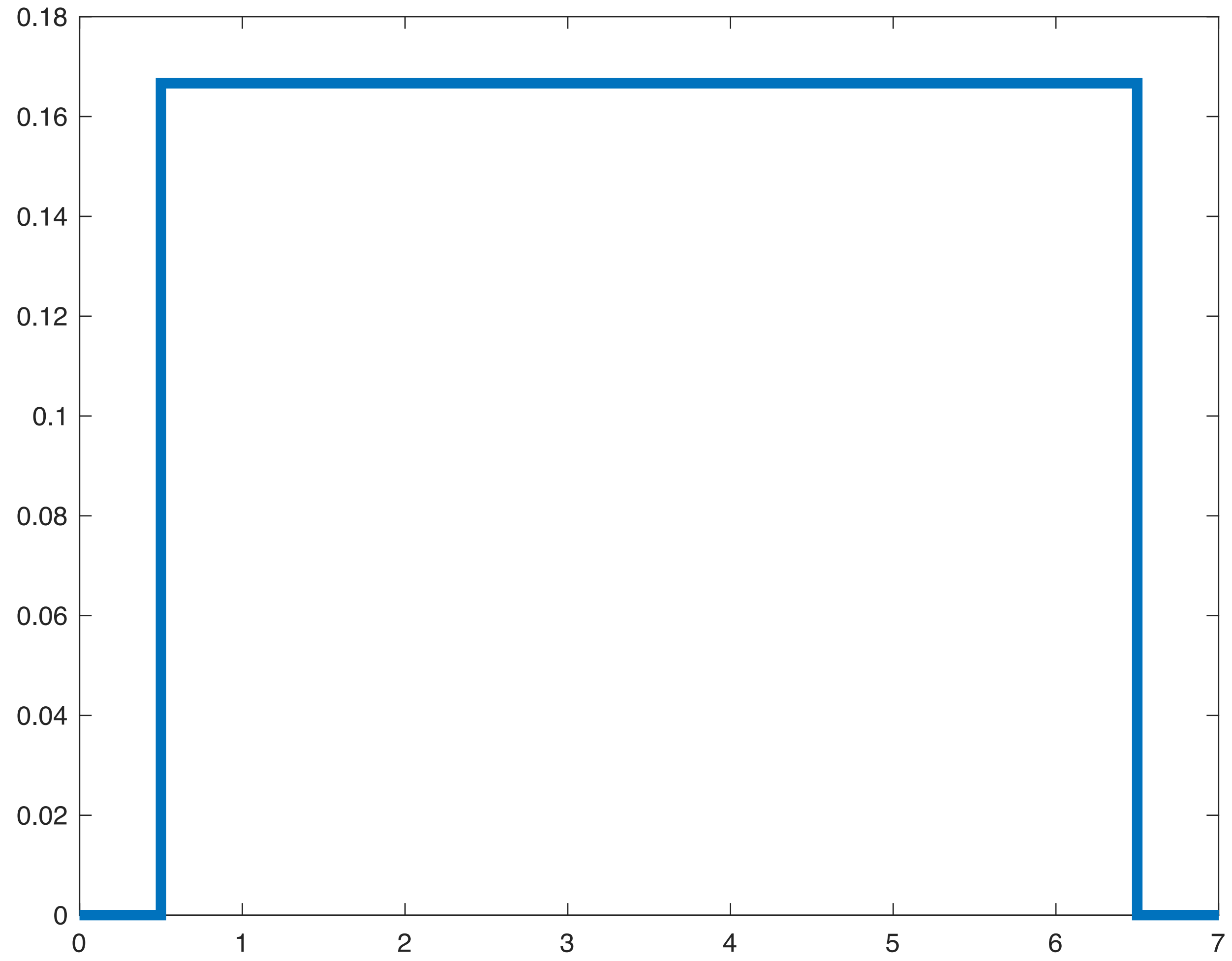
Class 13: stats mini review; the blob, parameters, and analysis dragons

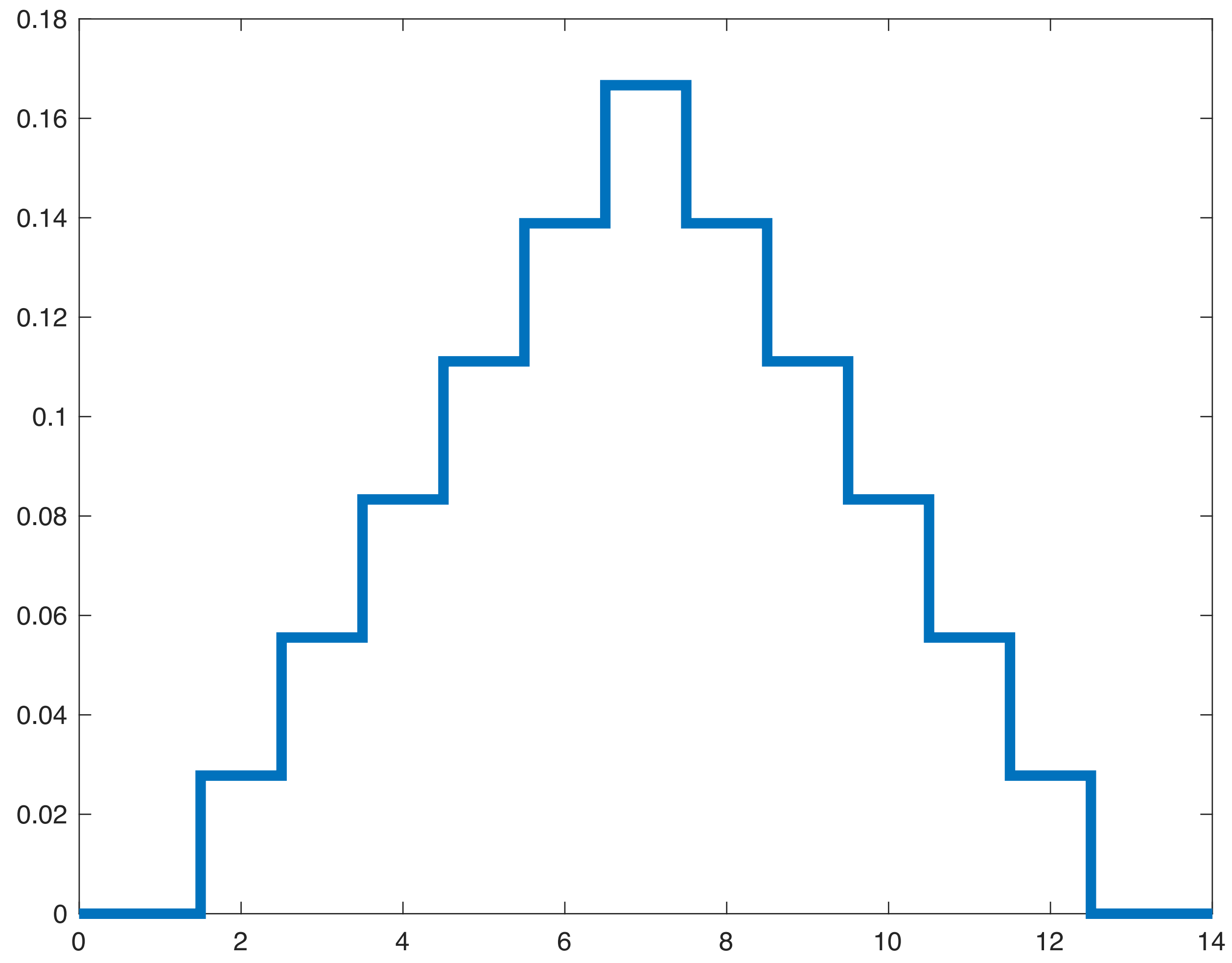
Miguel F. Morales
Bryna Hazelton

Small stats review

Convolution, six sided die

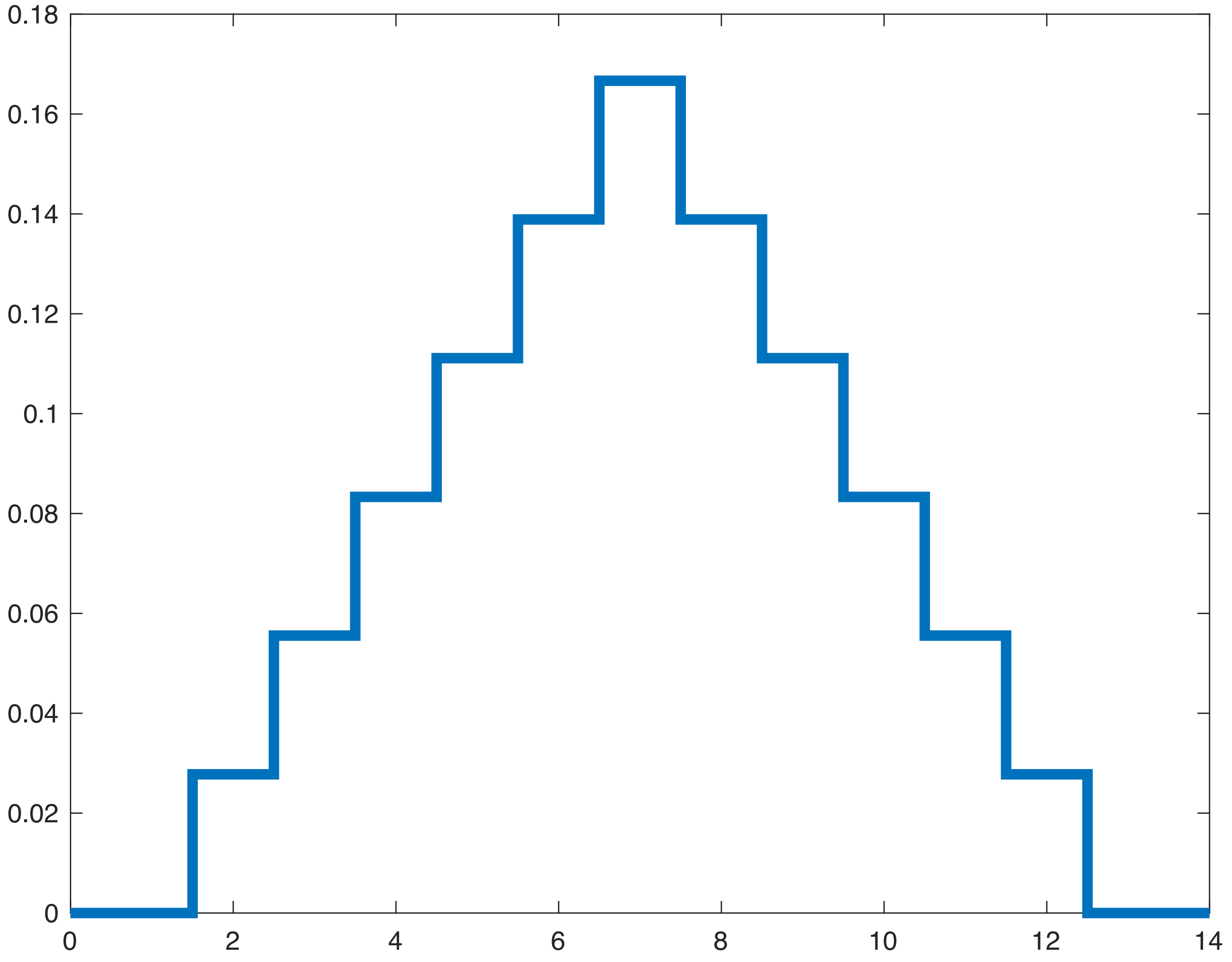


Convolution, six sided die

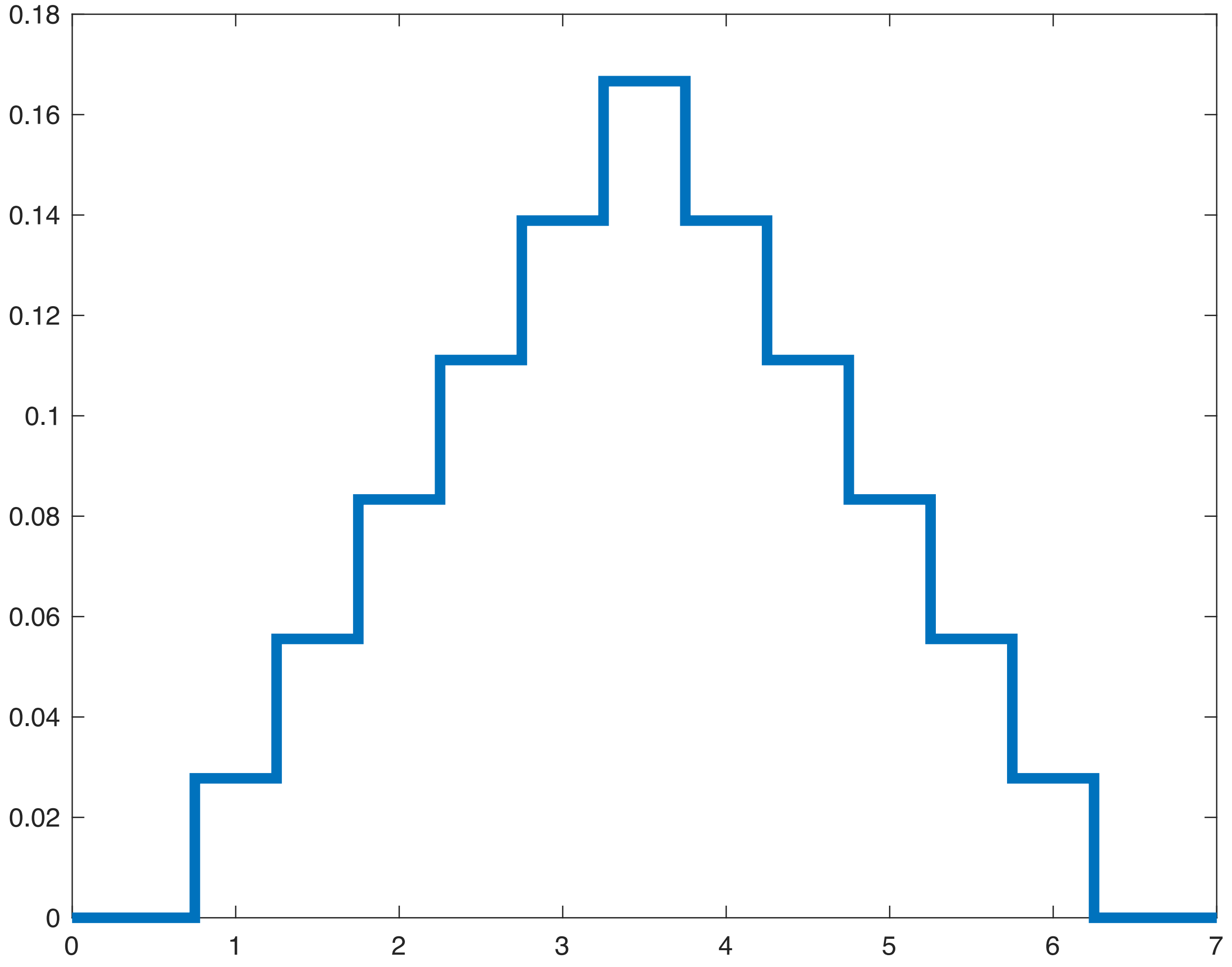


Convolution, six sided die

Sum

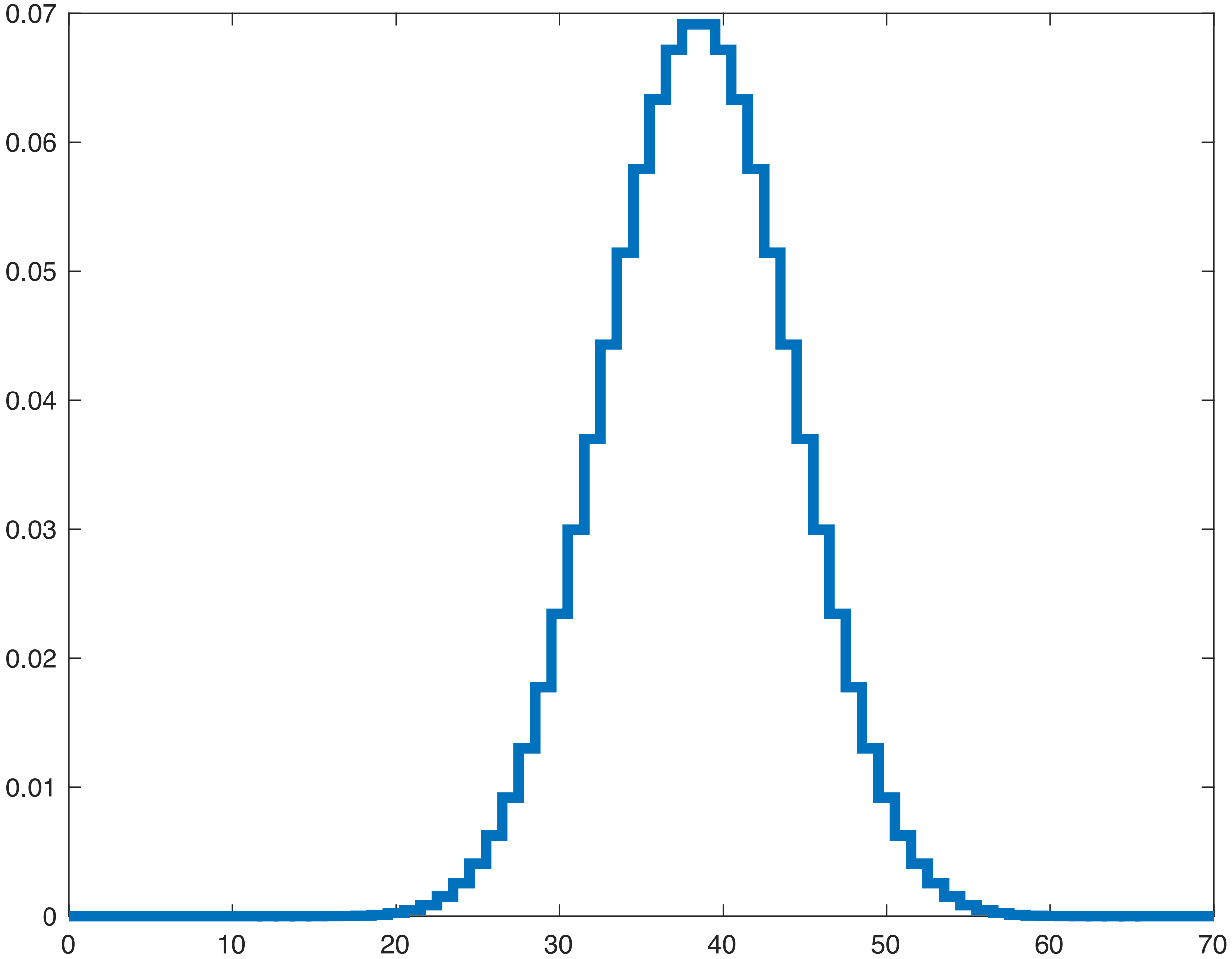


Average

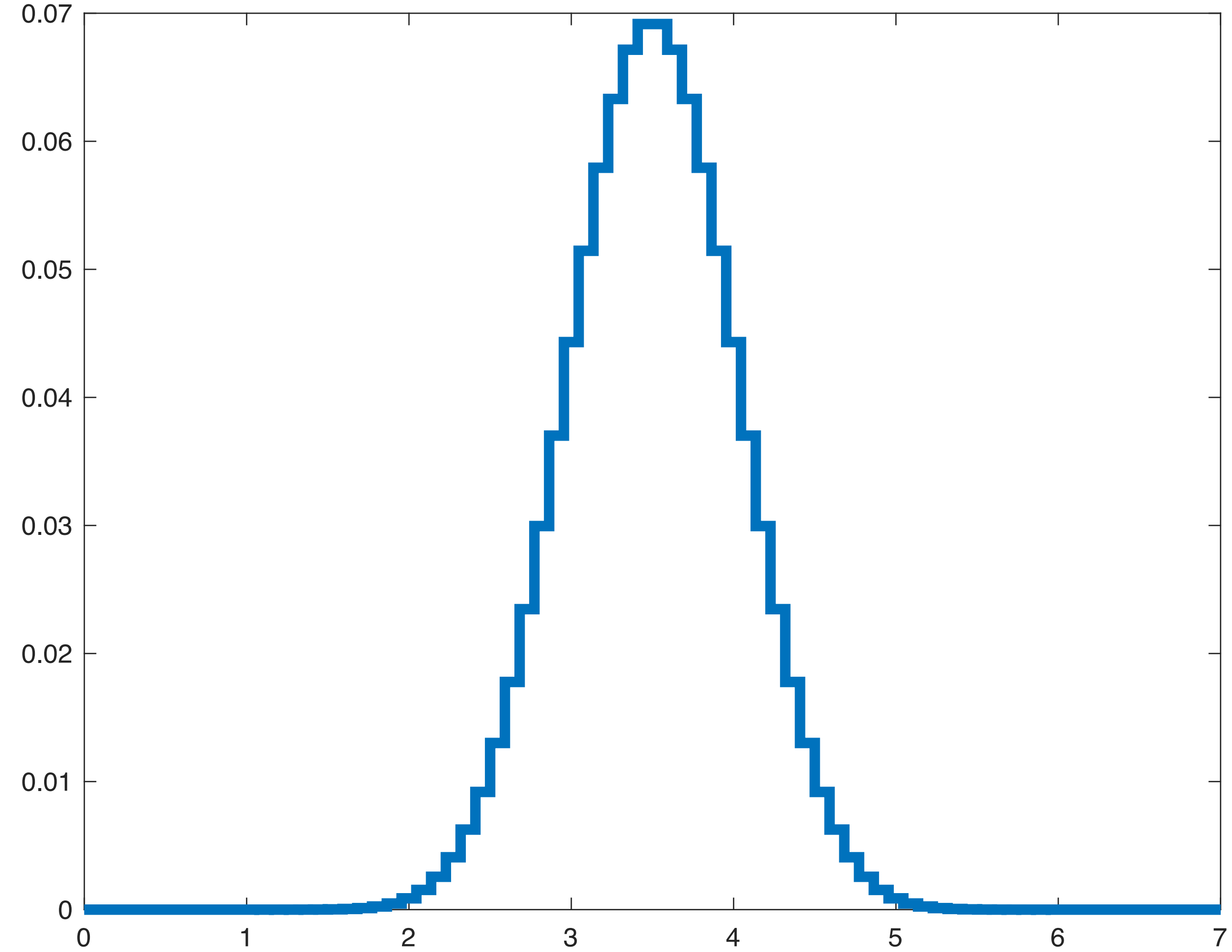


Convolution, six sided die

Sum of 11

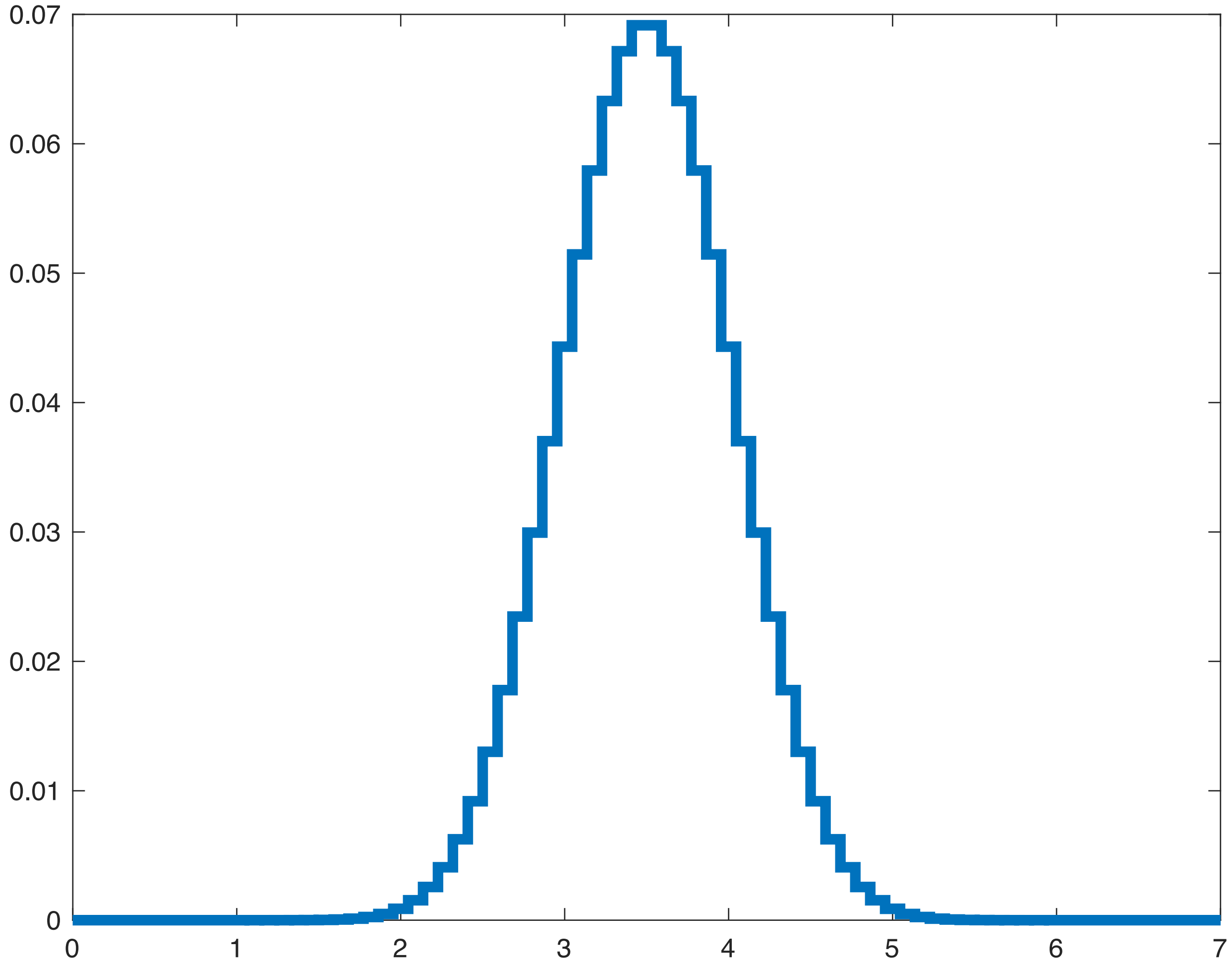


Average of 11

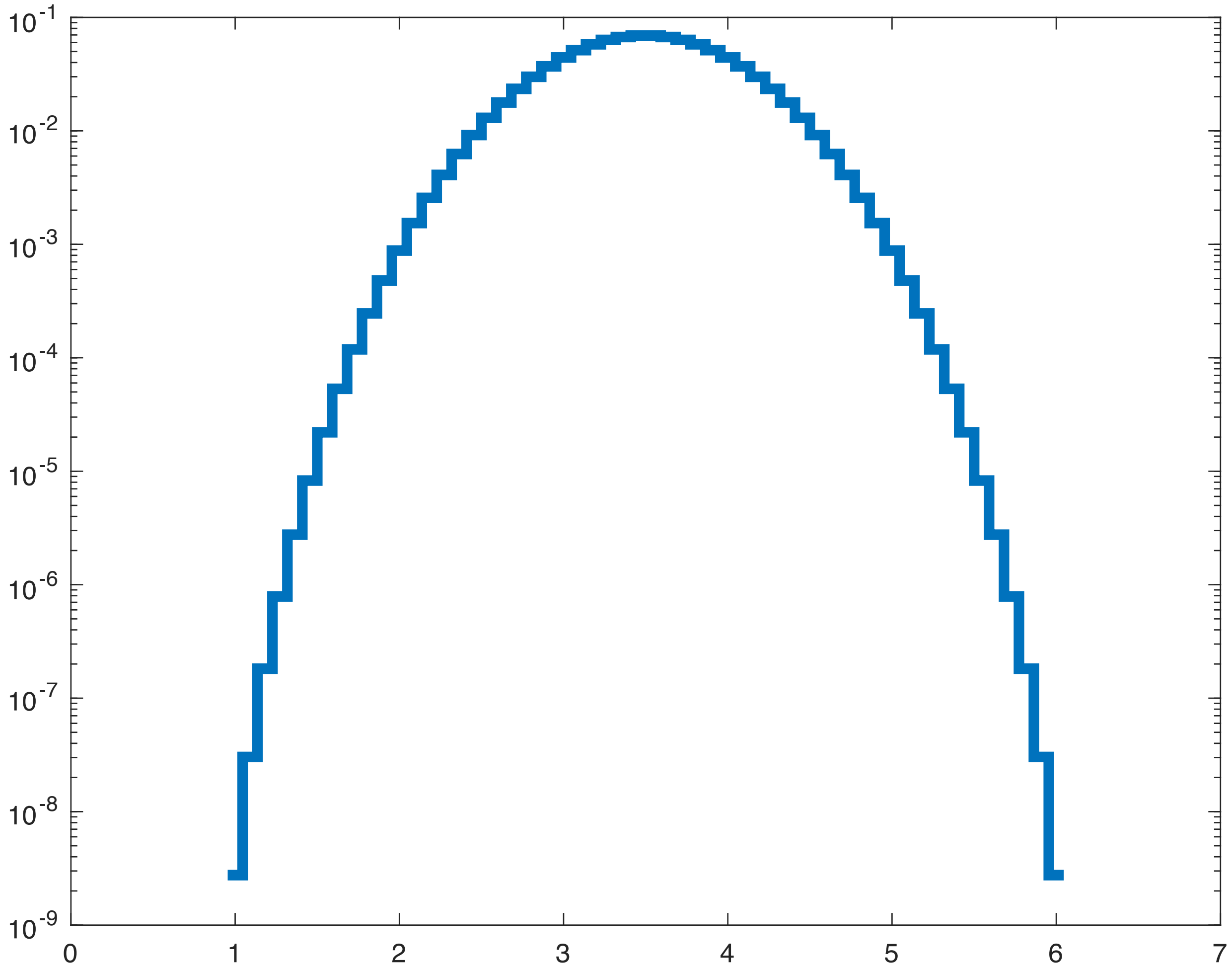


Convolution, six sided die

Average of 11

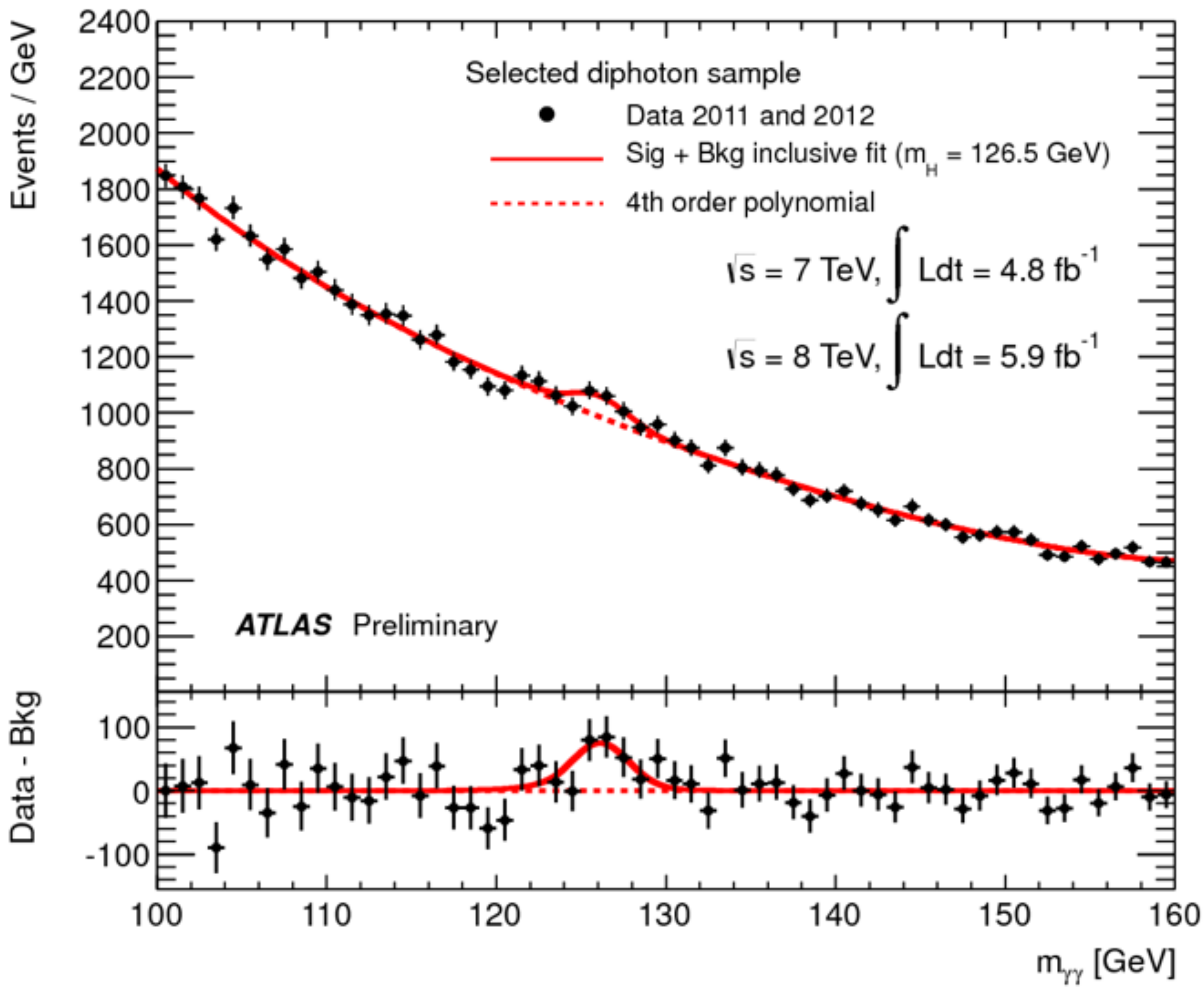


Average of 11

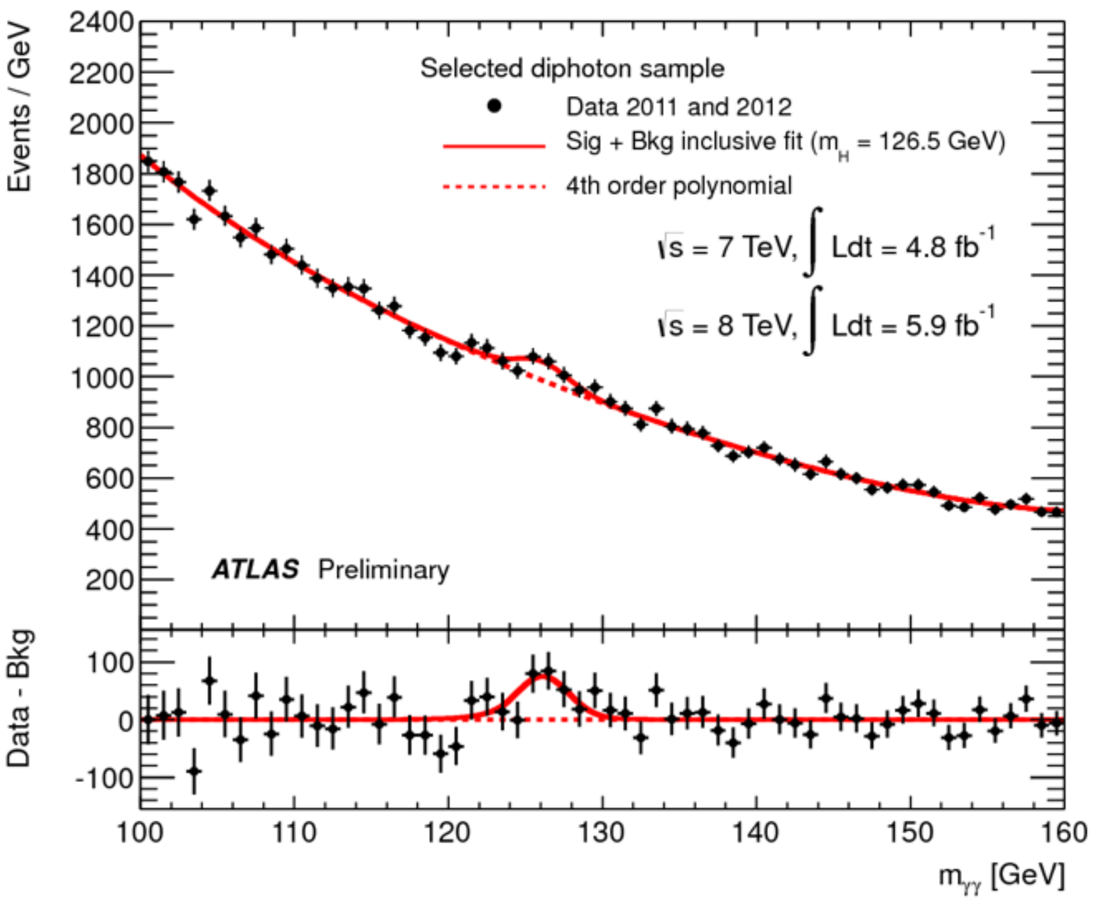


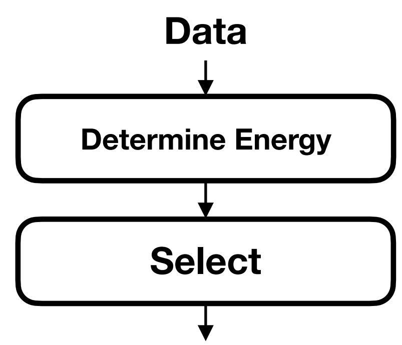
Background?

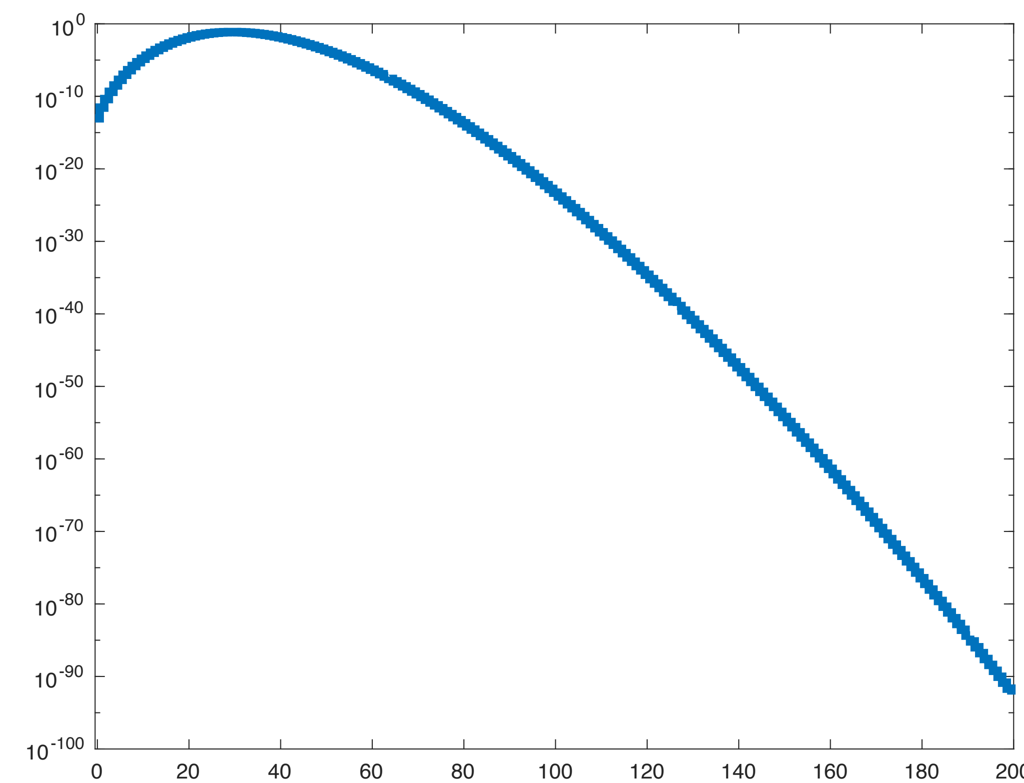
LHC Higgs discovery



LHC Higgs discovery, binned analysis



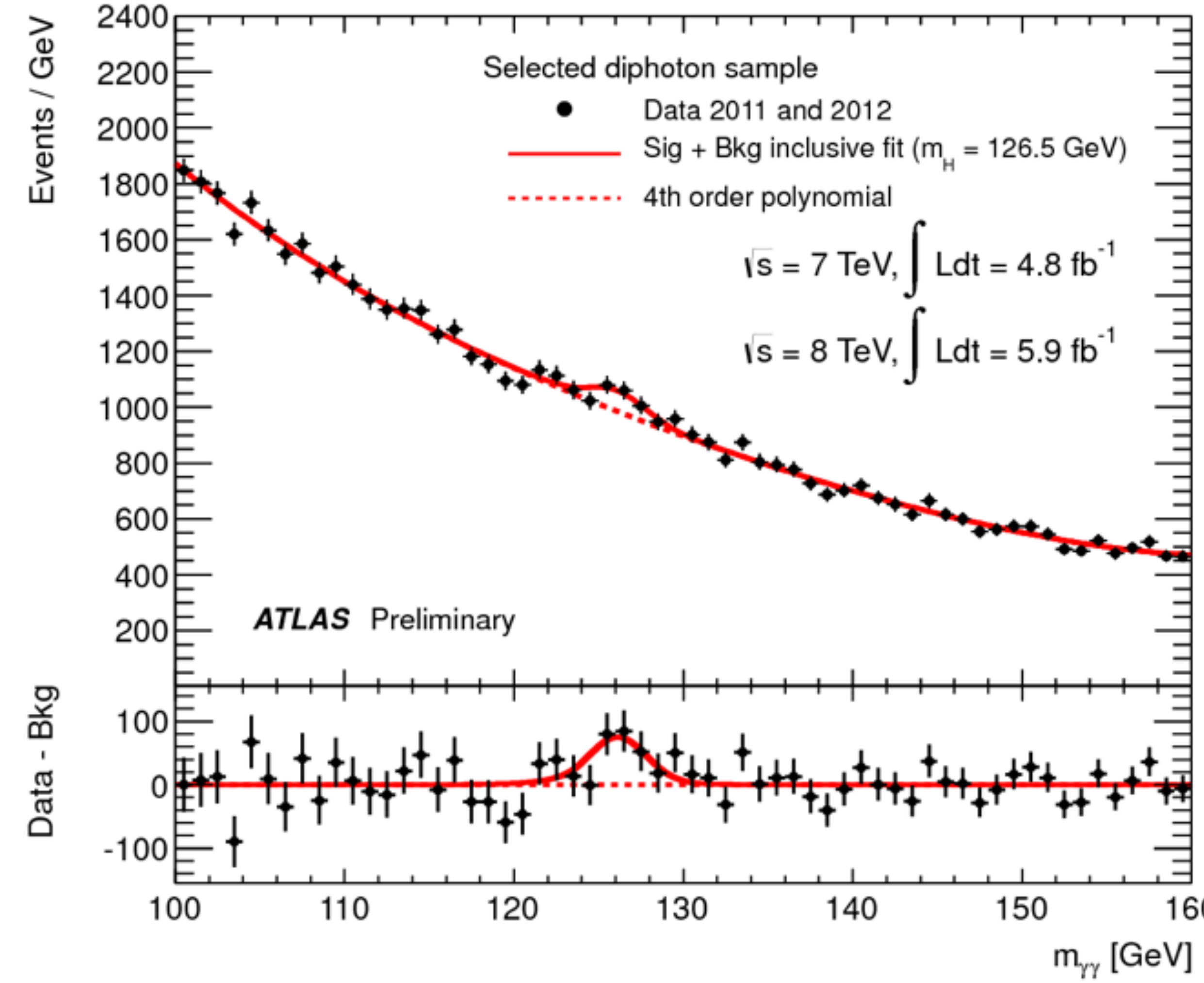




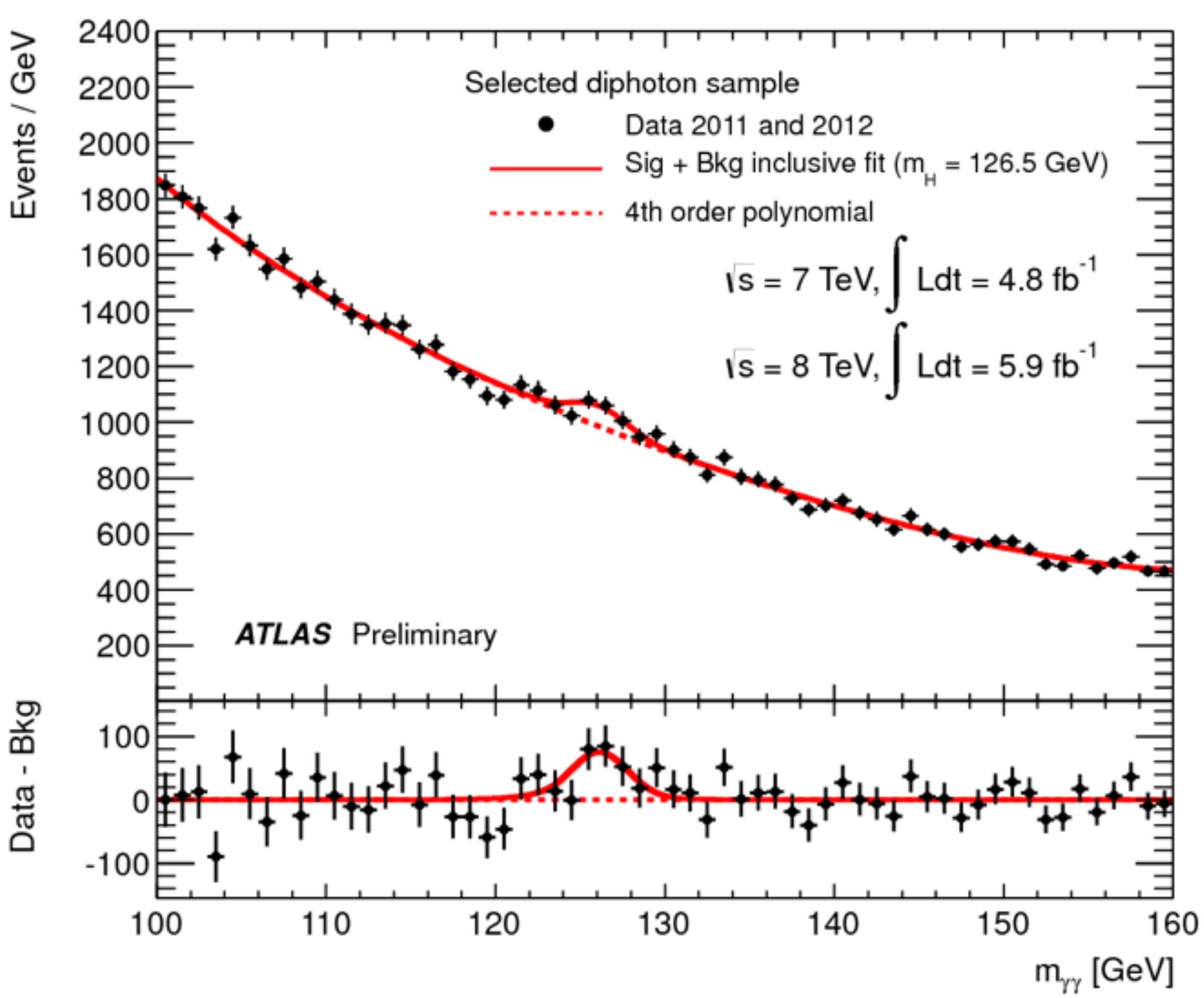
Poisson background

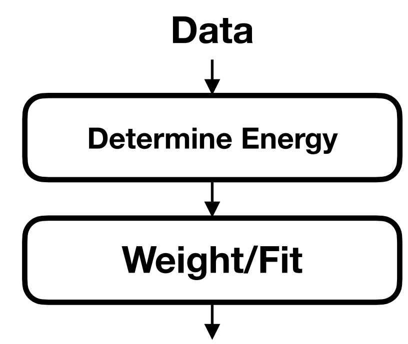
- Mean expected background is integral over spectrum in selection

- Signal strength also depends on selection



LHC Higgs discovery, weighted or Maximum Likelihood analyses



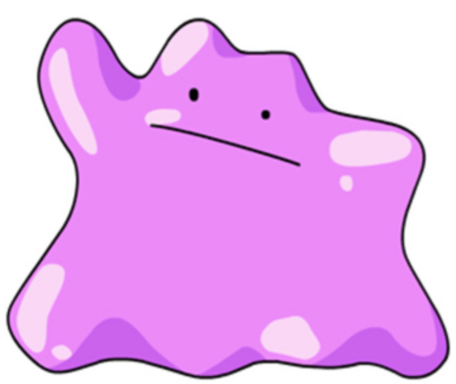


Why is ML (a little) better?

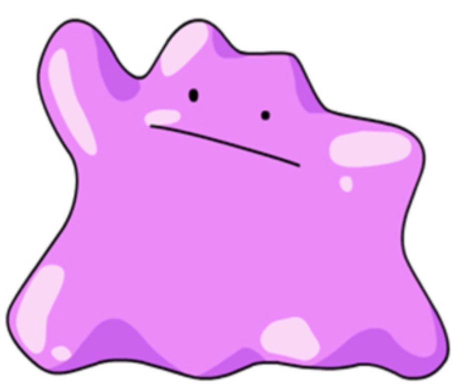
Review

- Multiple dimensions
- Key is what is probability of distribution of background after analysis steps
- Distribution depends on analysis choices

Common analysis problems (part 1 of ∞)

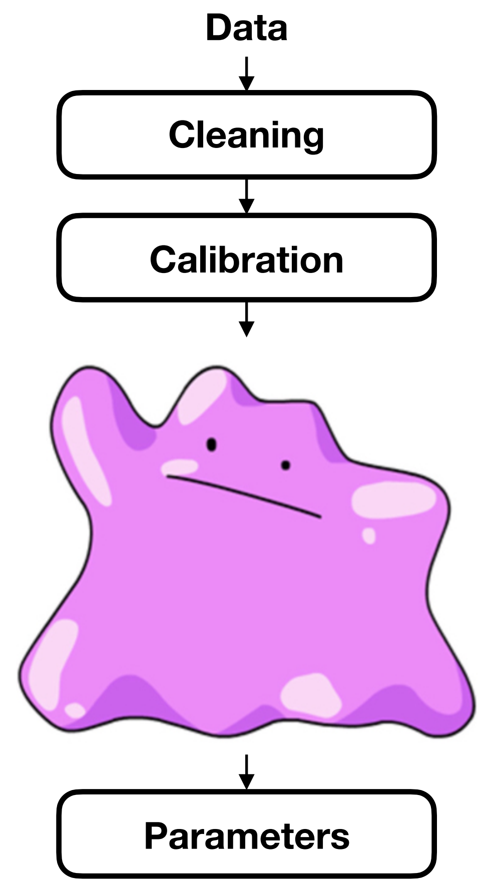


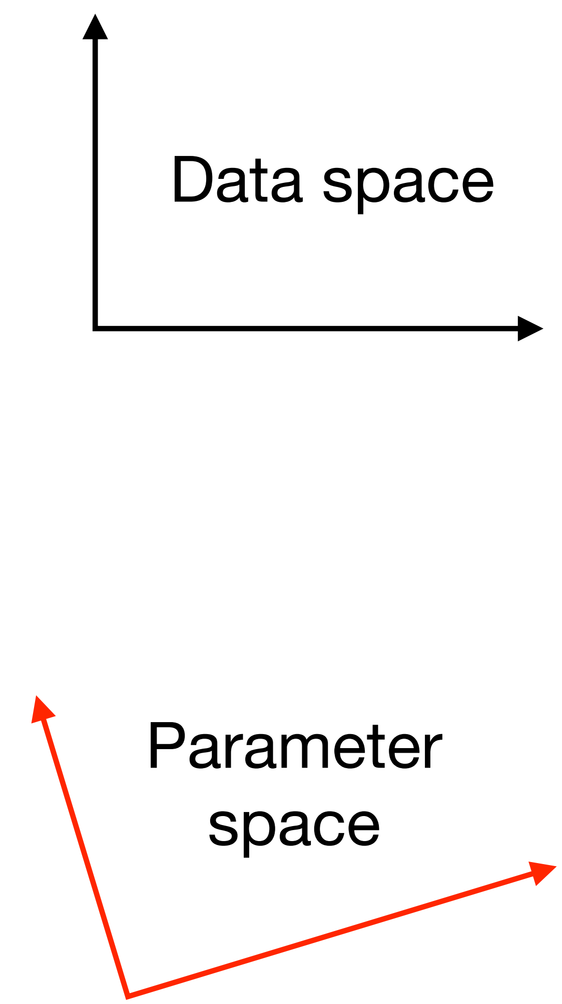
Inherited code, ML, and the analysis blob



Parameters as measurement

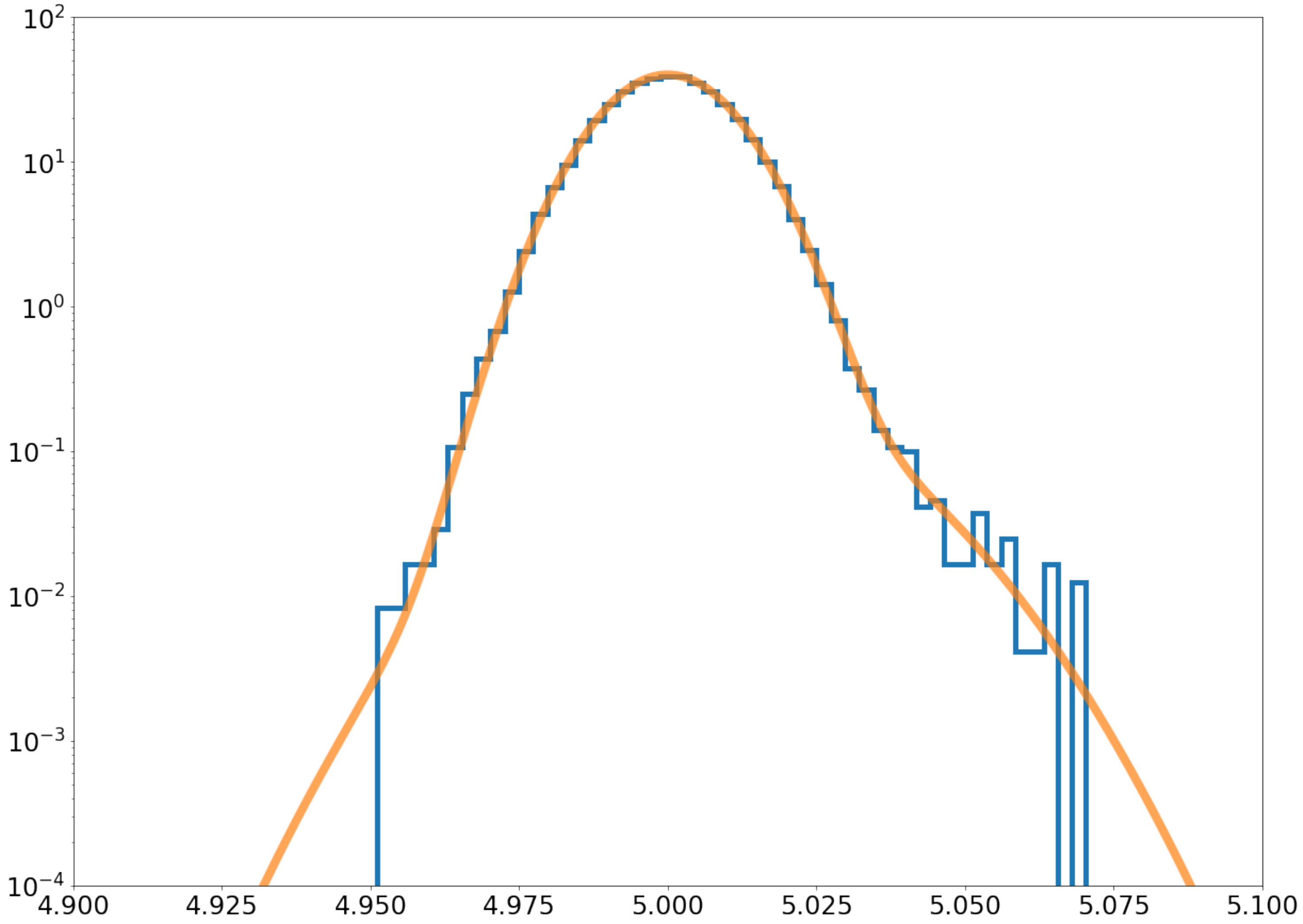
- New measurement ‘space’
- Horrible, non-linear, but treat as a new measurement





Treat data after blob as new measurement

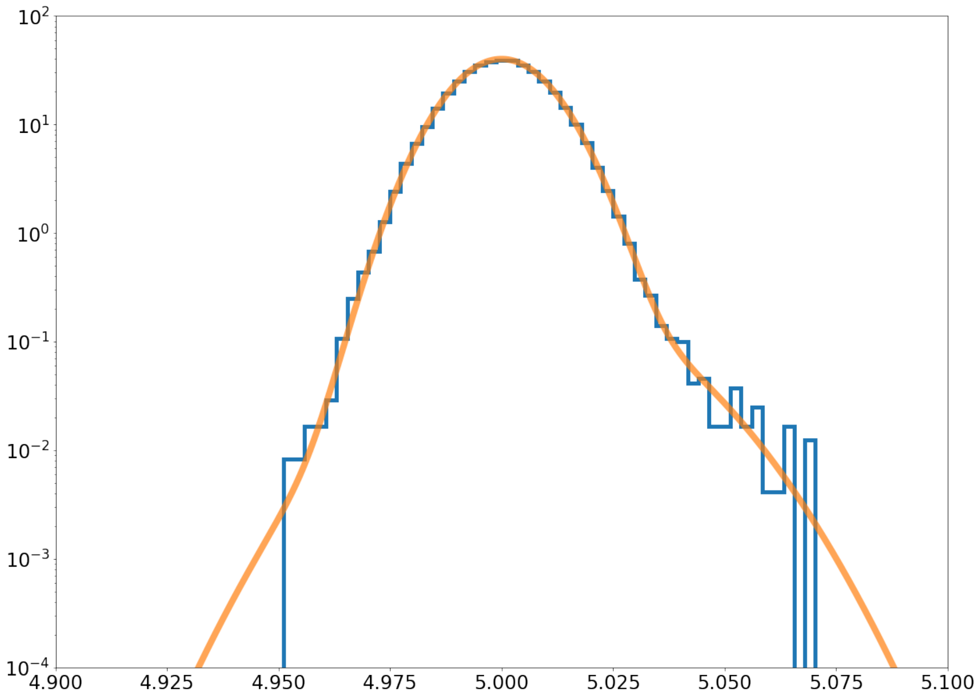
- Histogram background to determine the background pdf() in this new abstract space



Background & parameters

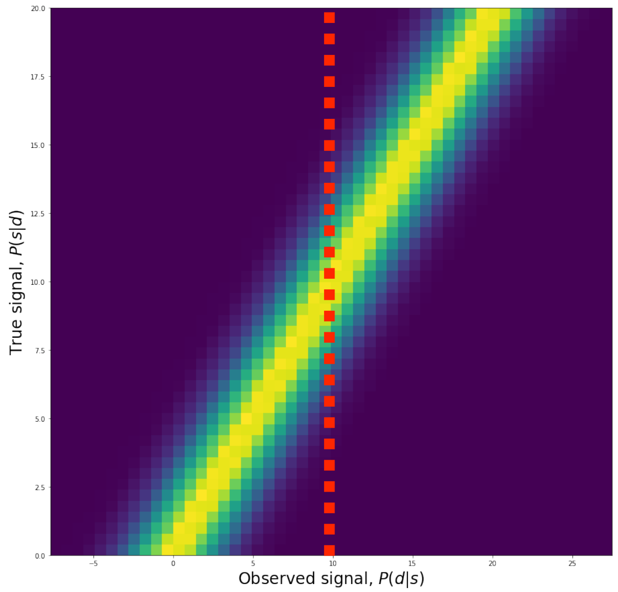
Zero signal examples can be

- Signal-free data
- MC simulation (signal turned off)



Statistics & Parameters

- Just treat like a measurement
- Background distribution for null hypothesis (σ significance)
- Confidence intervals
- All just like ‘measurement’ was the output of the blob
- Cost is no analytic errors



Parameter fitting

Problems with parameters

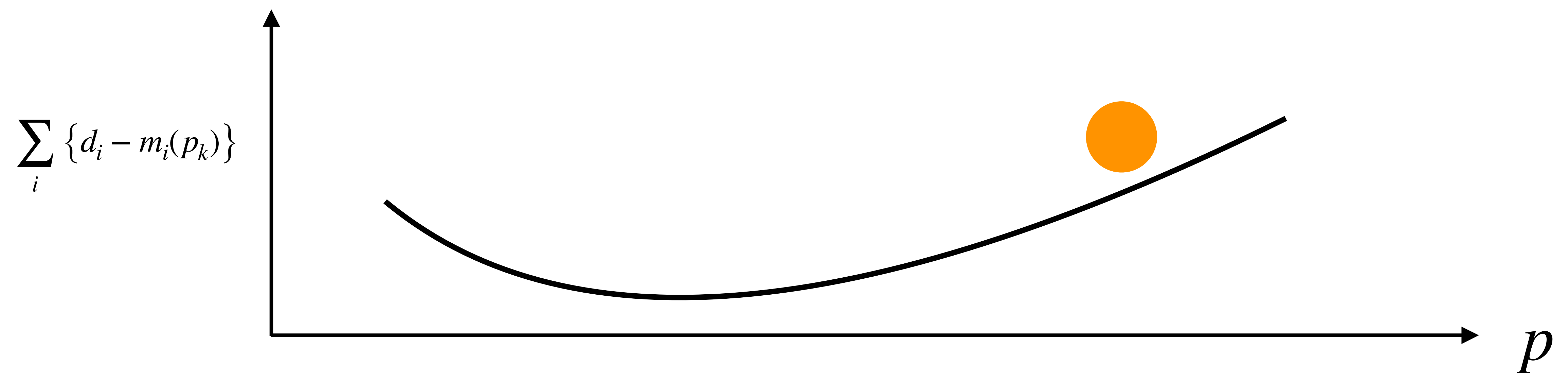
- Local minima
- Degeneracy, catastrophic errors, chaos

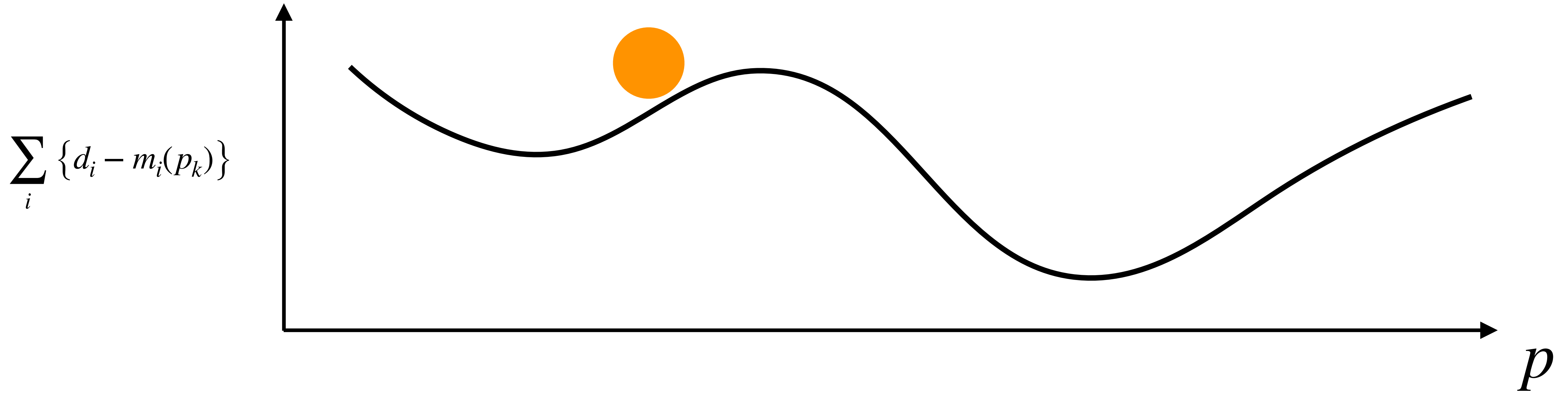
Fitting parameers

$$\sum_i \left\{ d_i - m_i(p_k) \right\}$$

$$\sum_i \frac{\left| d_i - m_i(p_k) \right|^2}{\sigma_i^2}$$

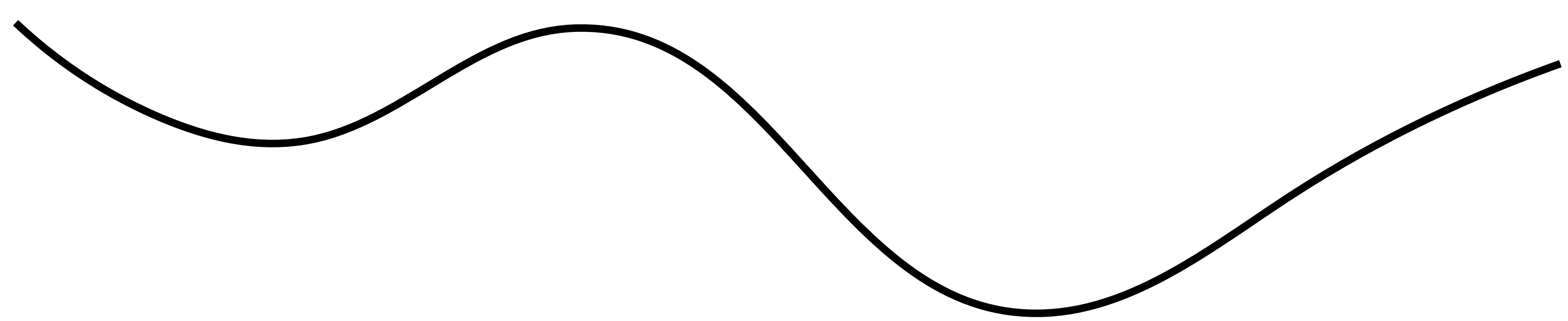
Local minima





Local minima

- Simulated annealing
- Genetic algorithms
- Markov Chain Monte Carlo sampling

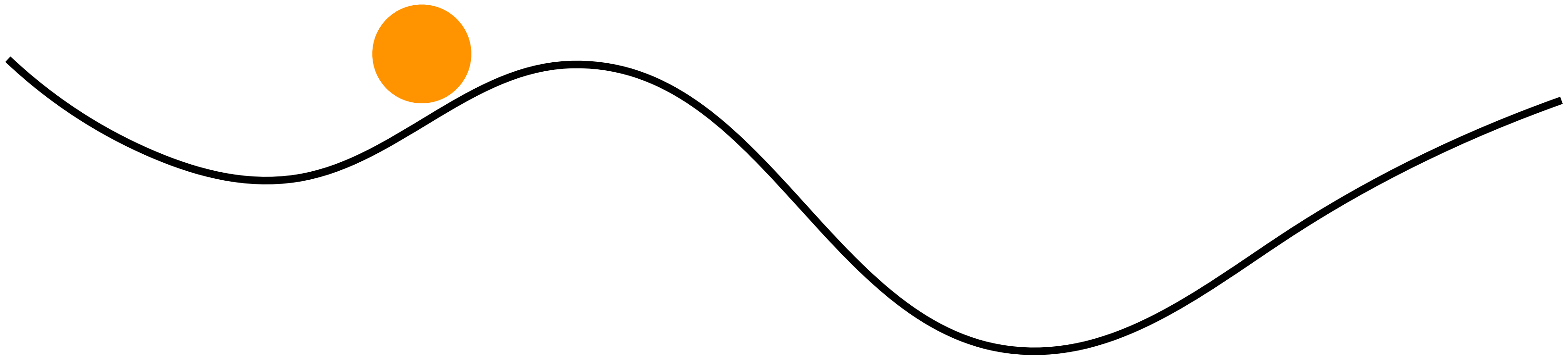


Local minima

Lots of options

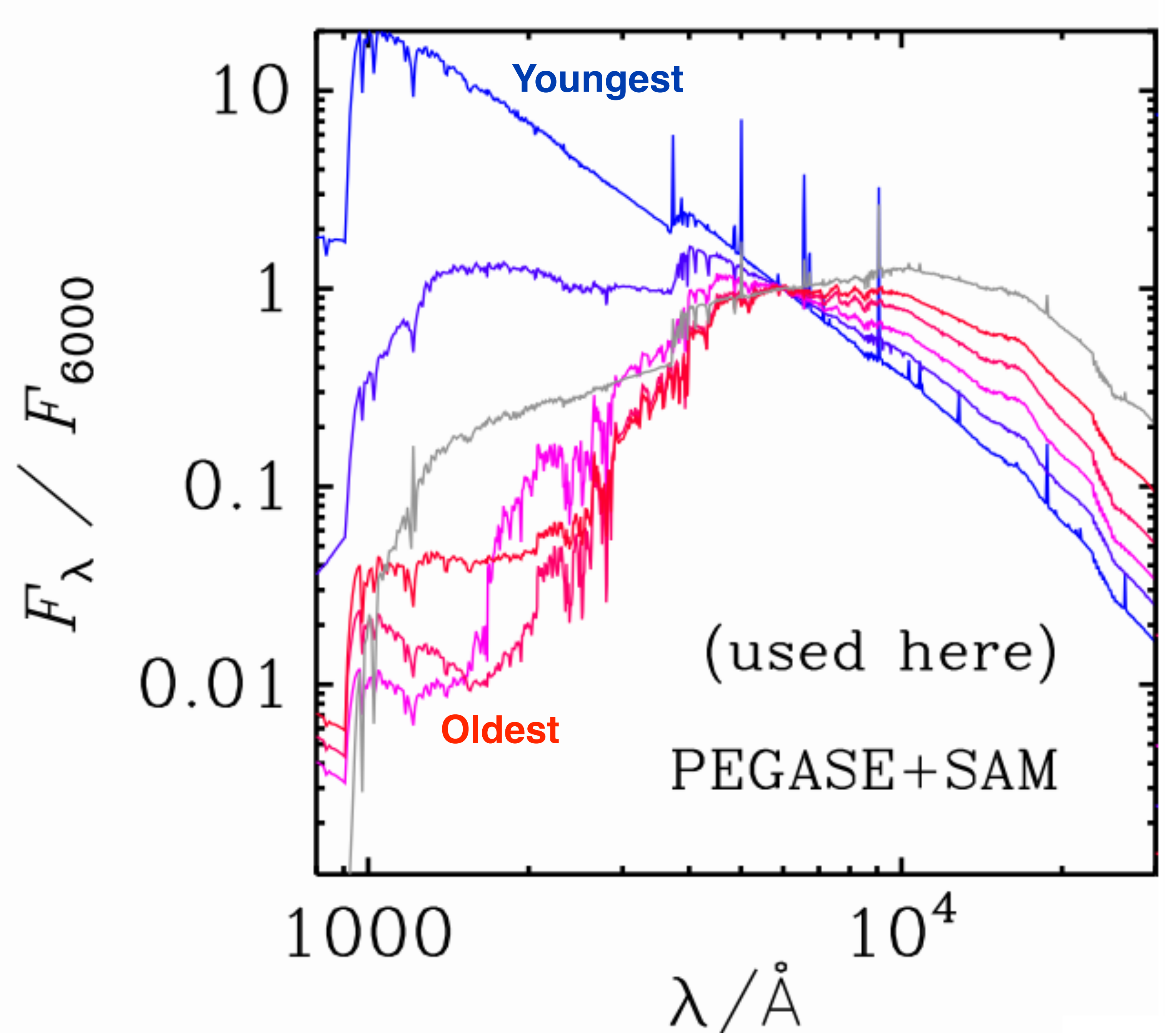
Make sure you actually need them…

- Related to degeneracies (telling you something)
- Adding more (or different) information can sometimes remove local minima

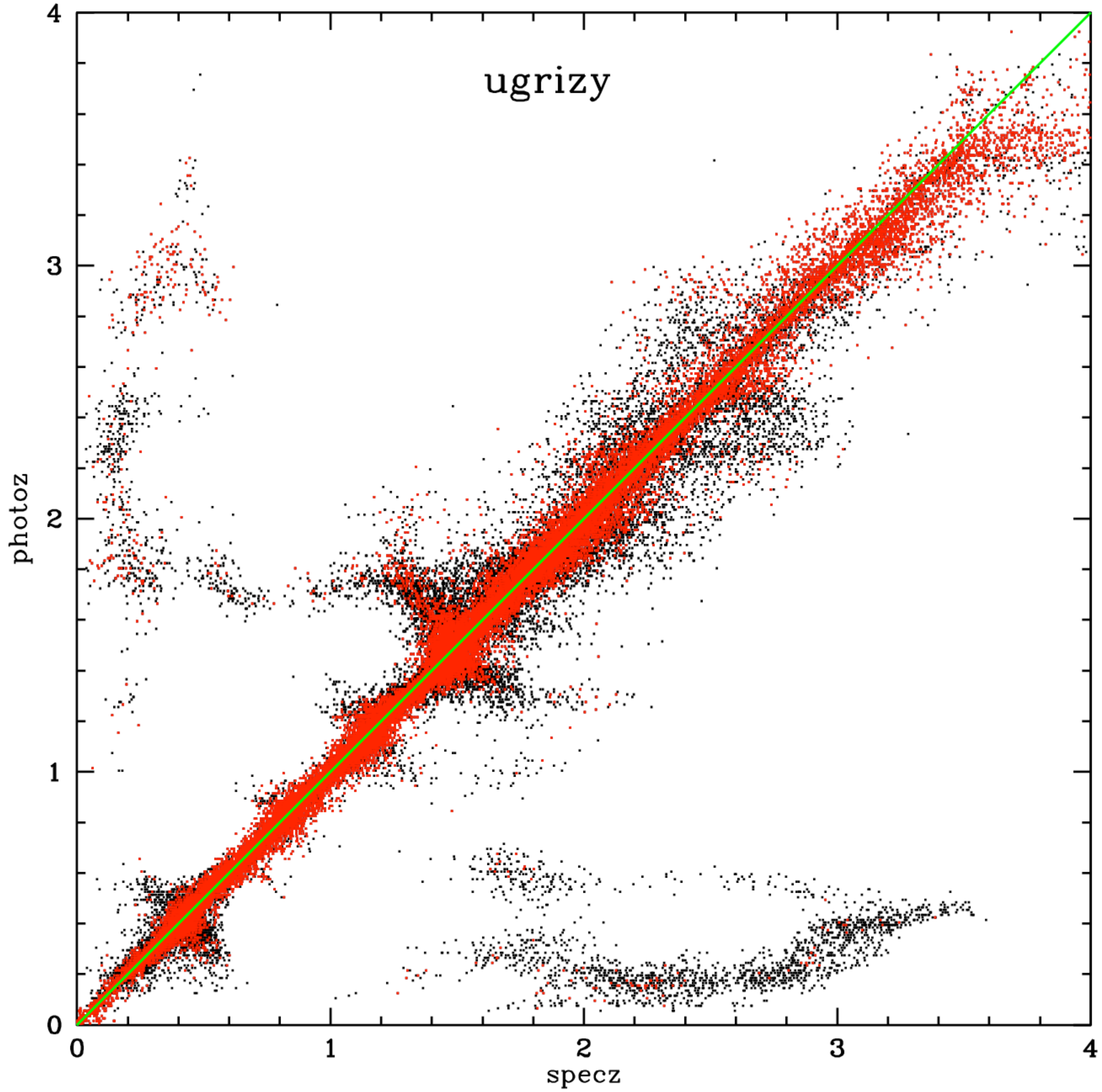


Degeneracy, catastrophic errors, chaos

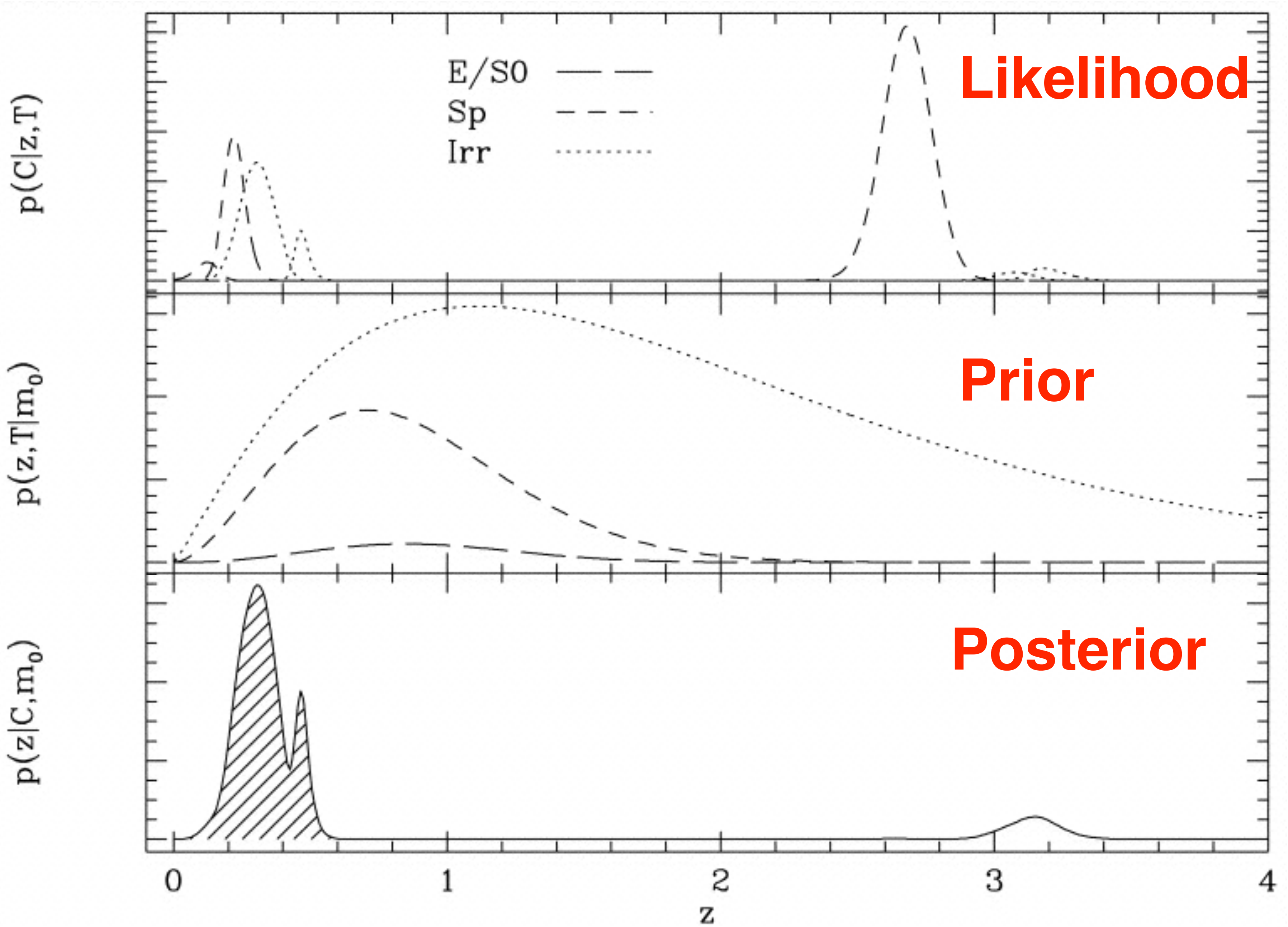
Catastrophic photo-z



Catastrophic photo-z



Catastrophic photo-z



$$P(\{s, n_1, n_2, \dots\} \,|\, d) = \frac{P(d \,|\, \{s, n_1, n_2, \dots\})\; P(s)\; P(n_1)\; P(n_2)\dots}{P(d)}$$

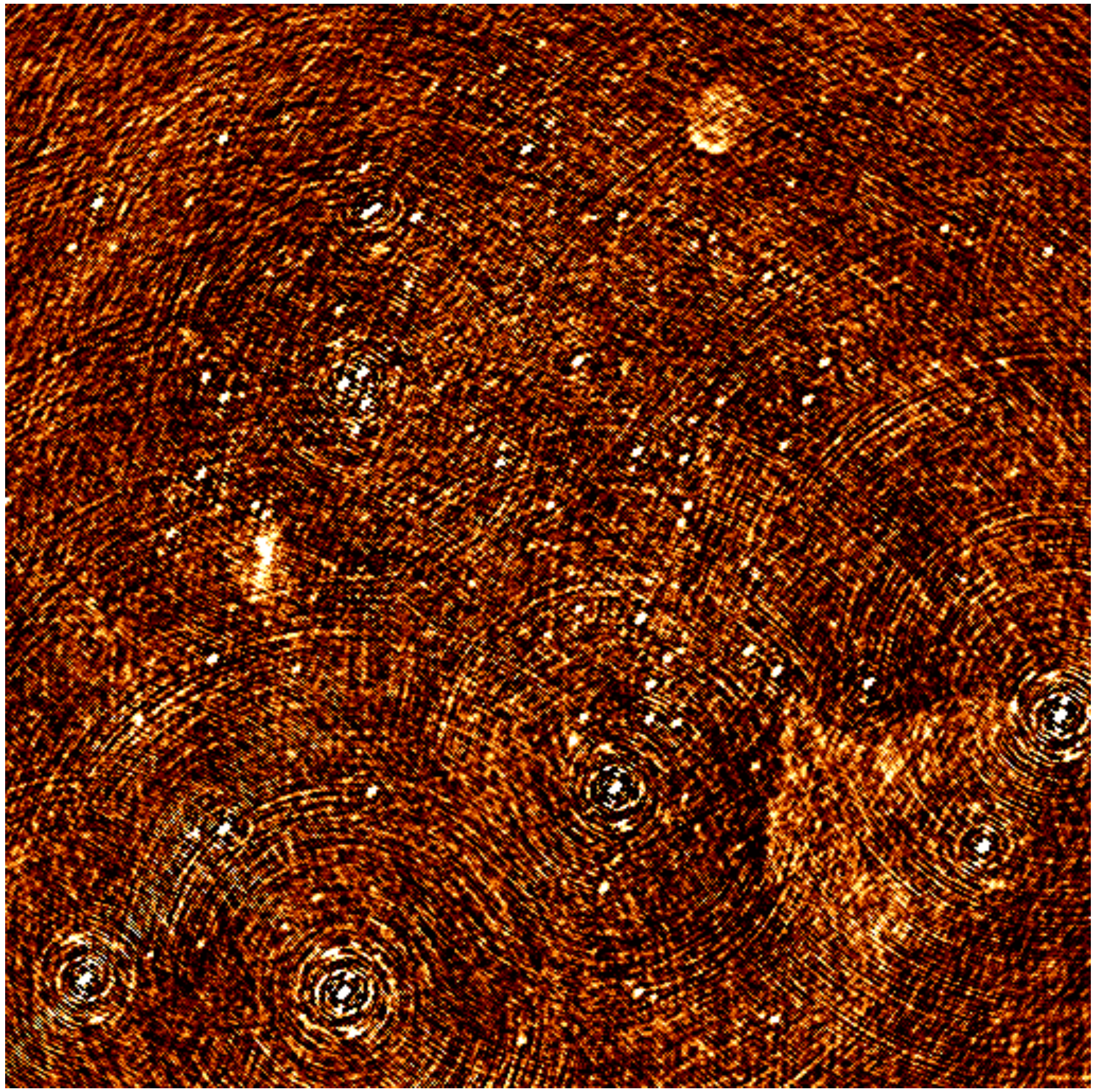
Benitez 2000

Degeneracy & catastrophic errors

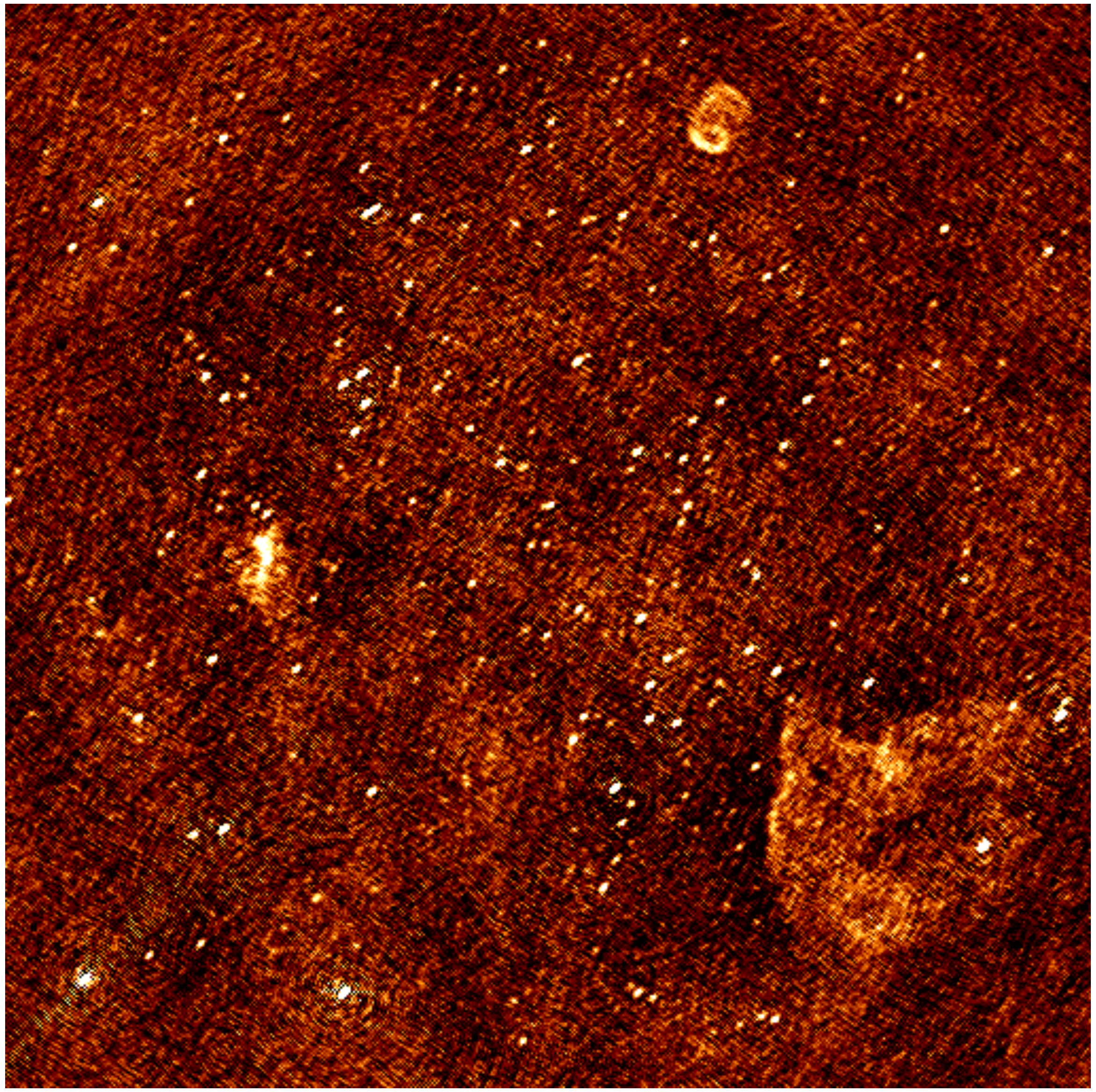
- Does it impact science?
- Can you get more information?
- Are priors your friend?

Chaotic algorithms (deconvolution; CLEAN)

clean-dirty.fits



clean-image.fits



Formally chaotic algorithms

- ‘Decision points’ in algorithm
- Gain or step control to help convergence
- Tend to have algorithm settings, which depend on implicit priors

Degeneracy, catastrophic errors, chaos

So what do you do?

- Be careful
- Test (implicit) assumptions
- Beware of walking out of applicability