Analysis of Fermi LAT data

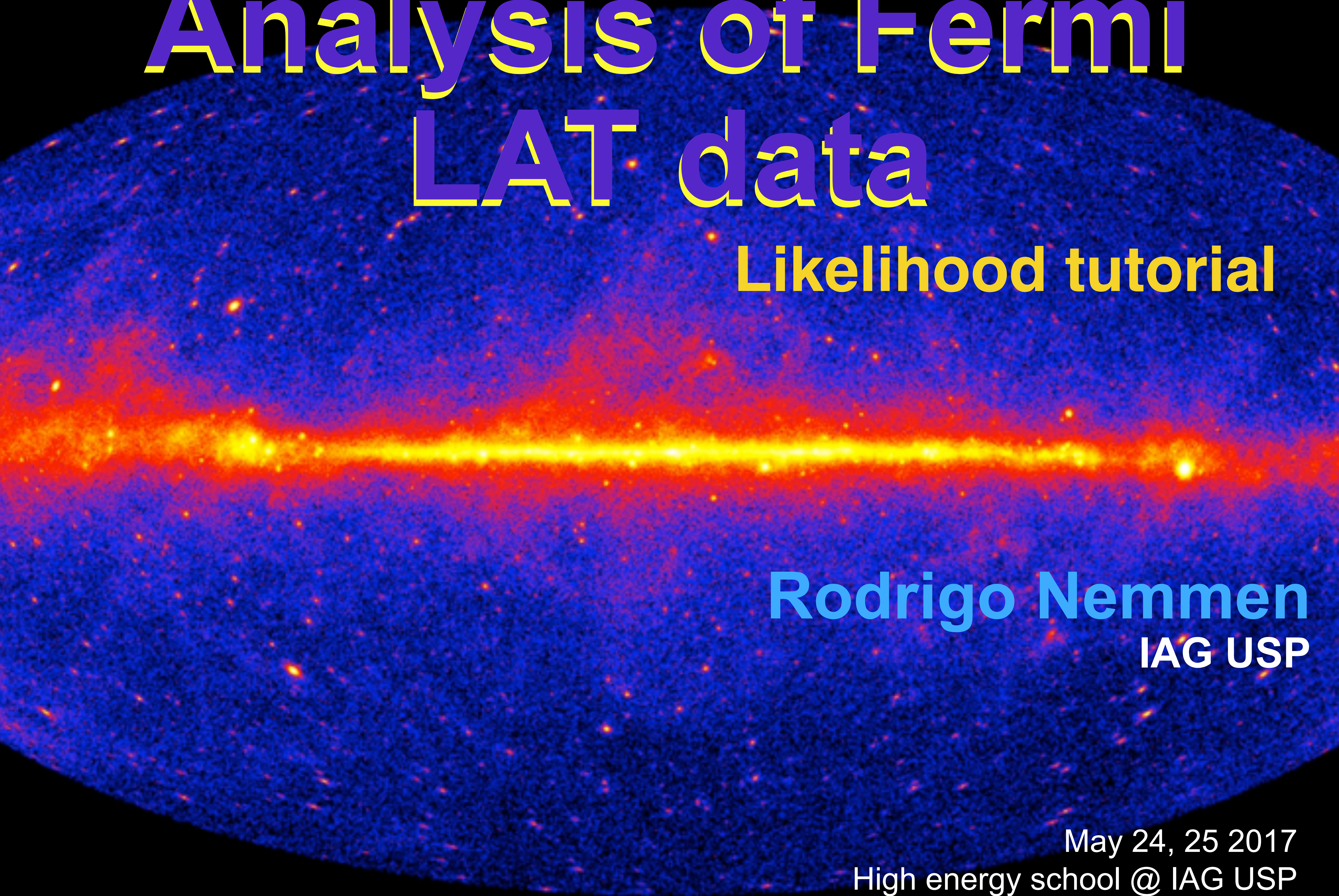
Likelihood tutorial

Rodrigo Nemmen

IAG USP

May 24, 25 2017

High energy school @ IAG USP



Overview of activities

Day 1

- Introduction, overview
- Obtaining and exploring LAT data
- Inspecting the data: count map

Day 2

- **Basics of modeling**
- Likelihood analysis of a blazar
- Create a SED
- Produce a light curve

Structure of this talk

- **Basic theory of likelihood analysis**

see jupiter notebook

- Likelihood fit → Characterize spectra of a source
- obtain spectral energy distribution (SED)
- build a light curve

hands-on

Fundamental questions of ~~science~~ statistics

Model fitting: given this model, what parameters best fit my data?

Model selection: given two potential models, which better describes my data?

Model fitting

given this model, what
parameters best fit my data?

Maximum likelihood approach

Frequentist

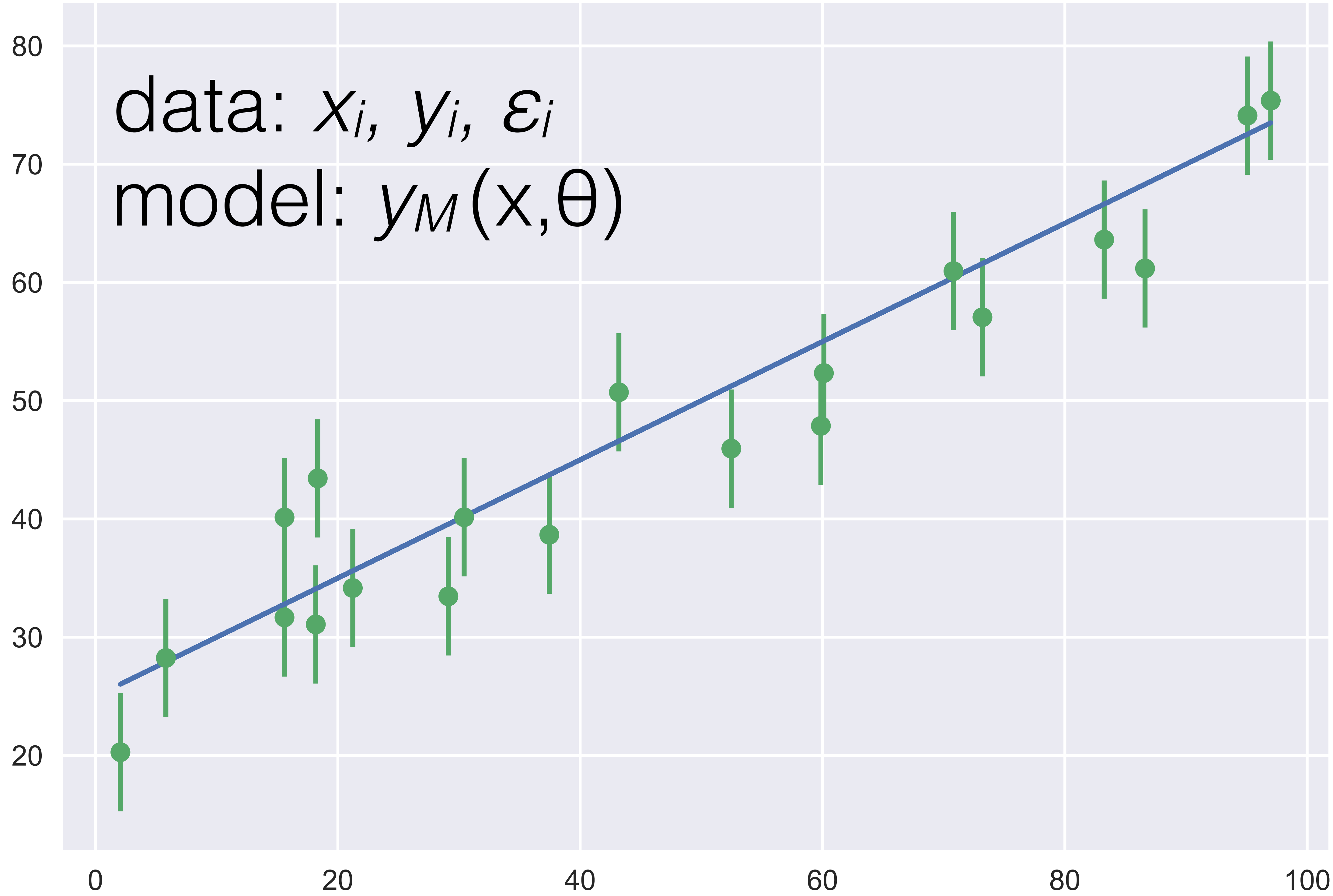
$$P(data \mid scientific\ model)$$

$$P(D \mid \theta)$$

likelihood

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0$$

How to write the likelihood function for the problem below?



Probability for a single measurement

$$P_i = \frac{1}{\sqrt{2\pi\epsilon_i^2}} e^{-\frac{(y_i - y_M)^2}{2\epsilon_i^2}}$$

Likelihood:

$$\mathcal{L} = \sum_i P_i = \sum_i \frac{1}{\sqrt{2\pi\epsilon_i^2}} e^{-\frac{(y_i - y_M)^2}{2\epsilon_i^2}}$$

Log-likelihood:

$$\log \mathcal{L} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \log(2\pi\varepsilon_i^2) + \frac{[y_i - y_M(x_i;\theta)]^2}{\varepsilon_i^2} \right\}$$

What about Fermi LAT data?

Likelihood for Poisson distribution

$$\mathcal{L} = \prod_i P_i = \prod_i \frac{m_i^{n_i} e^{-m_i}}{n_i!}$$

Binned likelihood

$$\mathcal{L} = \prod_i e^{-m_i} \prod_i \frac{m_i^{n_i}}{n_i!} = e^{-N_{\text{pred}}} \prod_i \frac{m_i^{n_i}}{n_i!}$$

For bin sizes infinitesimally small, n_i=1

$$\mathcal{L} = e^{-N_{\text{pred}}} \prod_i m_i$$

Unbinned likelihood

Log-likelihood

$$\log \mathcal{L} = \sum_i \log m_i - N_{\text{pred}}$$

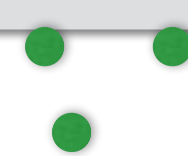
We need to maximize this likelihood function numerically

$$\log \mathcal{L} = \sum_i \log m_i - N_{\mathrm{pred}}$$

total predicted number of counts

$$\int_{\mathrm{ROI}} \left(\begin{array}{c}\text{instrument}\\ \text{response}\\ \text{function}\end{array}\right)\left(\begin{array}{c}\text{source}\\ \text{model}\end{array}\right) dE\, d\mathbf{p}\, dt$$

source model = point sources + diffuse sources + other stuff



Model selection

given two potential models, which better describes my data?

Is there a source in a given position of the sky?

maximum likelihood: model without an additional source (the "null hypothesis")

$$TS = -2\log\left(\frac{\mathcal{L}_{\max,0}}{\mathcal{L}_{\max,1}}\right)$$

maximum likelihood: model with additional source

Rule of thumb:

TS = 100 ⇒ ≈10σ

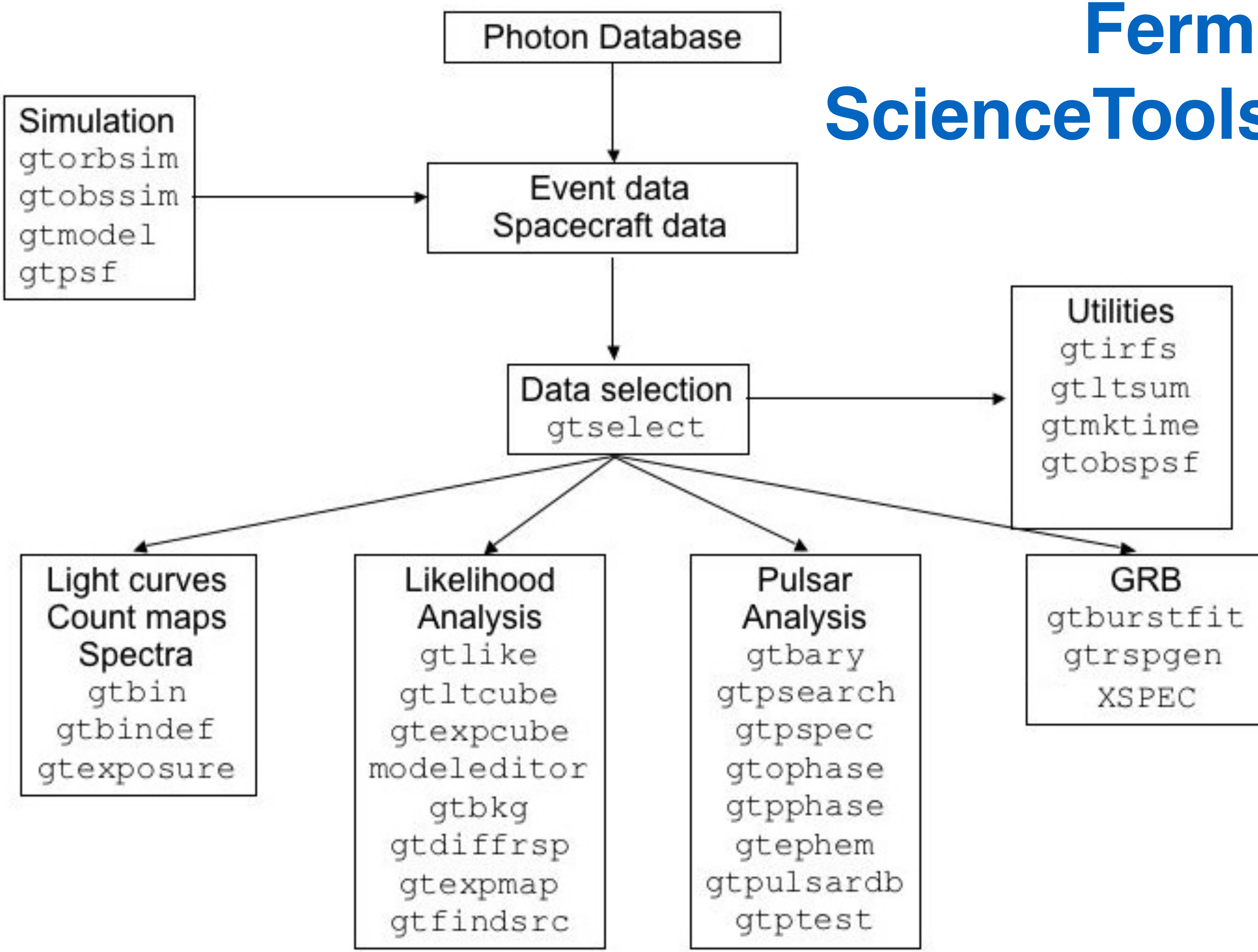
TS = 25 ⇒ ≈5σ

Hands-on activity

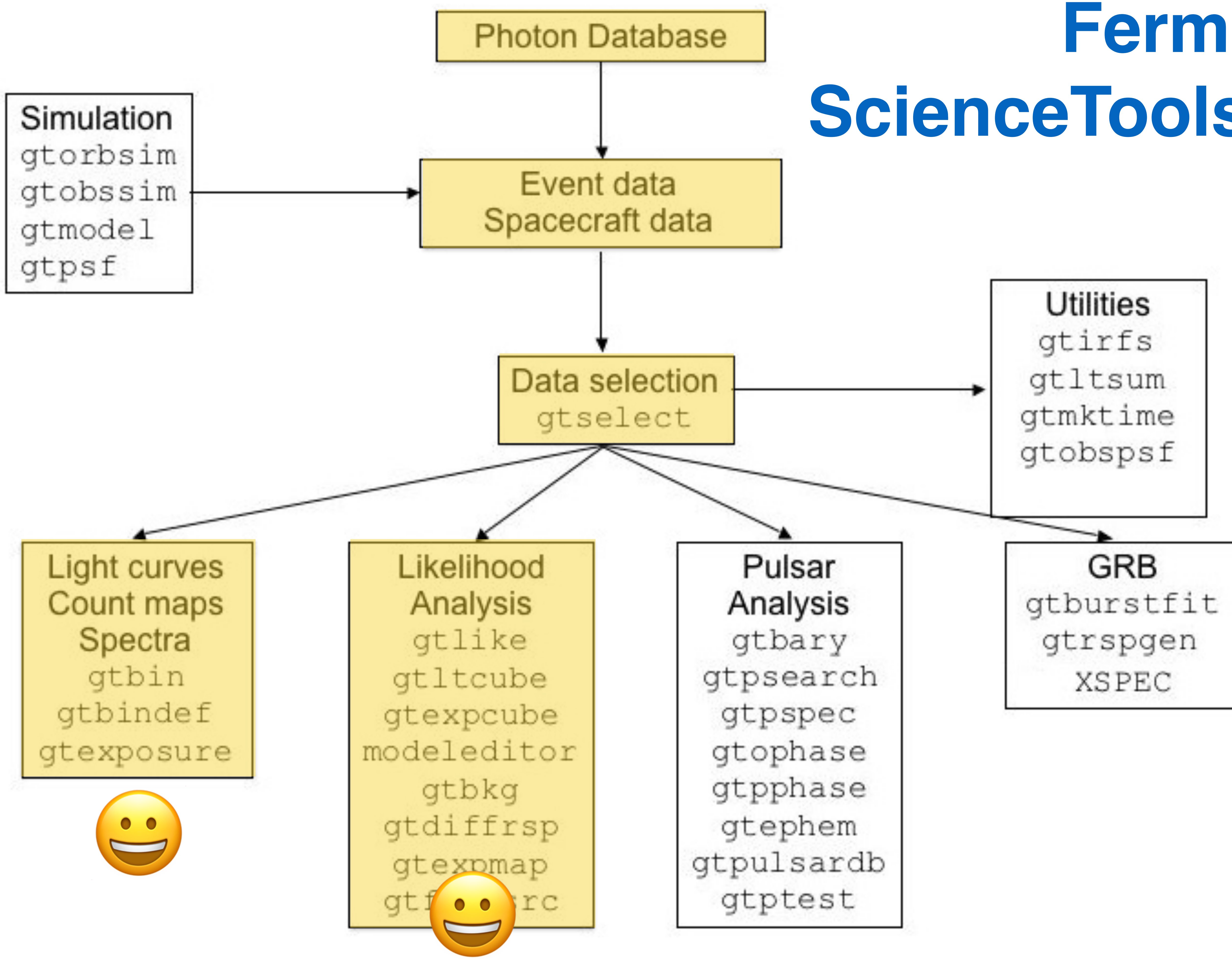
Steps for modeling of *Fermi* LAT data

1 Define your source model
2 Select data + ROI cuts gtselect
3 Select good time intervals (GTI) gtmktime
4 Bin data, create a count map gtbin
5 Compute useful quantities: live time cube, binned exposure cube gtltcube, gtexpcube2
6 Maximize likelihood numerically, get initial estimate of parameters gtlike
7 Optimize fit: improve parameter estimate

Fermi ScienceTools



Fermi ScienceTools



Python wrappers make analysis of Fermi data much easier: `Enrico` and `fermipy`

We will use Enrico:

http://enrico.readthedocs.io/en/latest/

Tutorial website

https://github.com/rsnemmen/Fermi-LAT-tutorial

Likelihood theory lesson: Tutorials → day two → jupiter notebook

Hands-on practice: Tutorials → day two → getting a flux

Please do not download large files during the tutorial or the WIFI network will *overload*

We will distribute the software and data you need via USB sticks

E-mail	rodrigo.nemmen@iag.usp.br
Web	http://rodrigonemmen.com
Twitter	@nemmen
Github	rsnemmen
Facebook	http://facebook.com/rodrigonemmen
Bitbucket	nemmen
Blog	http://astropython.blogspot.com
figshare	http://bit.ly/2fax2cT