Likelihood?

- Lets take an example:
- 40% of all humans have smelly pee after they eat asparagus
    - Ref: Scientific American Article.
- If you have a room with 56/100 people who have smelly pee
- What is the probability? (binomial)

$$P(56 \,|\, p = 40\%) = p^{56}(1-p)^{44}\frac{100!}{44!56!}$$

- What is the Likelihood of this happening? (given 40%)

What is Likelihood?

- Lets take an example:
- 40% of all humans have smelly pee after they eat asparagus
    - Ref: Scientific American Article.
- If you have a room with 56/100 people who have smelly pee
- What is the probability? (binomial)

$$P(56\,|\,p = 40\%) = p^{56}(1-p)^{44}\frac{100!}{44!56!} \qquad 4\text{x}10^{-4}$$

- What is the Likelihood of this happening? (given 40%)

What is Likelihood?

- Lets take an example:
- 40% of all humans have smelly pee after they eat asparagus
  - Ref: Scientific American Article.
- If you have a room with 56/100 people who have smelly pee
- What is the probability? (binomial)

$$P(56\,|\,p = 40\%) = p^{56}(1-p)^{44}\frac{100!}{44!56!} \qquad 4\text{x}10^{-4}$$

- What is the Likelihood of this happening? (given 40%)

$$\mathscr{L}(56\,|\,p = 40\%) = p^{56}(1-p)^{44}\frac{100!}{44!56!} \qquad 4\text{x}10^{-4}$$

What is the Max Prob/Like?

- The maximum probability of this distribution occurs at k=40:

$$P(40\,|\,p = 40\%) = p^{40}(1-p)^{60}\frac{100!}{40!60!} = 0.08$$

- The maximum likelihood occurs at p=0.56:

$$\mathscr{L}(56\,|\,p = ?) = p^{56}(1-p)^{44}\frac{100!}{44!56!}$$

- Likelihood we vary the underlying parameters to likelihood
- Probability we vary our tests to maximize probability

What is the Max Prob/Like?

- The maximum probability of this distribution occurs at k=40:

$$P(40\,|\,p = 40\%) = p^{40}(1-p)^{60}\frac{100!}{40!60!} = 0.08$$

- The maximum likelihood occurs at *p*=0.56:

$$\frac{\mathcal{L}(56\,|\,p)}{dp} = 0 = \frac{d}{dp}\left(p^{56}(1-p)^{44}\frac{100!}{44!56!}\right)$$

$$56\left(p^{55}(1-p)^{44}\right) - 44\left(p^{56}(1-p)^{43}\right) = 0 \rightarrow 56(1-p) - 44p = 0$$

$$56 - 100p = 0 \rightarrow p = 56/100$$

- Likelihood we vary the underlying parameters of our distribution

- Probability we vary our tests to maximize the probability

What is the Log Likelihood?

- The log like of the likelihood:

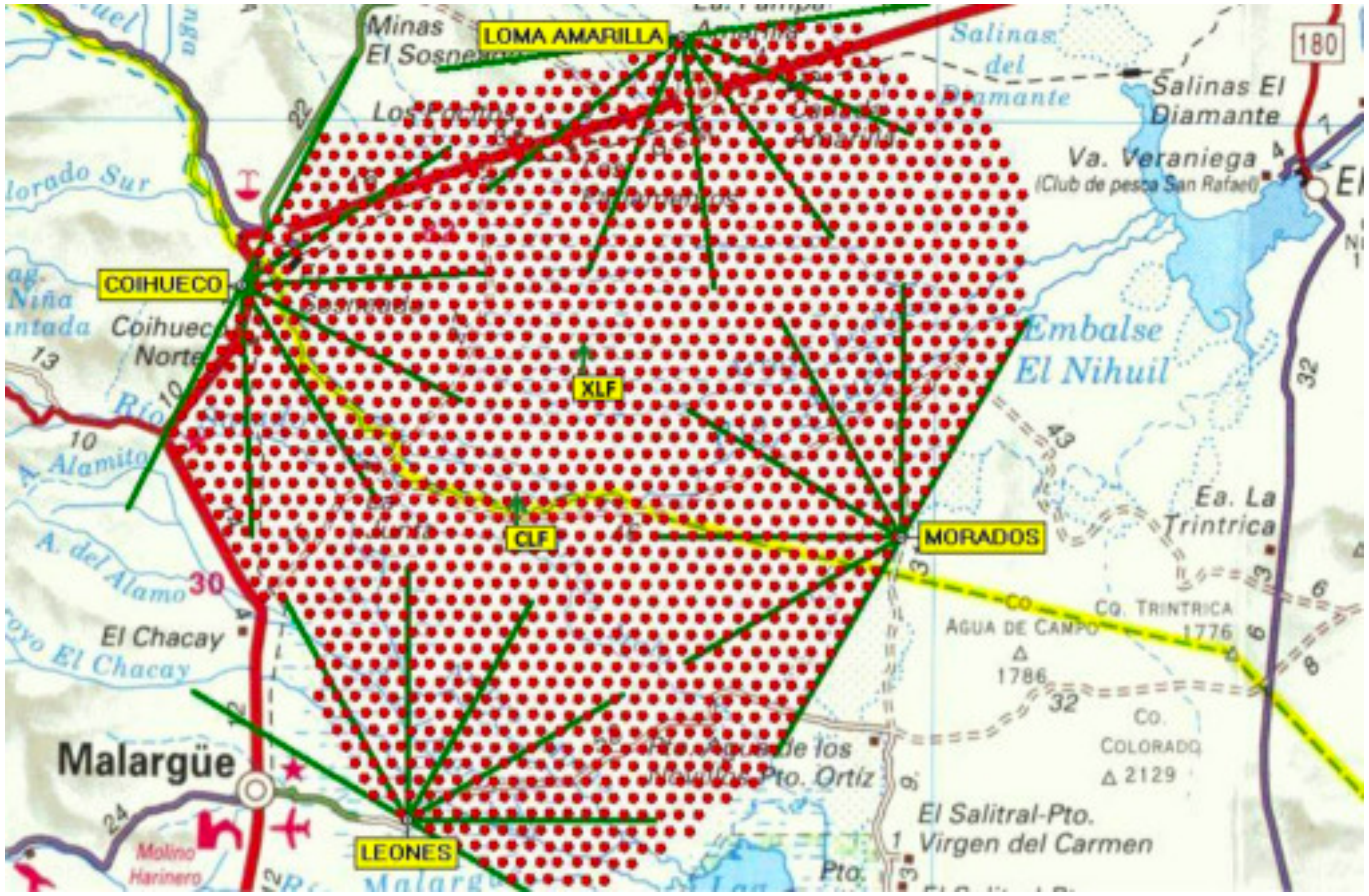
$$\log(\mathscr{L}(56\,|\,p = 40\%)) = 56\log(p) + 44\log(1-p) + \log\left(\frac{100\,!}{44\,!56\,!}\right)$$

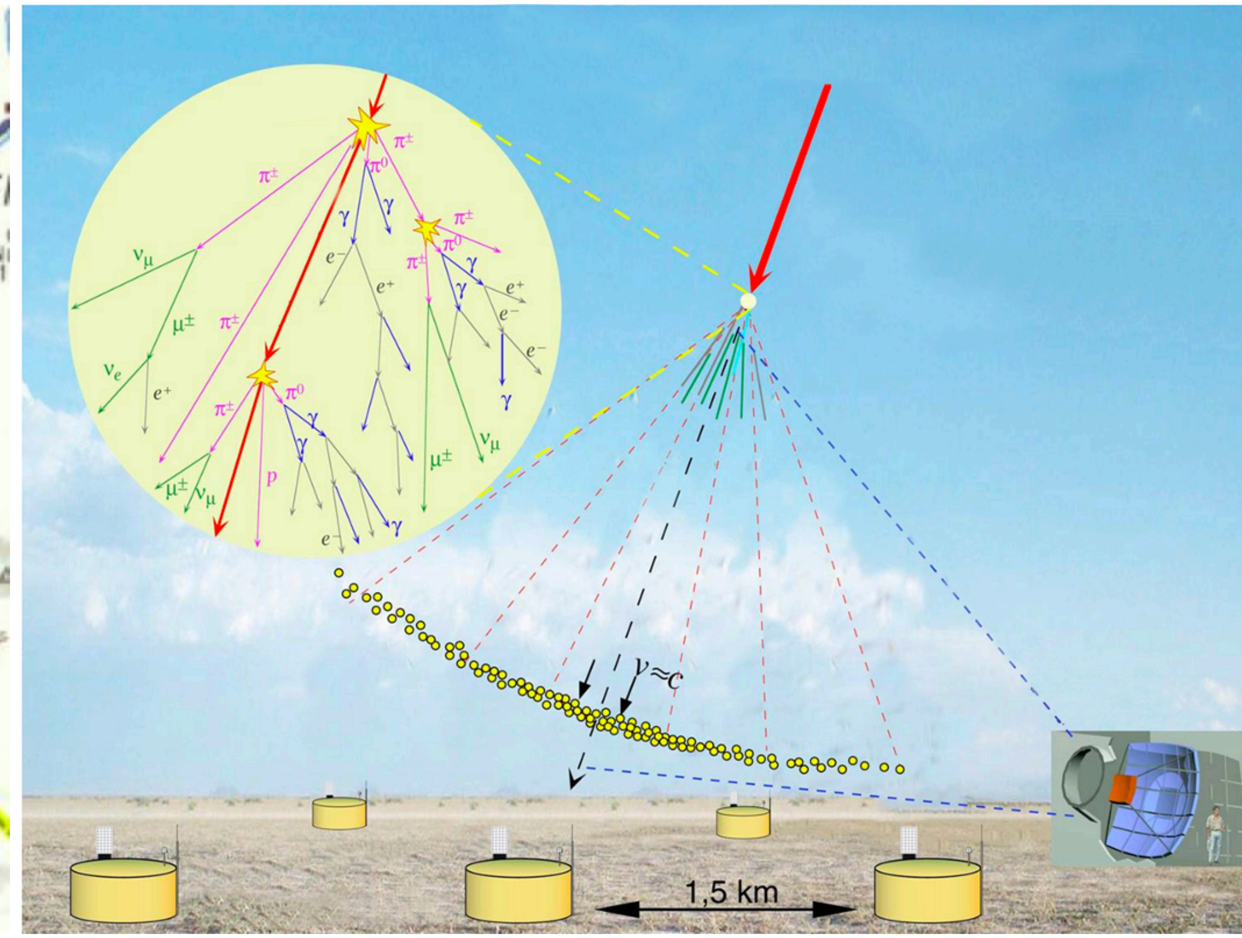
- Sometimes the likelihood can get very large
    - That's why we take the log
- Log is also a 1 to 1 mapping so

$$\max\left(\log(\mathscr{L})\right) = \max \mathscr{L}$$

Auger Experiment

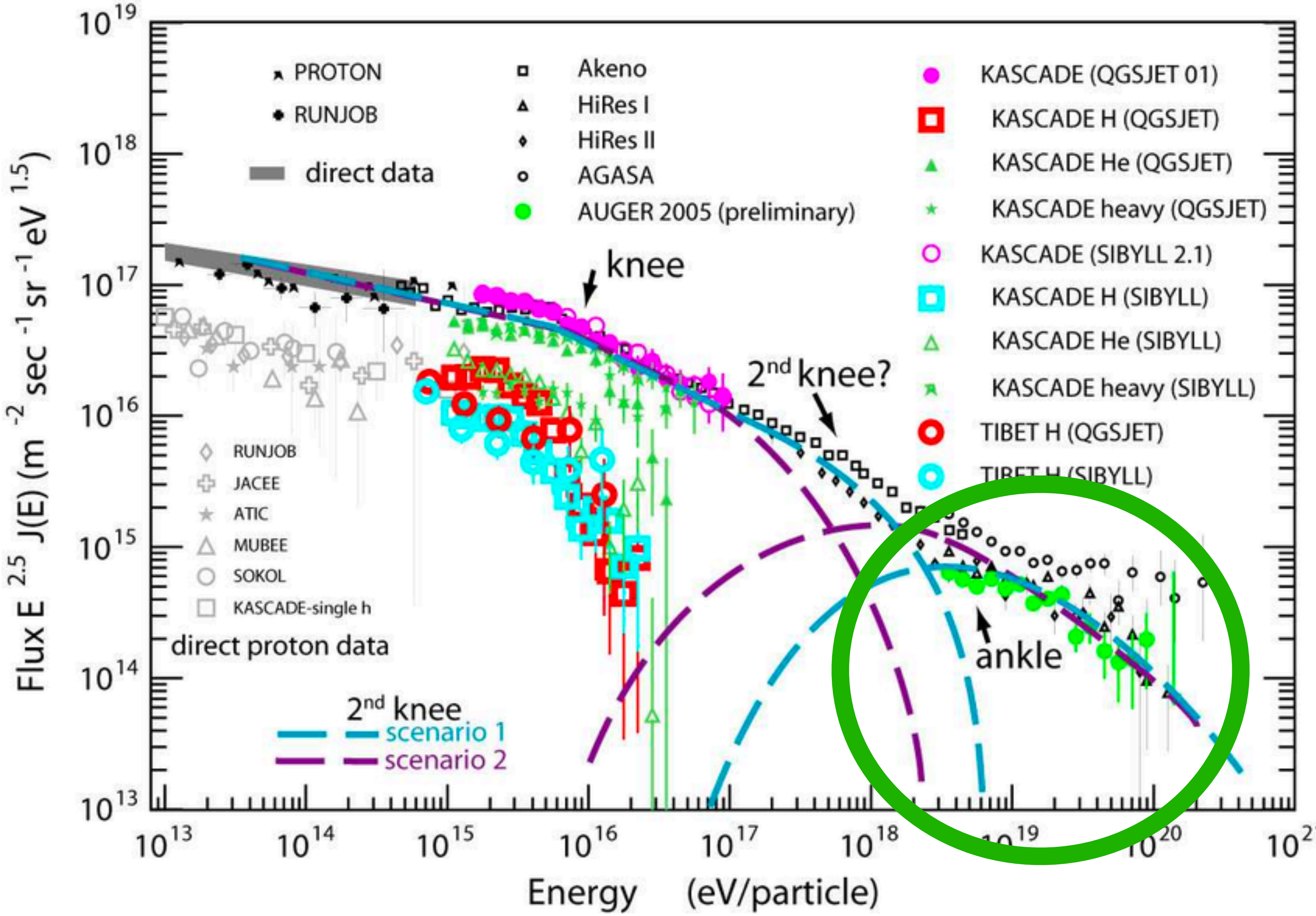
Auger Experiment





- Experiment is a giant array of detectors
- These detectors reconstruction particle showers
  - Particles are the highest energy particles in the world
- In the end you get a position and energy of each particle

What we learn w/Auger?



- We probe the highest energy particles in the world w/Auger

  - Nobody knows where they come from!

Auger Data

- Data can be found here :
  - https://www.auger.org/index.php/science/data
- We are going to analyze data from here :
  - https://www.auger.org/index.php/document-centre/viewdownload/115-data/4643-ra-data-event-lists-flux-map
- Here is the paper on it:
  - Science 357 (2017) 1266-1270
  - https://arxiv.org/abs/1709.07321

Auger Data

- Data format:

Two Datasets

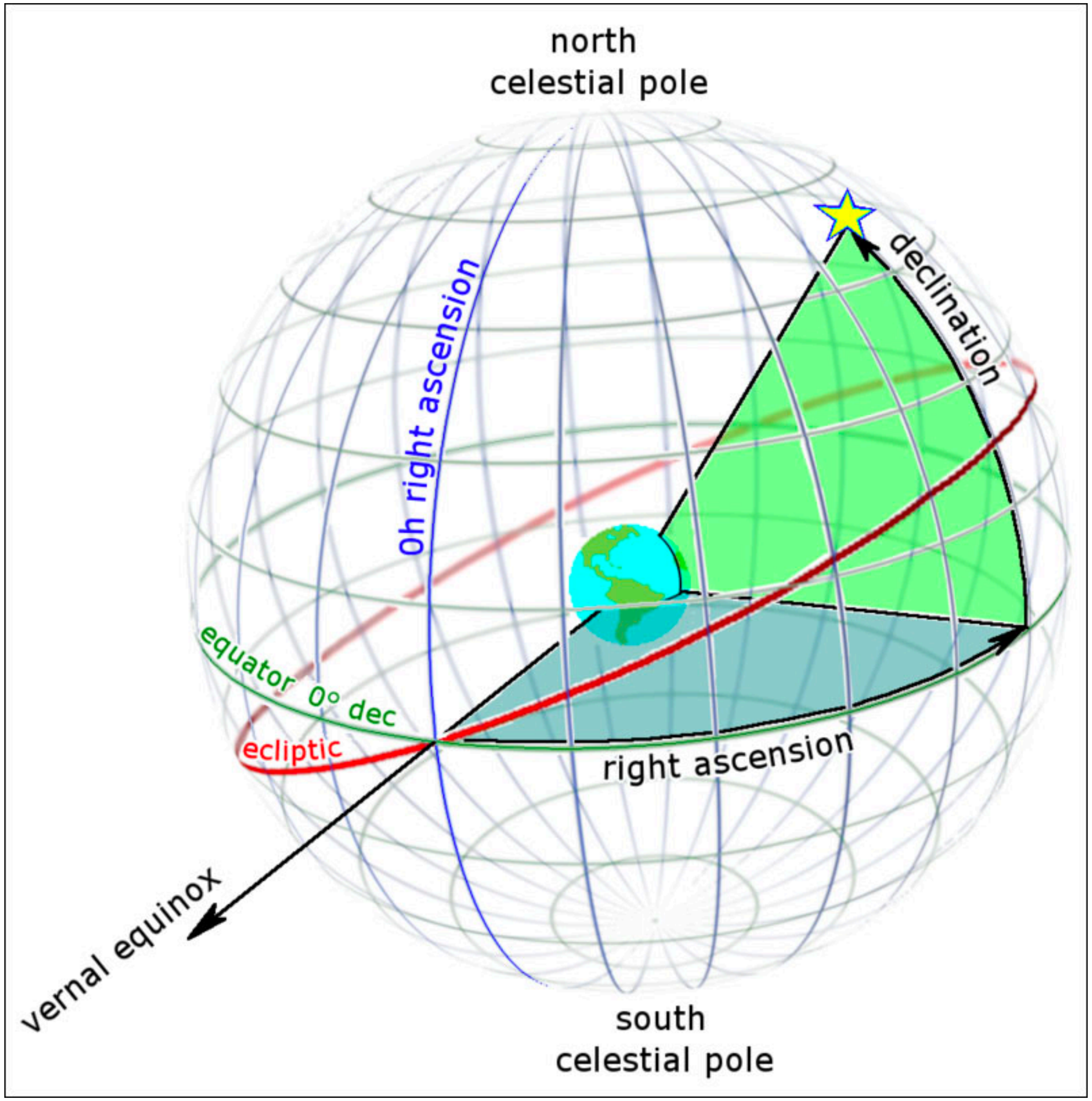
```
events_a4 : Particles with energy from 4-8 EeV
events_a8 : Particles with energy  > 8 EeV
EeV = Exa Electron Volt = 10^18 eV
```

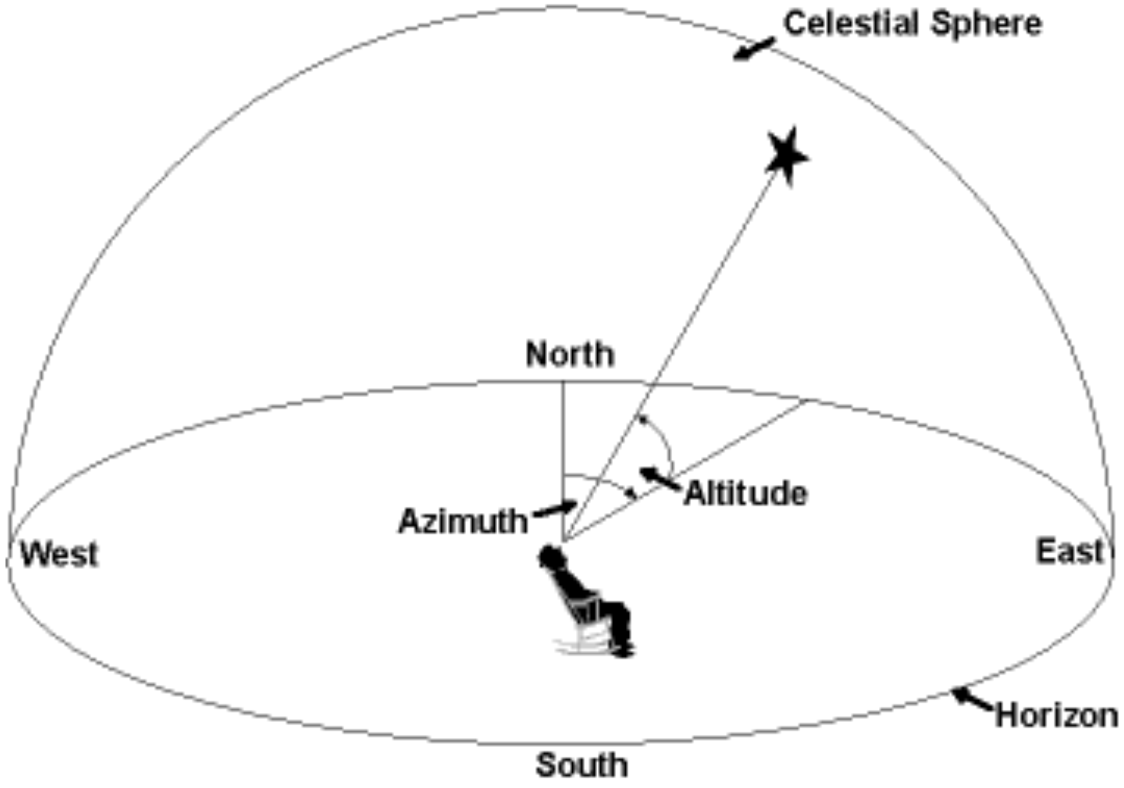
```
##
## Columns:
#
#  (1) year
#  (2) day
#  (3) declination (deg.)
#  (4) right ascention (deg.)
#  (5) local azimuth (deg.)
#  (6) weight
#
# year day   dec  RA  azimuth weight
#
2004    1   -11.8   208.9   164.7   1.0040
2004    2   -70.4    36.6  -111.6   1.0001
2004    3   -47.8   334.7  -143.8   0.9973
2004    3   -22.1    87.2  -162.8   1.0080
2004    7    -5.8   178.6   162.7   1.0046
2004    9     3.4    60.5    82.9   1.0072
2004   12     3.6   248.3   126.3   1.0007
```

Earth Coordinates in Equatorial form

Uncertainty correction for data
its close to 1 we will ignore it

Cosmic Rays

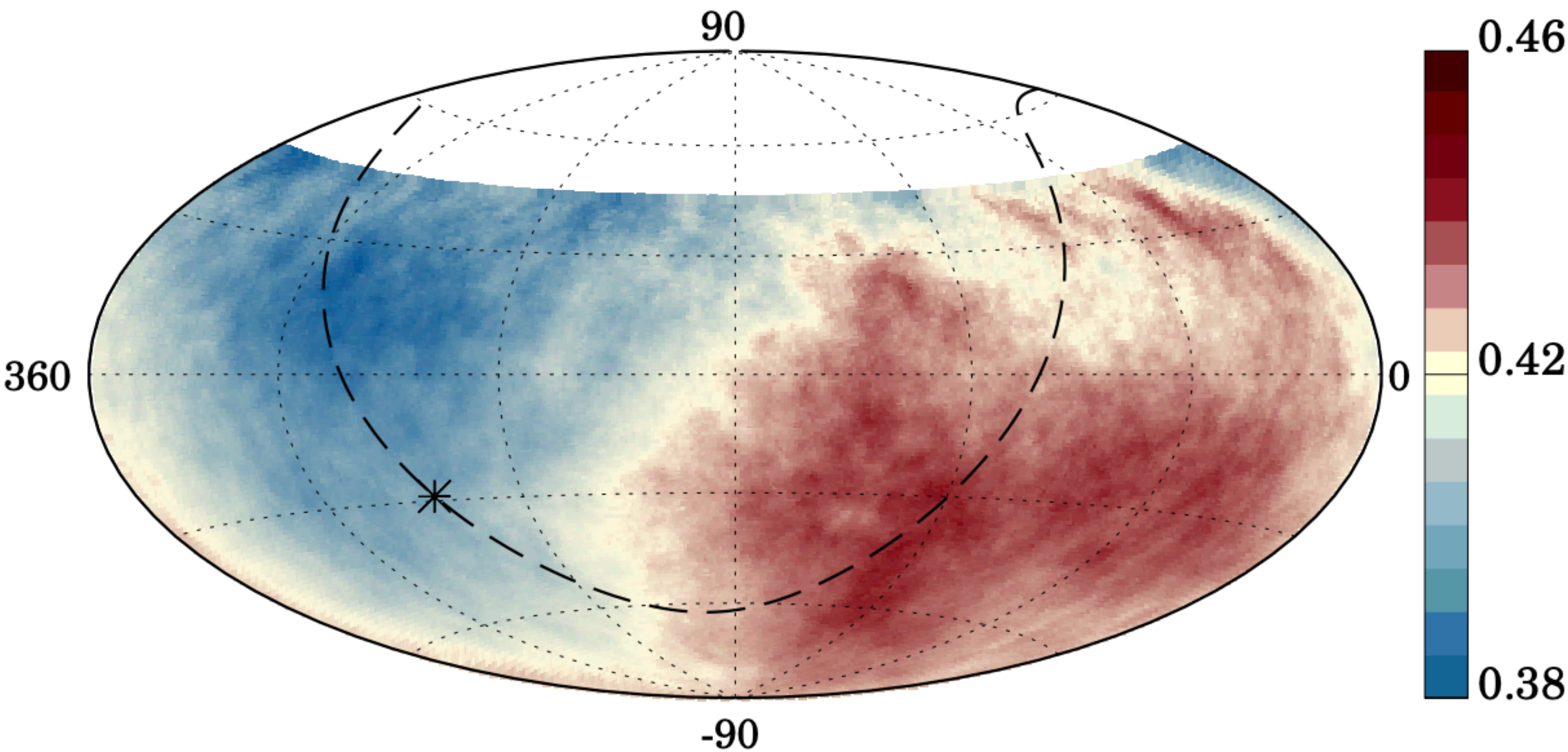




To understand Auger Data
We need to understand
Spherical Coordinates

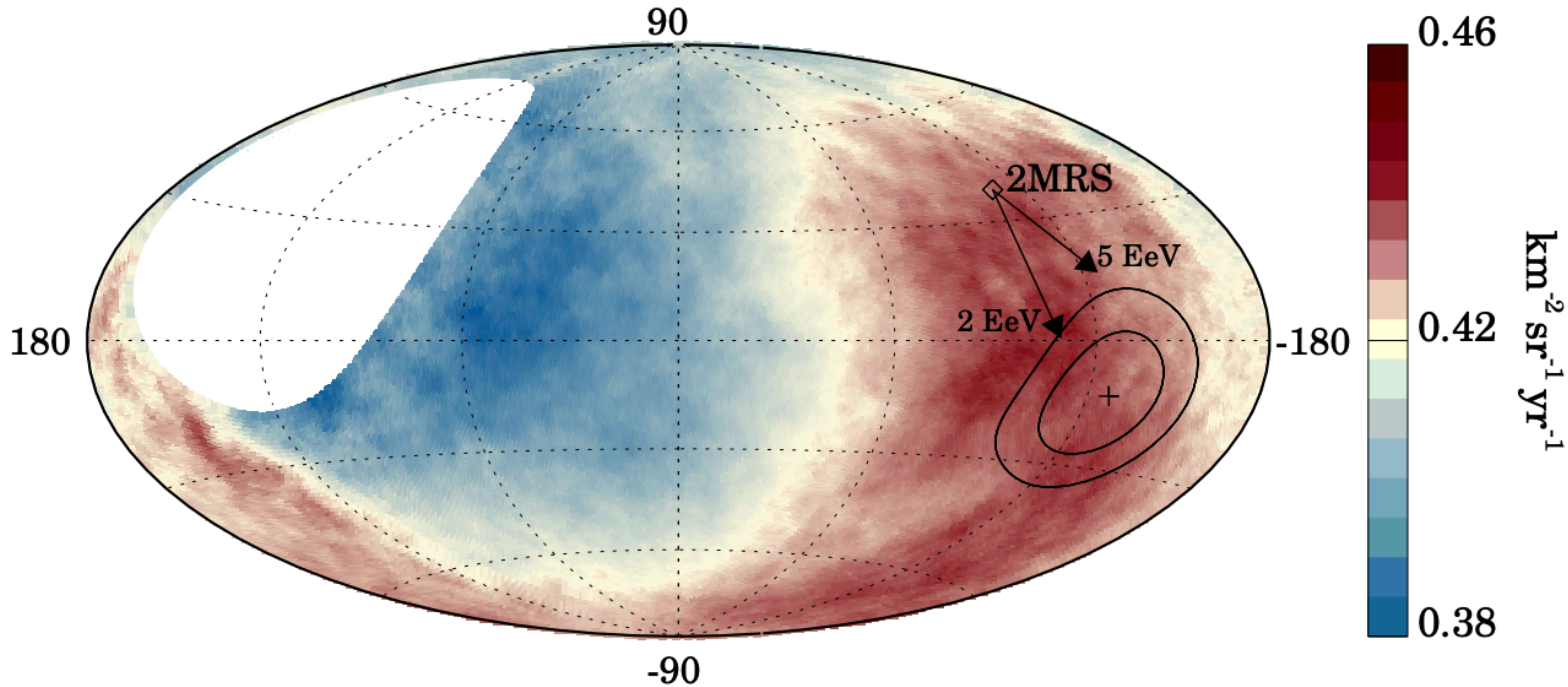
We also find

To get this you need to know details about acceptance
I have a guess in the notebook



https://arxiv.org/pdf/1012.3754.pdf

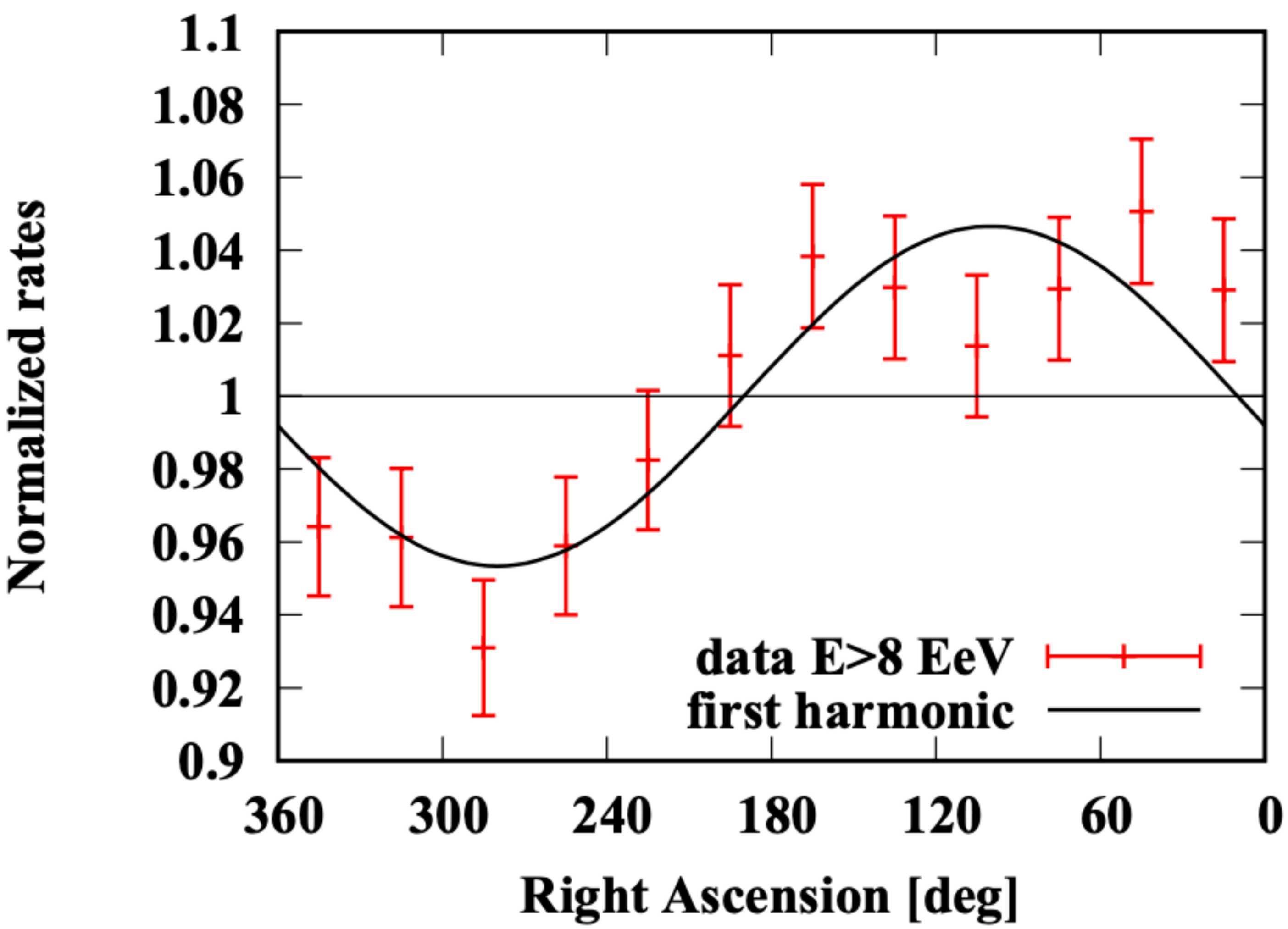
In Galatic Coordinates

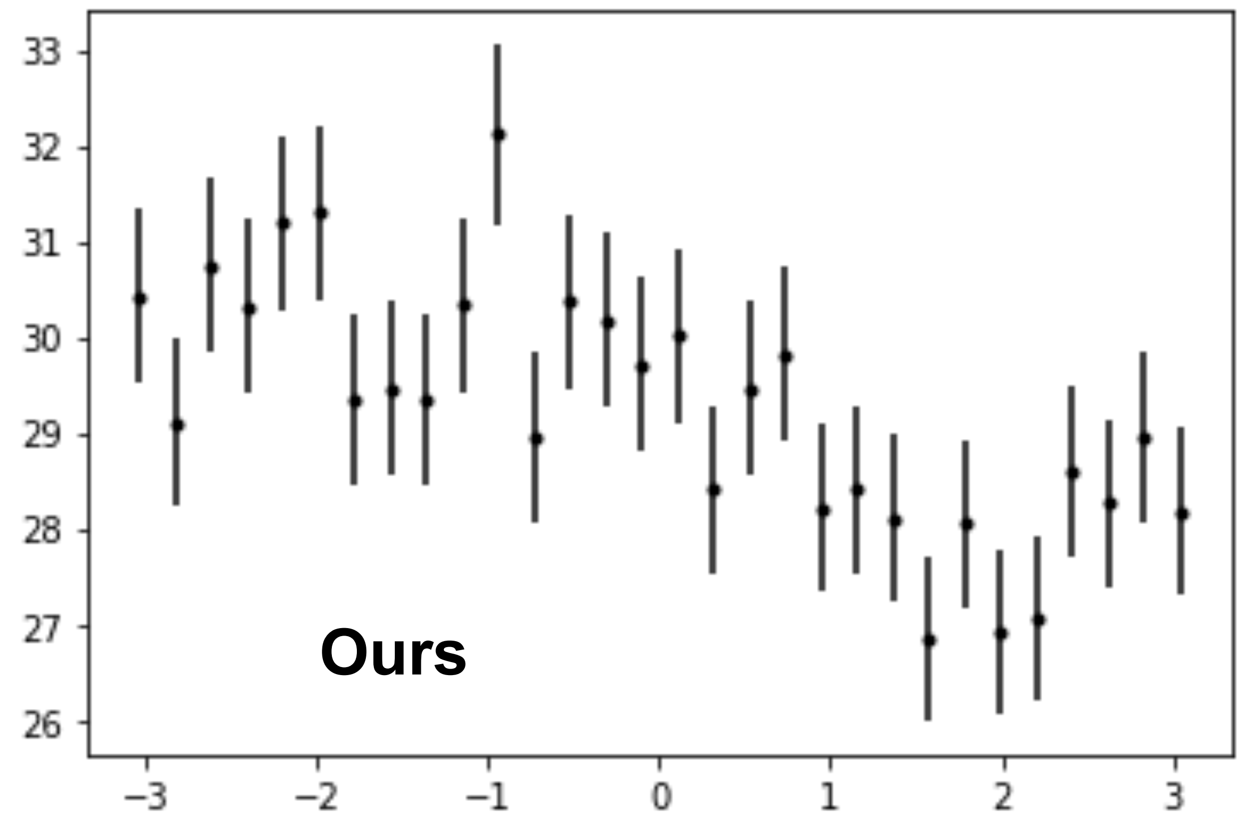


Challenge: Try to make this plot

https://arxiv.org/pdf/1012.3754.pdf

From the paper we find





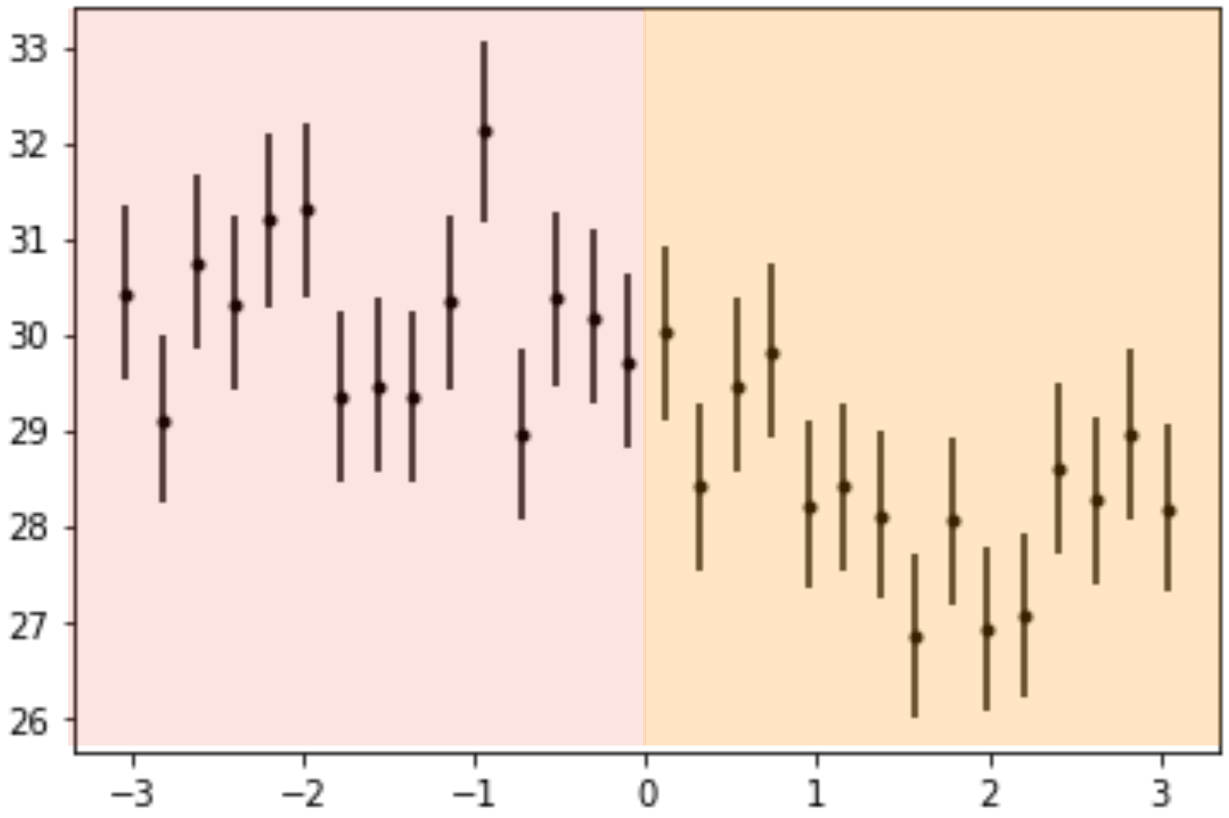
Our angle conversion is backward

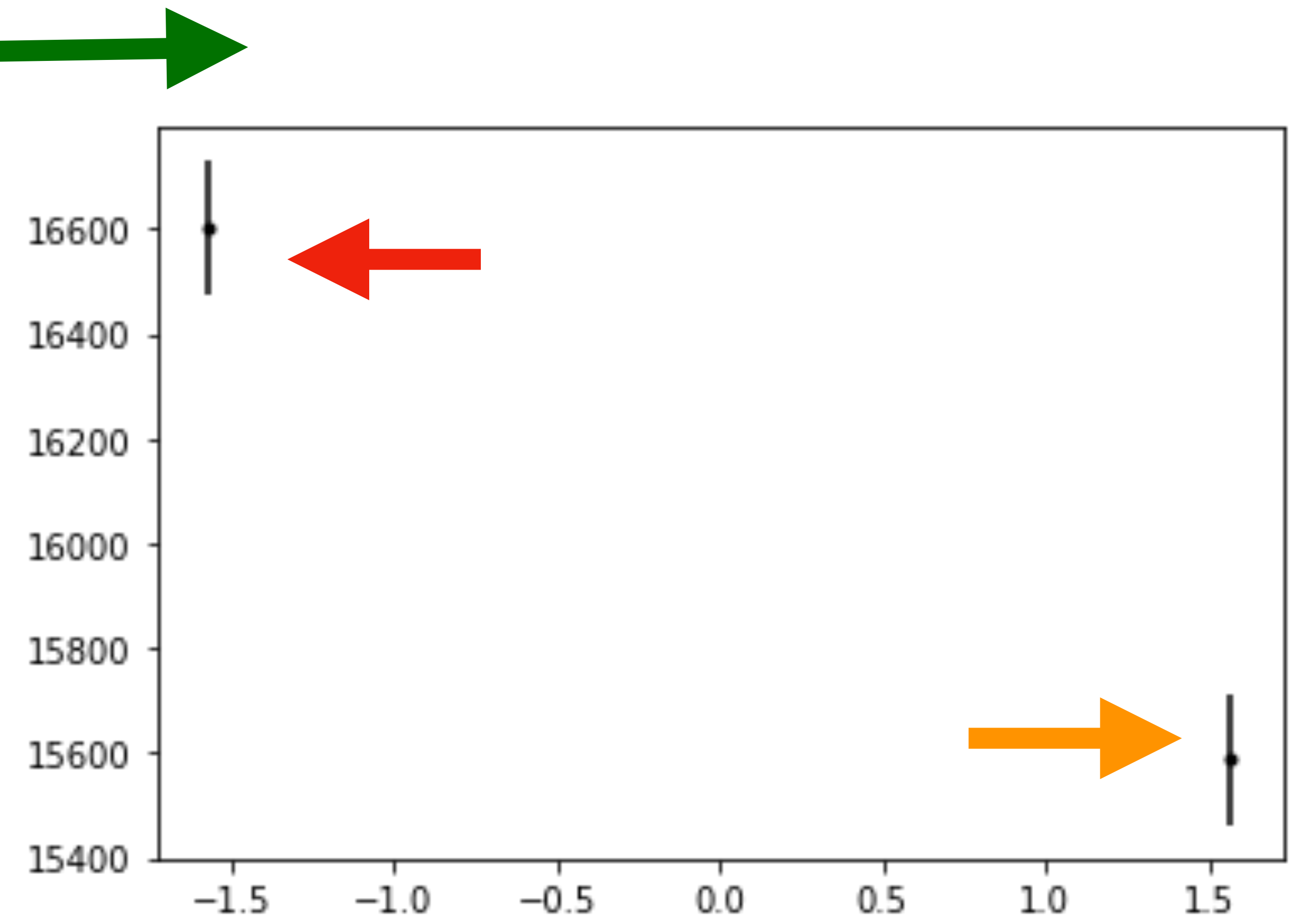
- An asymmetry in right ascension
- Our x-value is flipped and we didn’t normalize (same trend)

How do we quantify asymmetry?

Highest Energy > 8 EeV

First lets just make 2 bins





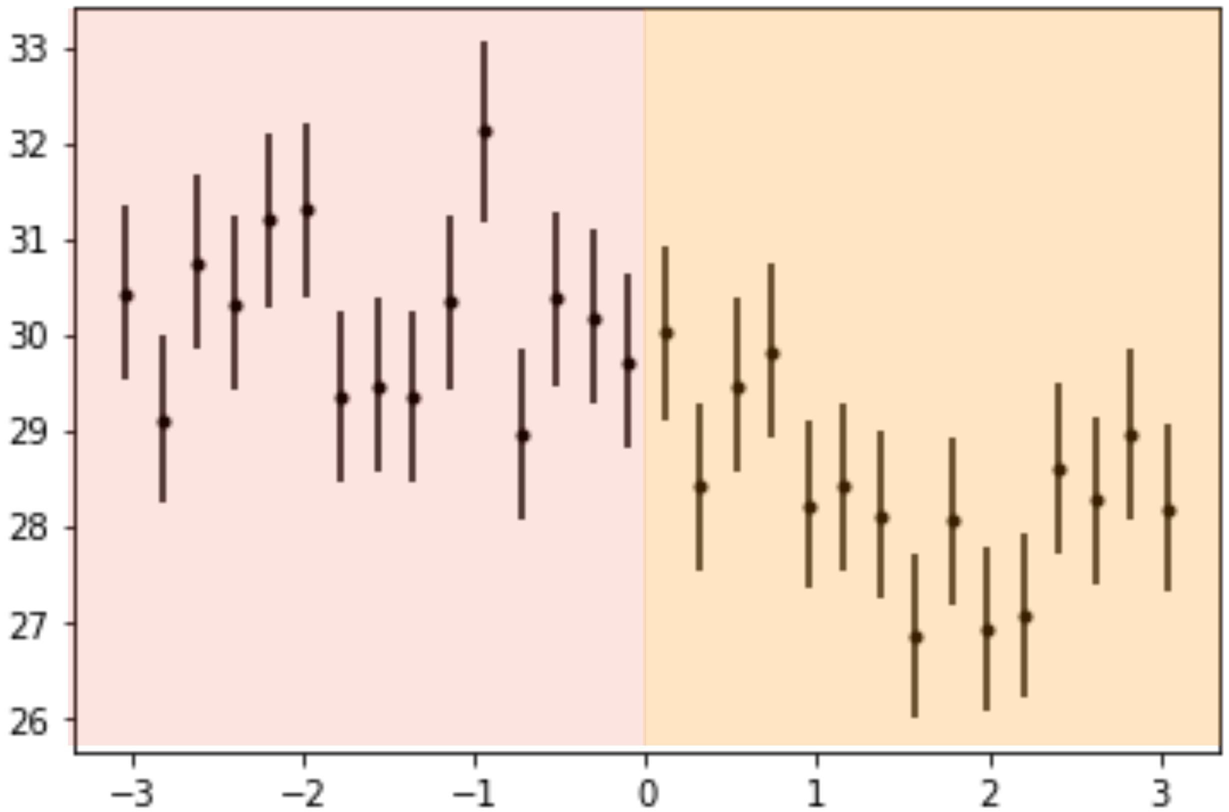
Given an underly poisson distribution for left and right what is the likelihood of each?

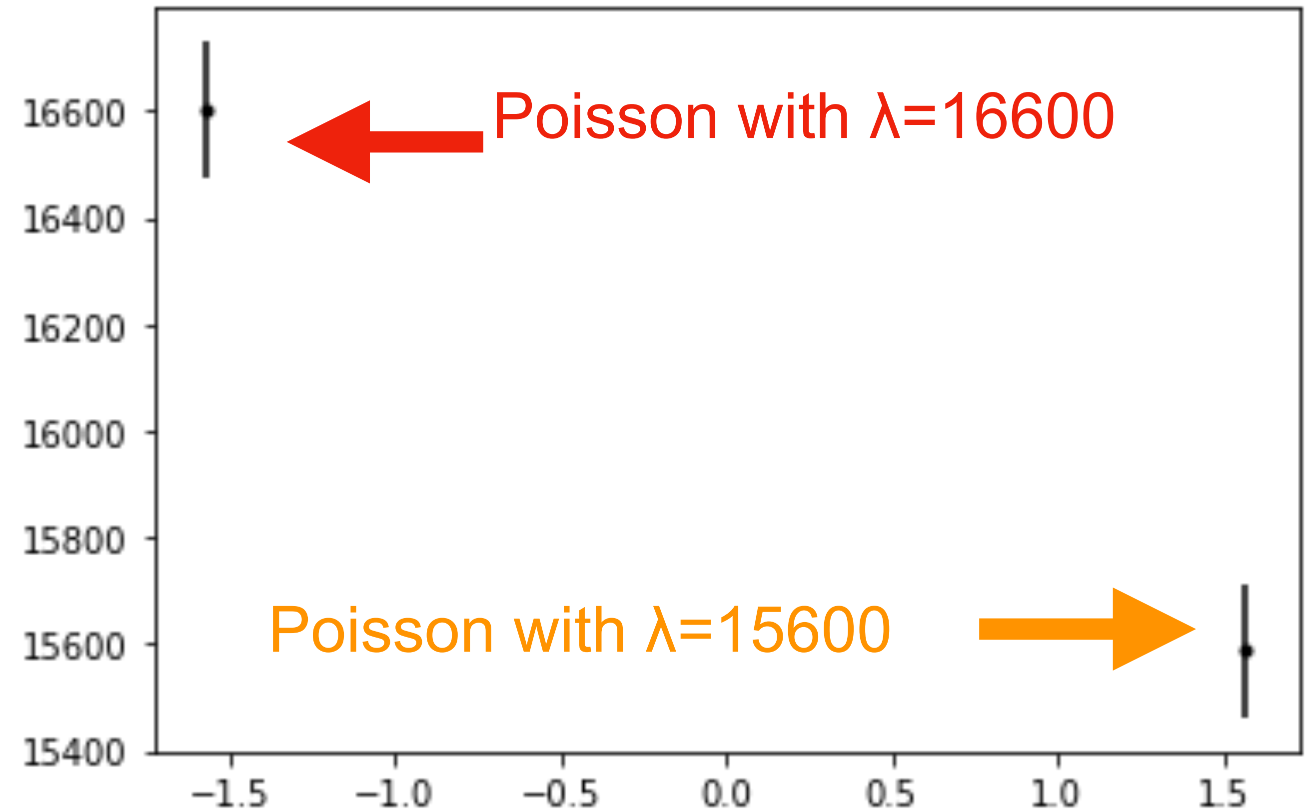
1/1000 times unlikely to fluctuate up and down like this

How do we quantify asymmetry?

Highest Energy > 8 EeV

First lets just make 2 bins





Given an underly poisson distribution for left and right what is the likelihood of each?

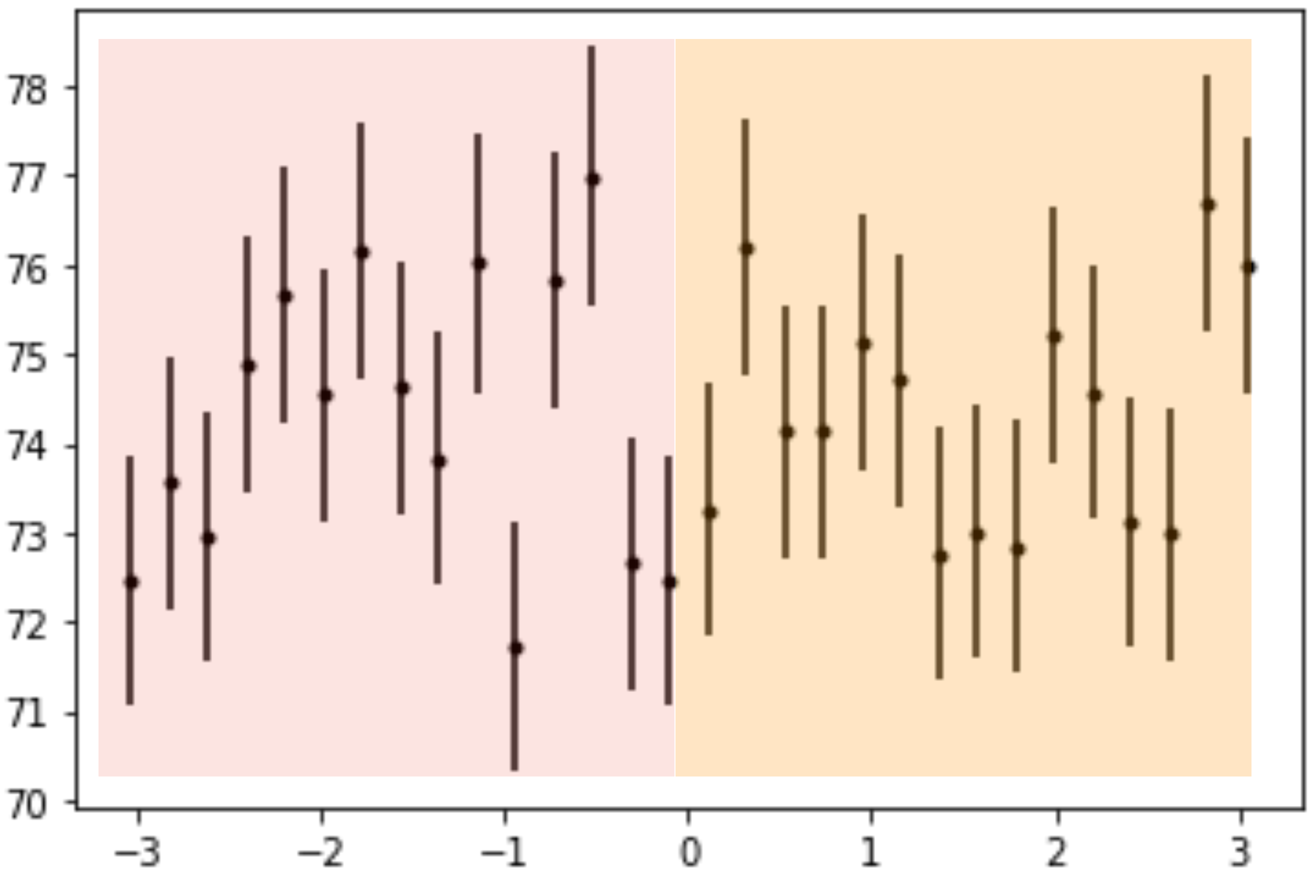
$$p(x = 16600\,|\,\lambda = 16100)/p(x = 16600\,|\,\lambda = 16600) = 10^{-3}$$

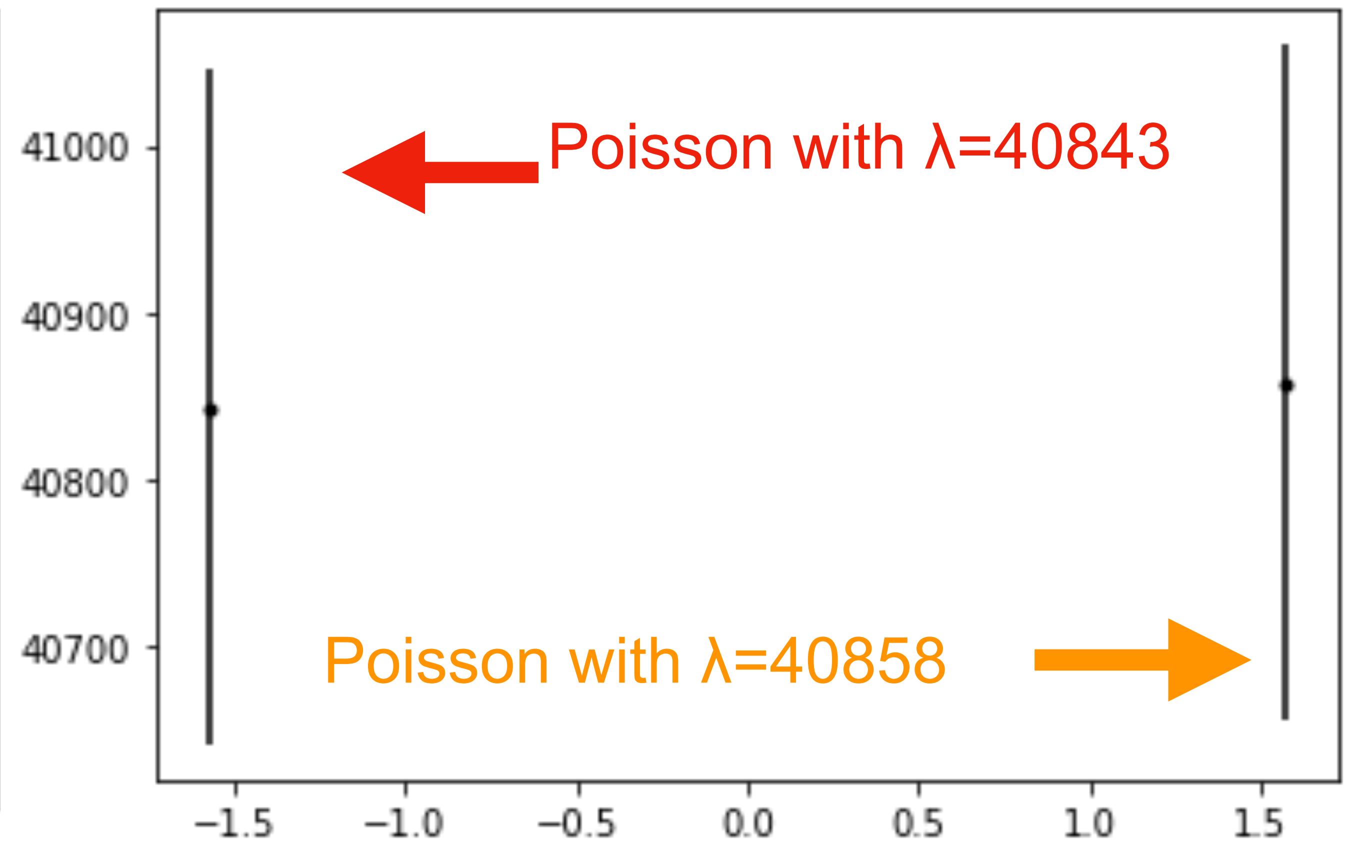
1/1000 times unlikely to fluctuate up and down like this

How do we quantify asymmetry?

Lower Energy 4-8 EeV

First lets just make 2 bins





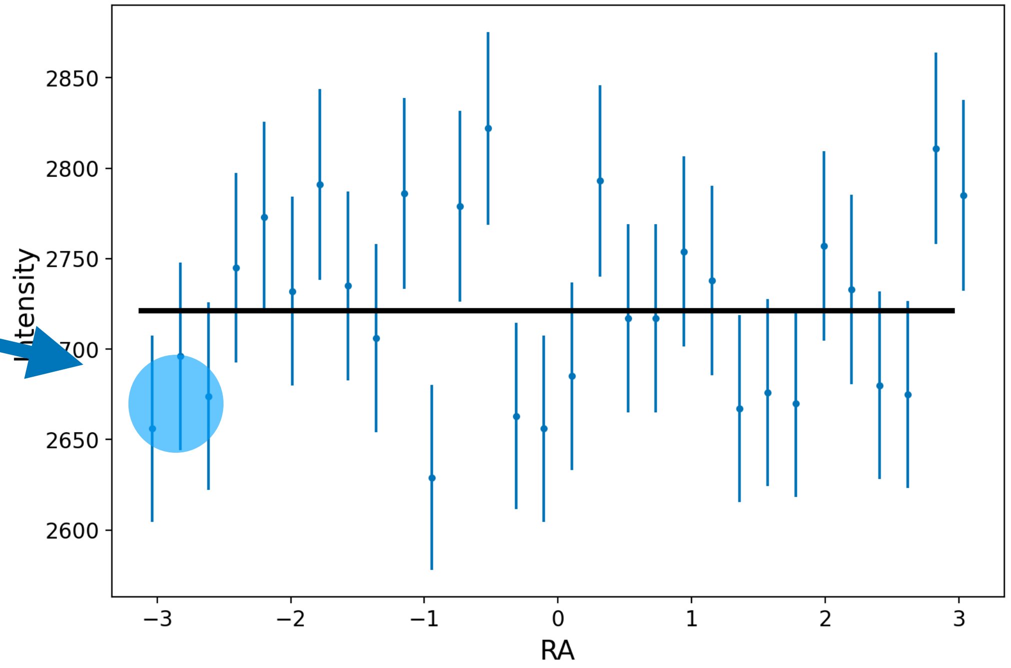
Given an underly poisson distribution for left and right what is the likelihood of each?

$$p = 0.999$$

Why would it change with energy?

What is the likelihood for a gaussian over *N* bins?

$$L(x|\lambda) = \Pi_{i=1}^{N} p(x_i|\lambda)$$



$$L(x|\lambda) = \Pi_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$-\log(L(x|\lambda)) = \sum_{i=1}^{N} -\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\boxed{\begin{gathered} L(x|\lambda) = \Pi_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda}} e^{\frac{-(x_i-\lambda)^2}{2\lambda}} \\ -\log(L(x|\lambda)) = \sum_{i-1}^{N} -\frac{1}{2}\log(2\pi\lambda) + \frac{(x_i-\lambda)^2}{2\lambda} \end{gathered}}$$

Using the Poission variance

What is the likelihood for a gaussian over *N* bins?

$$f(x|\lambda) = -\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$E[f(x)] = -\frac{1}{2}\log(2\pi\sigma^2) + E[g(x)]$$

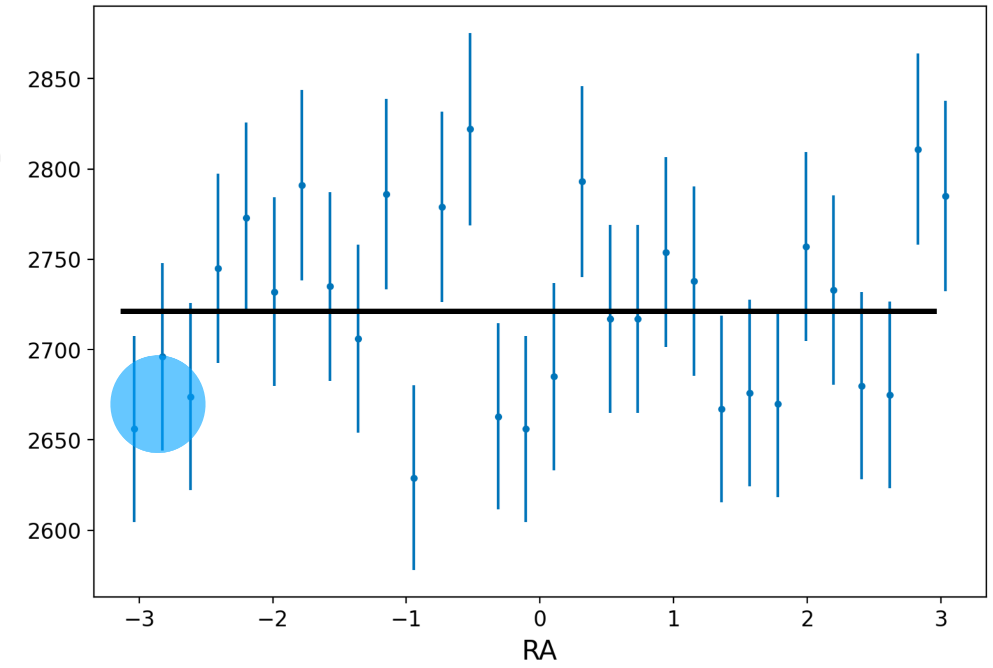
$$g(x) = \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$E[g(x)] = \frac{\mathrm{Var(x)}}{2\sigma^2}$$

$$E[g(x)] = \frac{\sigma^2}{2\sigma^2} = 1/2$$

$$g(x) = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma^2}$$

$$E[g(x)] = N$$



Expectation of log likelihood using a gaussian is N/2

N is the number of bins

χ^2 distribution

- χ^2 distribution is the sum of N independent variables X_i
  - Where the distributions X_i are distributed as normal
  - N denotes the number of degrees of freedom

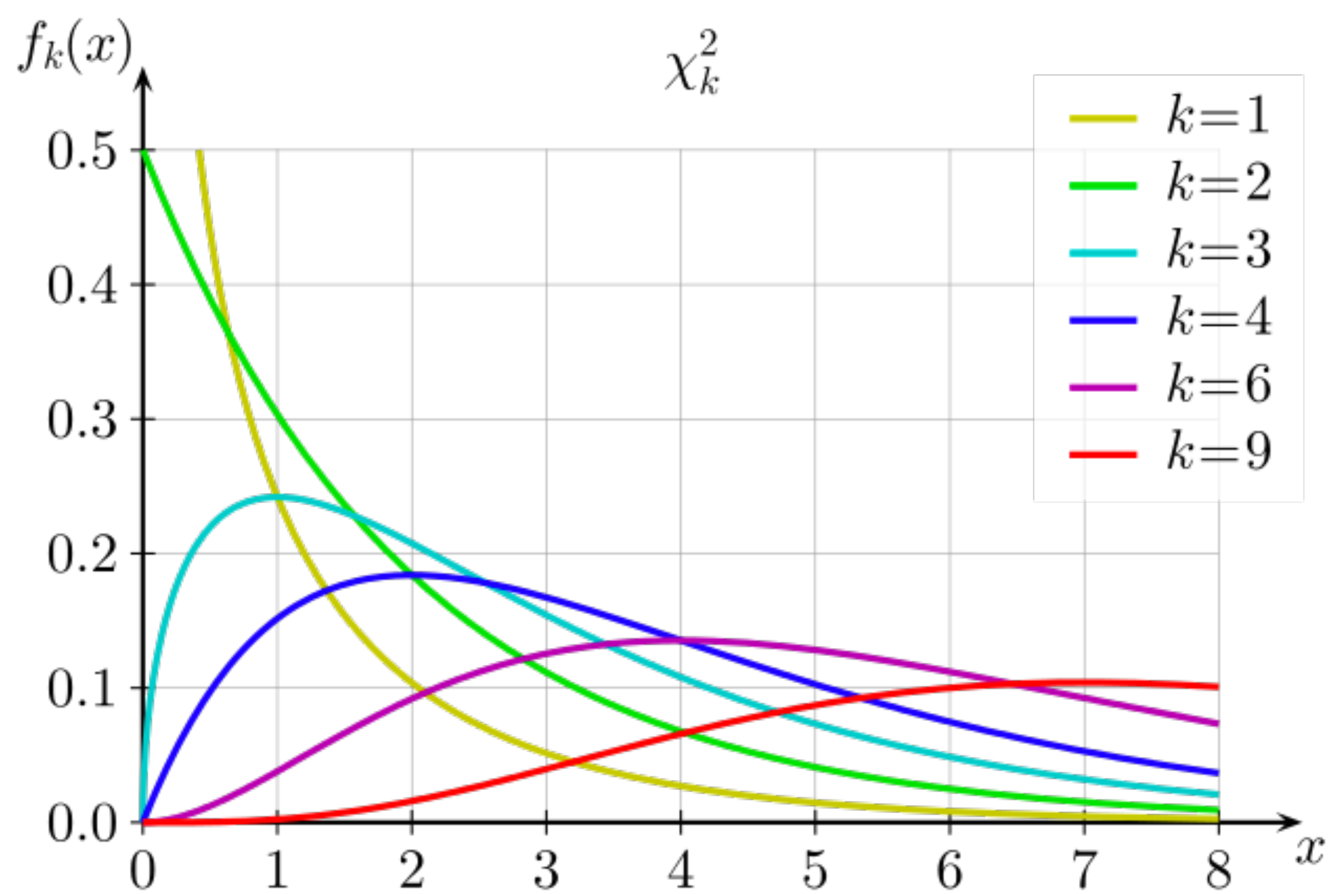
$$\chi^2(x, n) = \sum_{i=1}^{N} \frac{(x-\mu)^2}{\sigma^2}$$

$$\boldsymbol{E}[\chi^2(x)] \approx \boldsymbol{N}$$

$$\boldsymbol{E}[\chi^2(x)/N] \approx 1$$

$$\mathrm{Var}(\chi^2(\mathrm{x})) = 2\mathrm{N}$$

$$\Delta\chi^2(x) = |\chi^2(x) - \chi^2(x \pm \sqrt{2N})|$$



Stop Here!

The next part needs more time

Important Properties

- Taylor expand in our floated parameter (μ in this case)

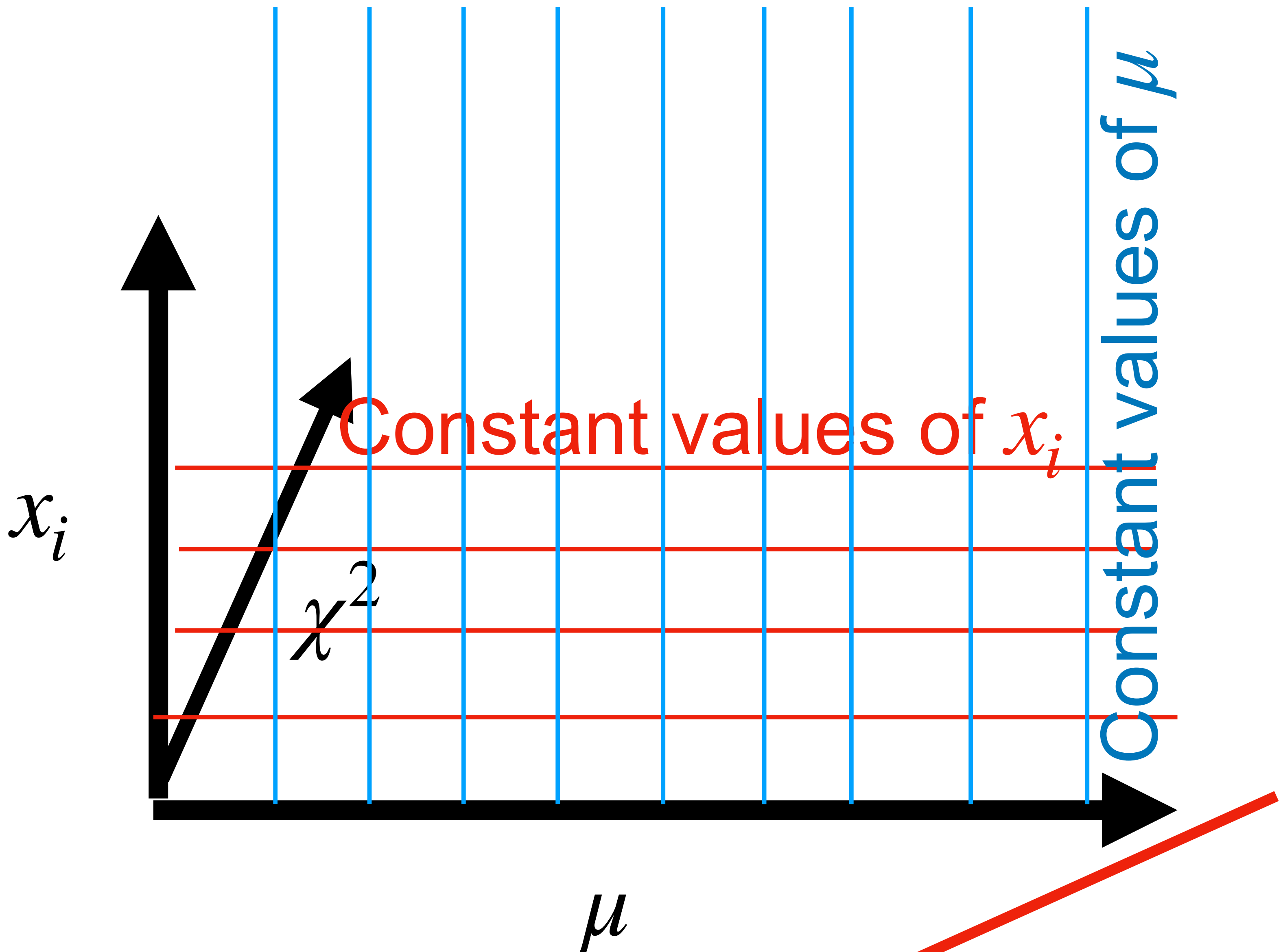
$$\chi^2(x,n) = \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{\sigma_i^2} = \sum_{i=2}^{N} \frac{(x_i-\mu)^2}{\sigma_i^2} + \frac{(x_1-\mu)^2}{\sigma_1^2}$$

Consider this as a fixed constant

Motion of this point by 1 standard deviation in σ_1 causes $\Delta\chi^2 = 1$ from minimum

Can view the distribution of just x_1 as just a χ^2 of 1 DOF

Visualization



$$\chi^2(x,n) = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma_i^2}$$

A function of x_i and μ

Visualization

$$\chi^2(x,n) = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma_i^2} + C_x$$

Chi-2 of N-degrees of freedom

$$\chi^2(x,1) = \frac{(\mu - \mu_0)^2}{\sigma_\mu^2} + C_{mu}$$

Chi-2 of 1-degree of freedom

Visualization

$$\chi^2(x, n) = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma_i^2} + C_x$$

Chi-2 of N-degrees of freedom

$$\chi^2(x, 1) = \frac{(\mu - \mu_0)^2}{\sigma_\mu^2} + C_{mu}$$

Chi-2 of 1-degree of freedom

These are the same formula

$$\chi^2(x_i, \mu) = \chi^2_{min}(x_i, \mu_0) + \frac{1}{2}\frac{\partial^2}{\partial \mu^2}\chi^2_{min}(x_i, \mu_0)(\mu - \mu_0)^2$$

Important Properties

- Taylor expand in our floated parameter (μ in this case)

$$\chi^2(x, n) = \sum_{i=1}^{N} \frac{(x-\mu)^2}{\sigma^2}$$

$$\chi^2(x_i, \mu) = \chi^2_{min}(x_i, \mu_0) + \frac{1}{2}\frac{\partial^2}{\partial \mu^2}\chi^2_{min}(x_i, \mu_0)(\mu - \mu_0)^2$$

Frozen Varying

$$\Delta\chi^2(x, n) = 1 \rightarrow (\mu \rightarrow \mu \pm \sigma)$$

$$1 = \frac{1}{2}\frac{d}{d\mu^2}\chi^2(\mu_0)\sigma^2$$

Important Properties

- Taylor expand in our floated parameter (μ in this case)

$$\chi^2(x,n) = \sum_{i=1}^{N} \frac{(x-\mu)^2}{\sigma^2}$$

$$\chi^2(x_i,\mu) = \underbrace{\chi^2_{min}(x_i,\mu_0)}_{\text{Frozen}} + \underbrace{\frac{1}{2}\frac{\partial^2}{\partial\mu^2}\chi^2_{min}(x_i,\mu_0)(\mu-\mu_0)^2}_{\text{Varying}}$$

$$\Delta\chi^2(x,n) = 1 \rightarrow (\mu \rightarrow \mu \pm \sigma)$$

Wilk's Theorem

$$1 = \frac{1}{2}\frac{d}{d\mu^2}\chi^2(\mu_0)\sigma^2 \longrightarrow \boxed{\sigma^2 = \frac{2}{\frac{d}{d\mu^2}\chi^2(\mu_0)}}$$

An Example

Recall that if we vary take the average over N

Our uncertainty on the mean goes as

$$\sigma_\mu = \sigma\sqrt{\frac{1}{N}}$$

An Example

$$\Delta\chi^2 = 1 = \sum_{i=1}^{N} \frac{(x-\mu_0+\sigma_\mu)^2}{\sigma^2} - \sum_{i=1}^{N} \frac{(x-\mu_0)^2}{\sigma^2}$$

$$1 \approx \sum_{i=1}^{N} \frac{(x-\mu_0+\sigma_\mu)^2}{\sigma^2} - \sum_{i=1}^{N} \frac{(x-\mu_0)^2}{\sigma^2}$$

$$1 = \frac{1}{\sigma^2} \sum_{i=1}^{N} (x-\mu_0+\sigma_\mu)^2 - (x-\mu_0)^2$$

$$1 = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sigma_\mu^2 + 2\sigma_\mu(x-\mu_0)$$

$$1 = \frac{N\sigma_\mu^2}{\sigma^2}$$

$$\sigma_\mu^2 = \frac{\sigma^2}{N}$$

For a poisson distribution we recover the variance per bin

Important Properties

- Taylor expand in our floated parameter (μ in this case)

$$\chi^2(x_i, \mu) = \boxed{\chi^2_{min}(x_i, \mu_0) + \frac{1}{2}\frac{\partial^2}{\partial \mu^2}\chi^2_{min}(x_i, \mu_0)(\mu - \mu_0)^2}$$

χ^2 distribution of 1 degree of freedom

V[$\chi^2(x)$]=1

$$\Delta\chi^2 = 2\Delta \log \mathrm{L} = 1$$

For one degree of freedom

Important Properties

- Taylor expand in our floated parameter (μ in this case)

$$\chi^2(x_i,\mu) = \boxed{\chi^2_{min}(x_i,\mu_0) + \boxed{\frac{1}{2}\frac{\partial^2}{\partial\mu^2}\chi^2_{min}(x_i,\mu_0)}(\mu-\mu_0)^2}$$

$$\frac{1}{2}\frac{d^2\chi^2}{d\mu^2} \rightarrow \frac{1}{\sigma^2}$$

$$\boxed{\frac{\partial^2\chi^2}{\partial\theta^2} = \frac{2}{\sigma_\theta^2}}$$

For any floated parameter uncertainty of that parameter is given by the 2nd derivative of χ^2

This is known as Wilk's Theorem

$$\boxed{\sigma_\theta^2 = \left(\frac{\partial^2 \log L}{\partial\theta^2}\right)^{-1}}$$

Important Properties

- Taylor expand in our floated parameter (μ in this case)

$$\chi^2(x_i, \mu) = \chi^2_{min}(x_i, \mu_0) + \frac{1}{2}\frac{\partial^2}{\partial \mu^2}\chi^2_{min}(x_i, \mu_0)(\mu - \mu_0)^2$$

$$\frac{1}{2}\frac{d^2\chi^2}{d\mu^2} \rightarrow \frac{1}{\sigma^2}$$

χ^2 distribution of 1 degree of freedom

V[$\chi^2(x)$]=1

$$\frac{\partial^2 \chi^2}{\partial \theta^2} = \frac{2}{\sigma_\theta^2}$$

For any floated parameter uncertainty of that parameter is given by the 2nd derivative of χ^2

$$\Delta \chi^2 = 2\Delta \log \mathrm{L} = 1$$

For one degree of freedom

This is known as Wilk's Theorem

$$\sigma_\theta^2 = \left(\frac{\partial^2 \log L}{\partial \theta^2}\right)^{-1}$$

Multiple Dimensions

- For N variables the expansion is the same

$$\chi^2(x_i,\vec{\theta}) = \chi^2_{min}(x_i,\vec{\theta}) + \boxed{\frac{1}{2}(\theta_i-\theta_0)^T \boxed{\frac{\partial^2}{\partial\theta_i\theta_j}\chi^2_{min}(x_i,\vec{\theta}_0)}(\theta_j-\theta_0)}$$

χ^2 distribution of 1 degree of freedom

V[$\chi^2(x)$]=1

Hessian of
the χ^2 distribution

$$\Delta\chi^2 = 2\Delta\log\mathrm{L} = 1$$

For one degree of freedom

This is known as Wilk's Theorem

$$\sigma^2_{ij} = \left(\frac{\partial^2 \log L}{\partial\theta_i\theta_j}\right)^{-1}$$

2D examples

$$\chi^2(x,\vec{\theta}) = \chi^2_{min}(x,\vec{\theta}) + \frac{1}{2}\begin{pmatrix}\theta_a-\theta_{a-min} & \theta_b-\theta_{b-min}\end{pmatrix}\begin{pmatrix}\frac{\partial^2\chi^2}{\partial\theta_a^2} & \frac{\partial^2\chi^2}{\partial\theta_a\partial\theta_b}\\ \frac{\partial^2\chi^2}{\partial\theta_a\partial\theta_b} & \frac{\partial^2\chi^2}{\partial\theta_b^2}\end{pmatrix}\begin{pmatrix}\theta_a-\theta_{a-min}\\ \theta_b-\theta_{b-min}\end{pmatrix}$$

$$\frac{\partial^2\chi^2}{\partial\theta_a\partial\theta_b} \approx 0$$

$$\chi^2(x,\vec{\theta}) = \chi^2_{min}(x,\vec{\theta}) + \frac{1}{2}\begin{pmatrix}\Delta\theta_a & \Delta\theta_b\end{pmatrix}\begin{pmatrix}\frac{\partial^2\chi^2}{\partial\theta_a^2} & 0\\ 0 & \frac{\partial^2\chi^2}{\partial\theta_b^2}\end{pmatrix}\begin{pmatrix}\Delta\theta_a\\ \Delta\theta_b\end{pmatrix}$$

$$\chi^2(x,\vec{\theta}) = \chi^2_{min}(x,\vec{\theta}) + \frac{1}{2}\left(\Delta\theta_a^2\frac{\partial^2\chi^2}{\partial\theta_a^2} + \Delta\theta_b^2\frac{\partial^2\chi^2}{\partial\theta_b^2}\right)$$

$$= \boxed{\chi^2_{min}(x,\vec{\theta}) + \left(\frac{\Delta\theta_a^2}{\sigma^2_{\theta_a}} + \frac{\Delta\theta_b^2}{\sigma^2_{\theta_b}}\right)}$$

Ellipse

Relating all the 2Ds

$$\frac{2}{\sigma^2} = \frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}$$

$$\sigma^2 = 2\left(\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right)^{-1}$$

Wilk's Theorem

$$\begin{pmatrix} \frac{\partial^2 \chi^2}{\partial\theta_a^2} & 0 \\ 0 & \frac{\partial^2 \chi^2}{\partial\theta_b^2} \end{pmatrix} \rightarrow \begin{pmatrix} \frac{2}{\sigma_a^2} & 0 \\ 0 & \frac{2}{\sigma_b^2} \end{pmatrix}$$

Wilk's For Uncorrelated Parameters

For correlated Paramters Can always Diagonalize

$$A^{-1} 2 \begin{pmatrix} \frac{\partial^2 \chi^2}{\partial\theta_a^2} & \frac{\partial^2 \chi^2}{\partial\theta_a \partial\theta_b} \\ \frac{\partial^2 \chi^2}{\partial\theta_a \partial\theta_b} & \frac{\partial^2 \chi^2}{\partial\theta_b^2} \end{pmatrix}^{-1} \quad A = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

2D Terminology

Covariance Matrix

$$\begin{pmatrix} \sigma_a^2 & \mathrm{COV}(a,b) \\ \mathrm{COV}(a,b) & \sigma_b^2 \end{pmatrix} = \sum_{i=1}^{N} \begin{pmatrix} (a_i - \bar{a})^2 & (a_i - \bar{a})\left(b_i - \bar{b}\right) \\ (a_i - \bar{a})\left(b_i - \bar{b}\right) & \left(b_i - \bar{b}\right)^2 \end{pmatrix}$$

$$\rho = \begin{pmatrix} 1 & \dfrac{\mathrm{COV}(a,b)}{\sigma_a \sigma_b} \\ \boxed{\dfrac{\mathrm{COV}(a,b)}{\sigma_a \sigma_b}} & 1 \end{pmatrix}$$

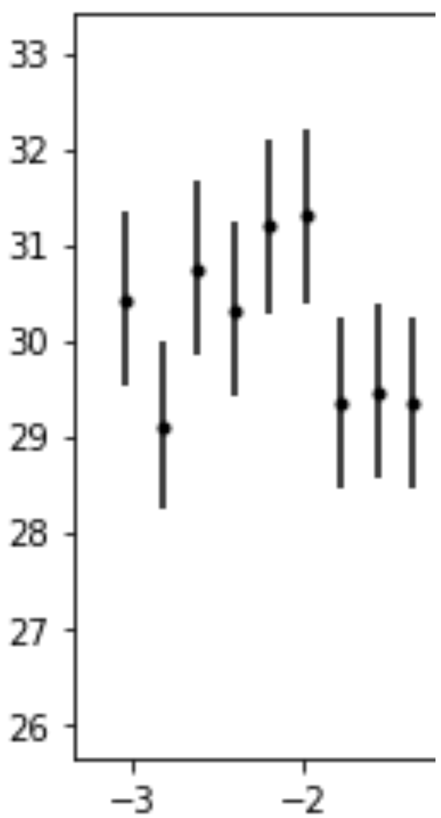
Correlation Matrix

Correlation Coefficient

How do we quantify asymmetry?

First lets just make 2 bins

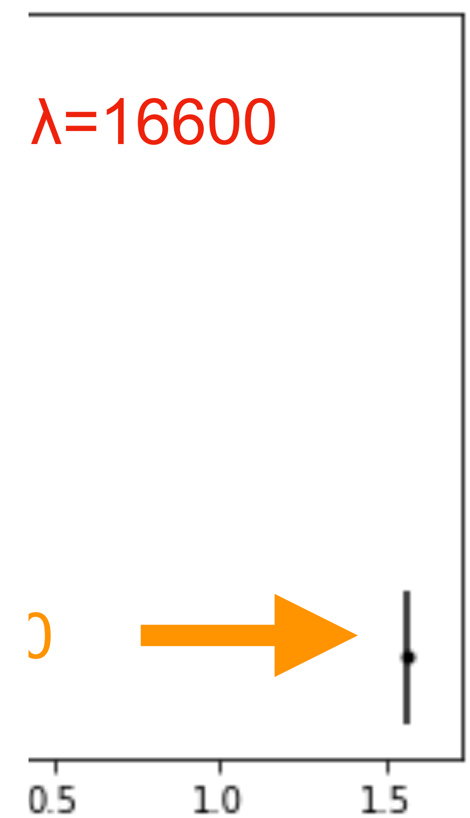
$$L(x|\lambda) = \Pi_{i=1}^{N} p(x_i|\lambda)$$



$$L(x|\lambda) = \Pi_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$-\log(L(x|\lambda)) = \sum_{i-1}^{N} -\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$L(x|\lambda) = \Pi_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda}} e^{\frac{-(x_i-\lambda)^2}{2\lambda}}$$



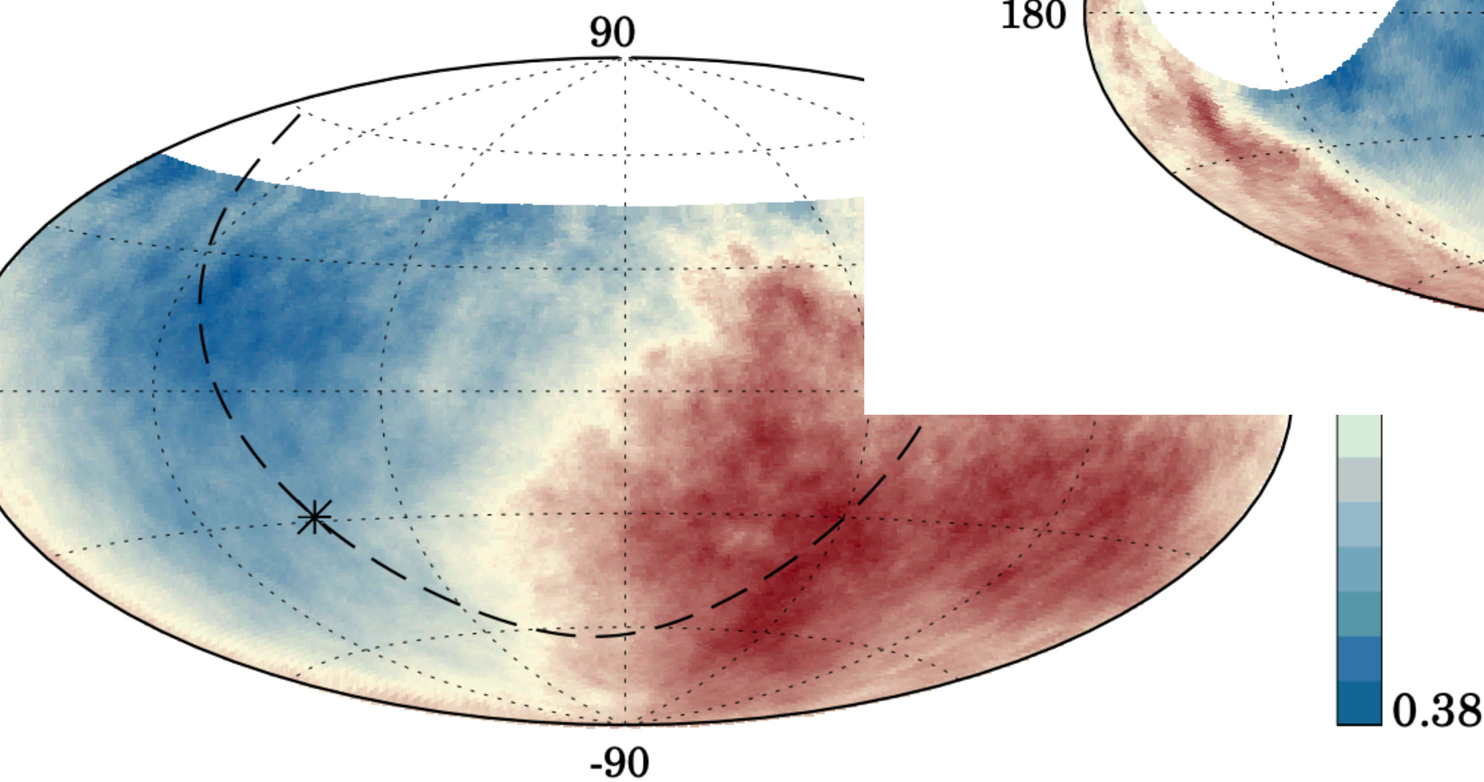
$$-\log(L(x|\lambda)) = \sum_{i-1}^{N} -\frac{1}{2}\log(2\pi\lambda) + \frac{(x_i-\lambda)^2}{2\lambda}$$

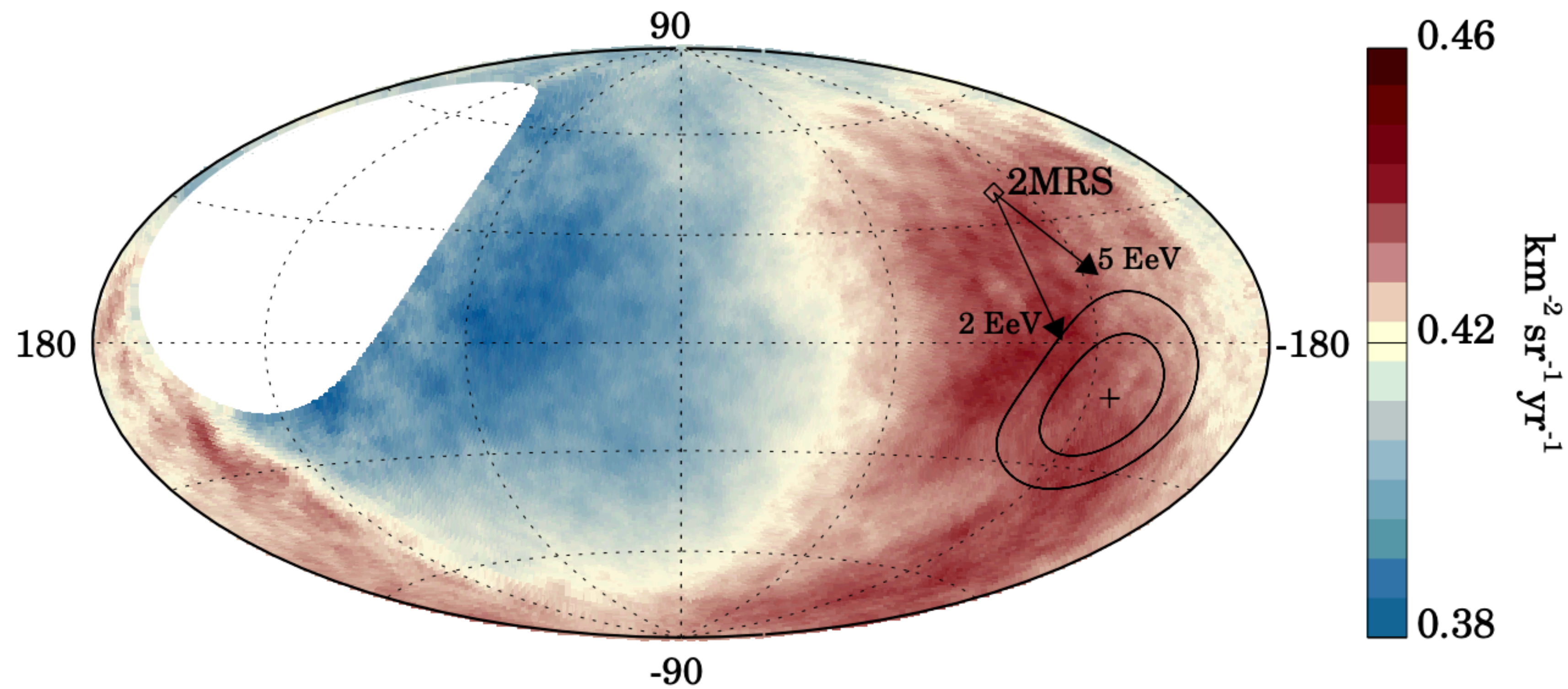
Given an unde … od of each?

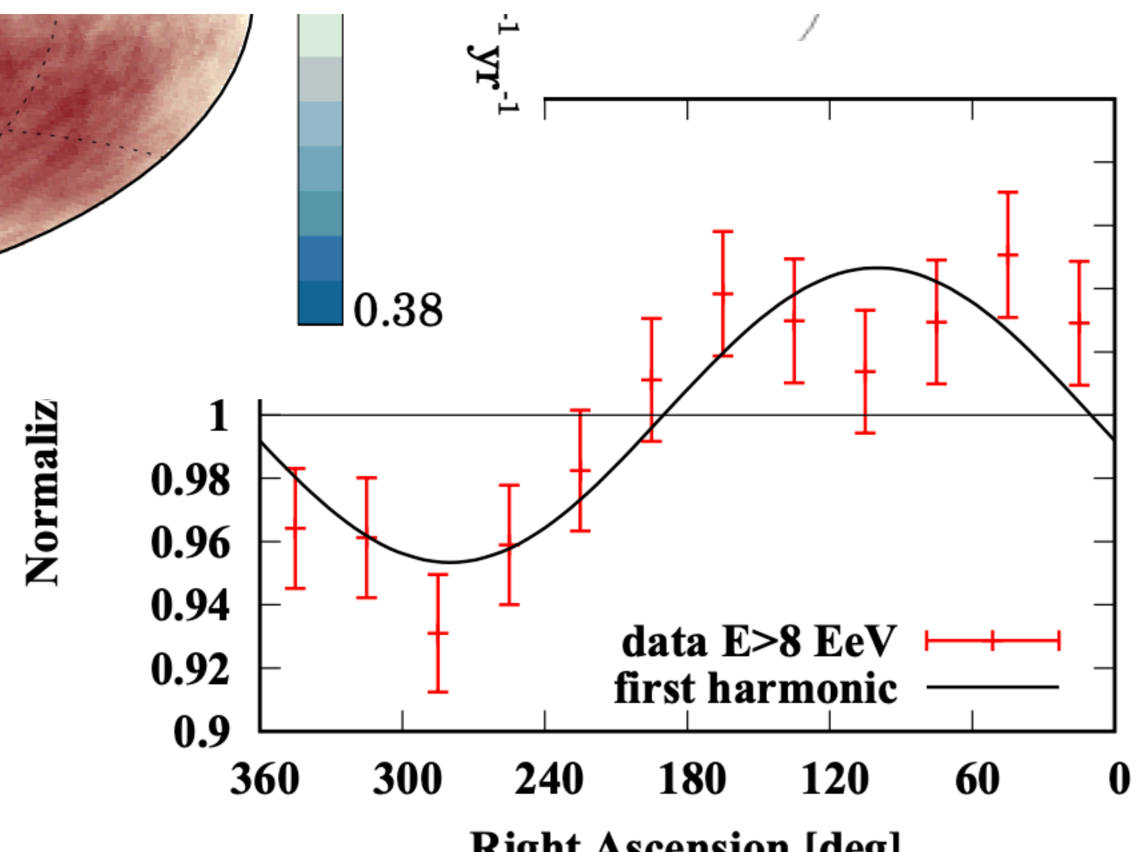
$p(x =$ … -3

1/1000 times unlikely to fluctuate up and down like this

We also find







https://arxiv.org/pdf/1012.3754.pdf