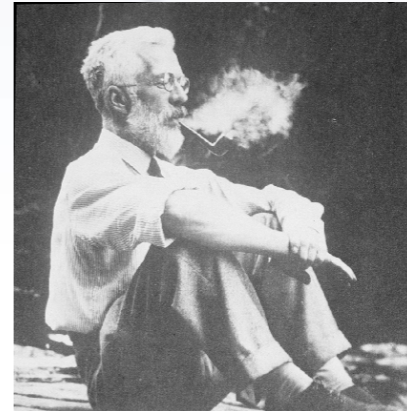
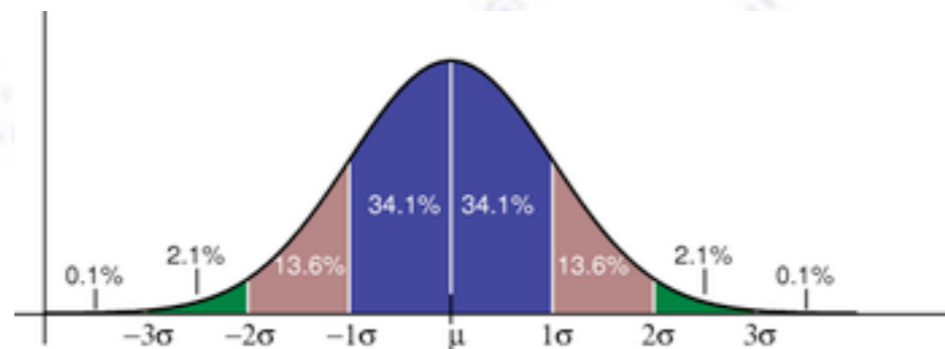


# Applied Statistics

## Fitting and Significance



Troels C. Petersen (NBI)



*"Statistics is merely a quantisation of common sense"*

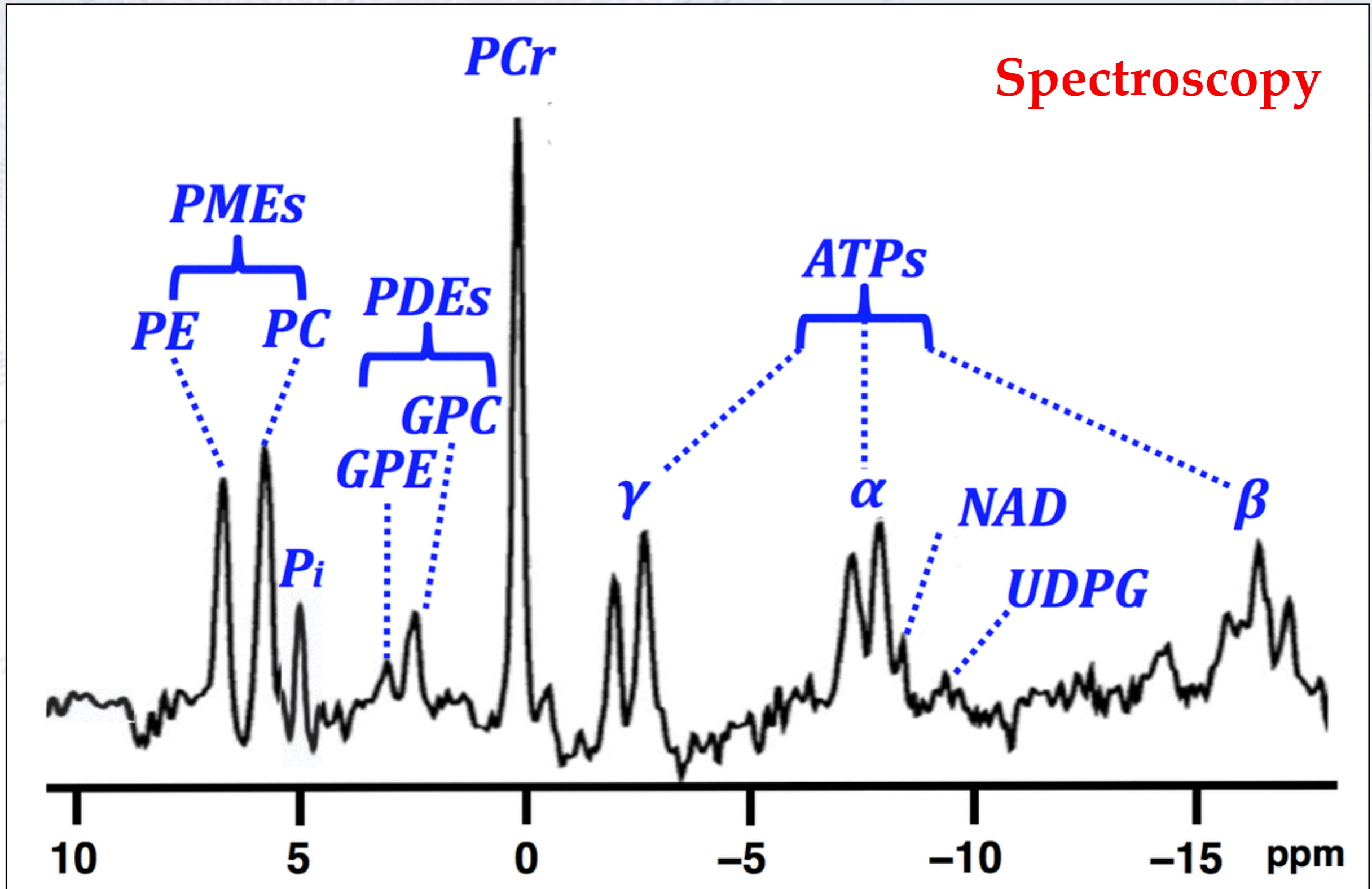
# Peaks on a background

Often data boils down to a single histogram, from which conclusions are drawn.

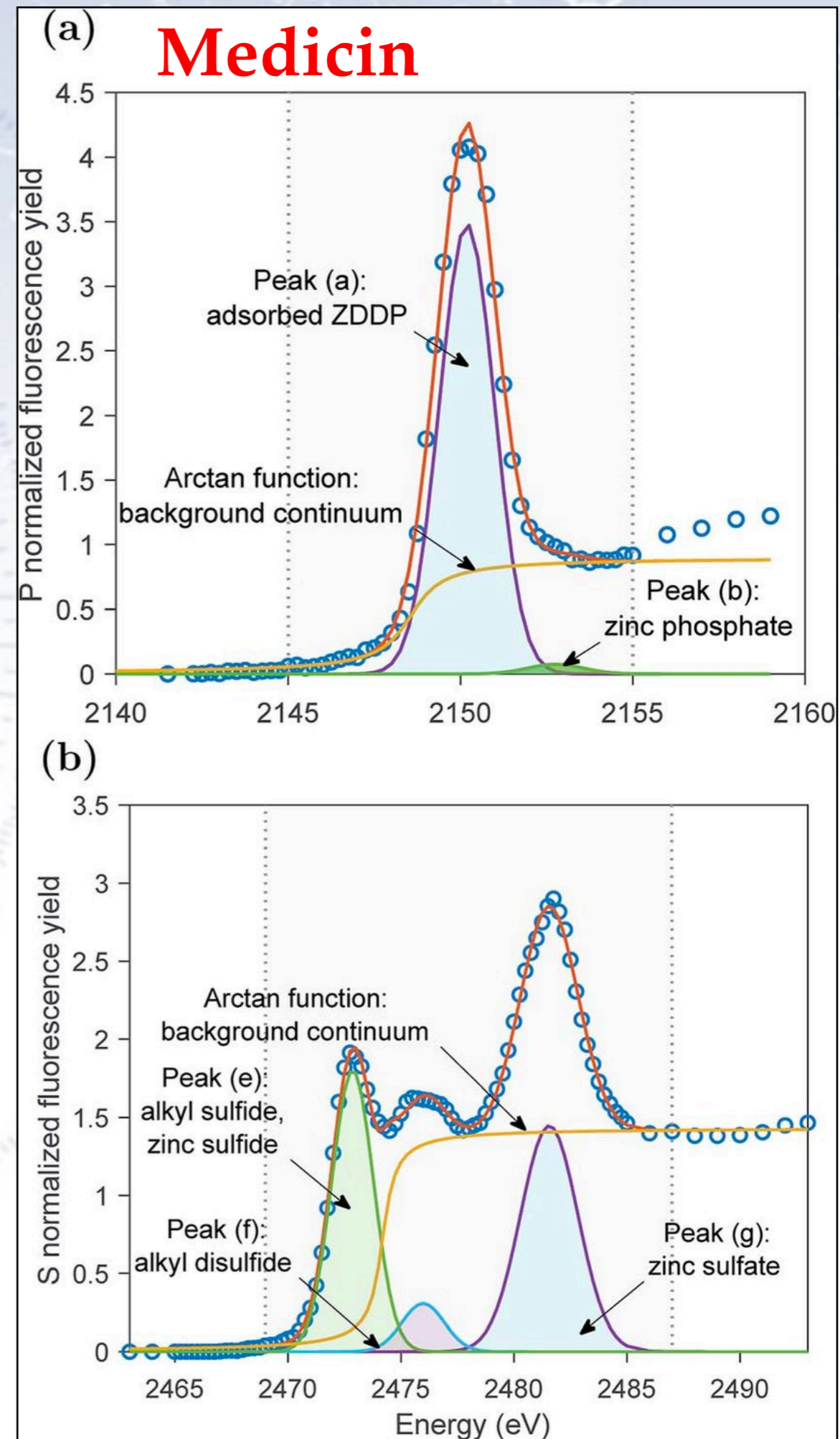
Though more complex representations might display sharper results, one often seeks simplicity, when making conclusions and attempting to convince others.

In such histograms, the most prominent and easily interpretable feature is a **peak**. Due to finite detector resolution these are mostly Gaussian, and so fitting one or more **Gaussian peaks on a background is a very typical case.**

# Fitting peaks - outside physics!

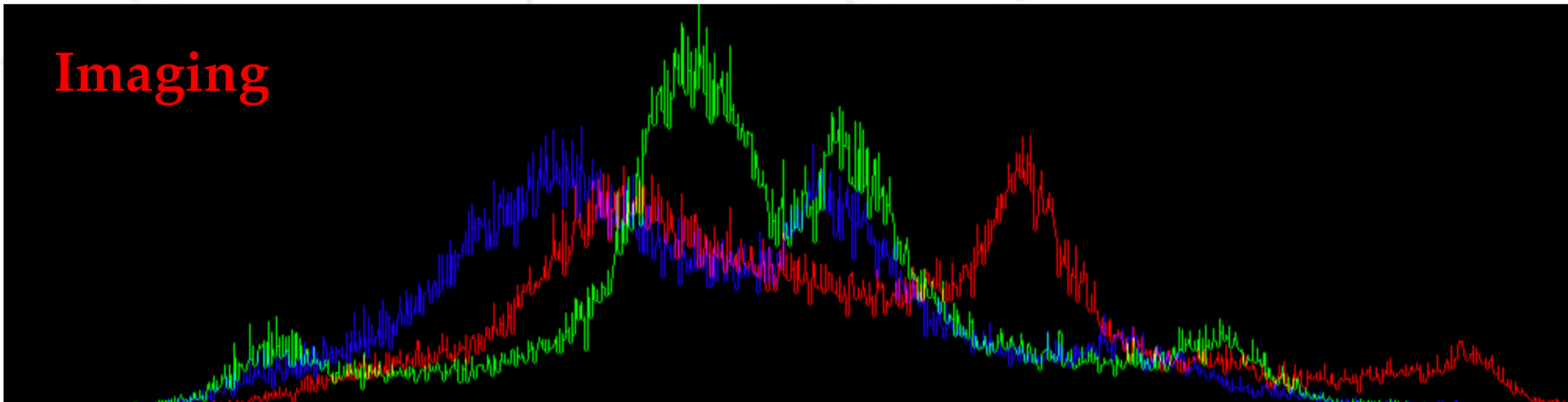


# Fitting peaks... outside physics!

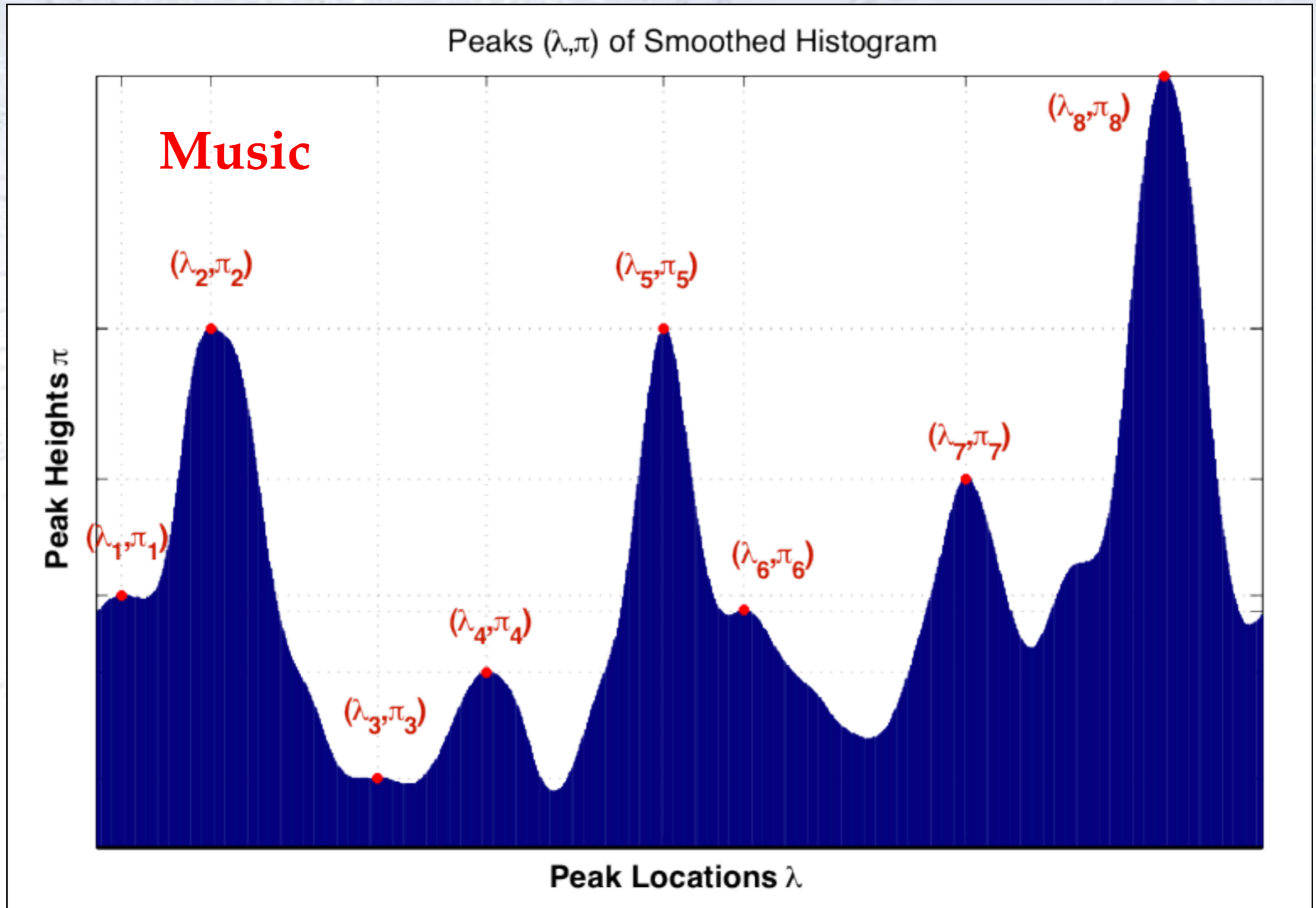


# Fitting peaks... outside physics!

**Imaging**



# Fitting peaks - outside physics!



# Peaks on a background

Often data boils down to a single histogram, from which conclusions are drawn.

Though more complex representations might display sharper results, one often seeks simplicity, when making conclusions and attempting to convince others.

In such histograms, the most prominent and easily interpretable feature is a peak. Due to finite detector resolution these are mostly Gaussian, and so fitting one or more **Gaussian peaks on a background is a very typical case.**

Of course one should make sure to give good starting values, or else the fit won't find the peak(s).

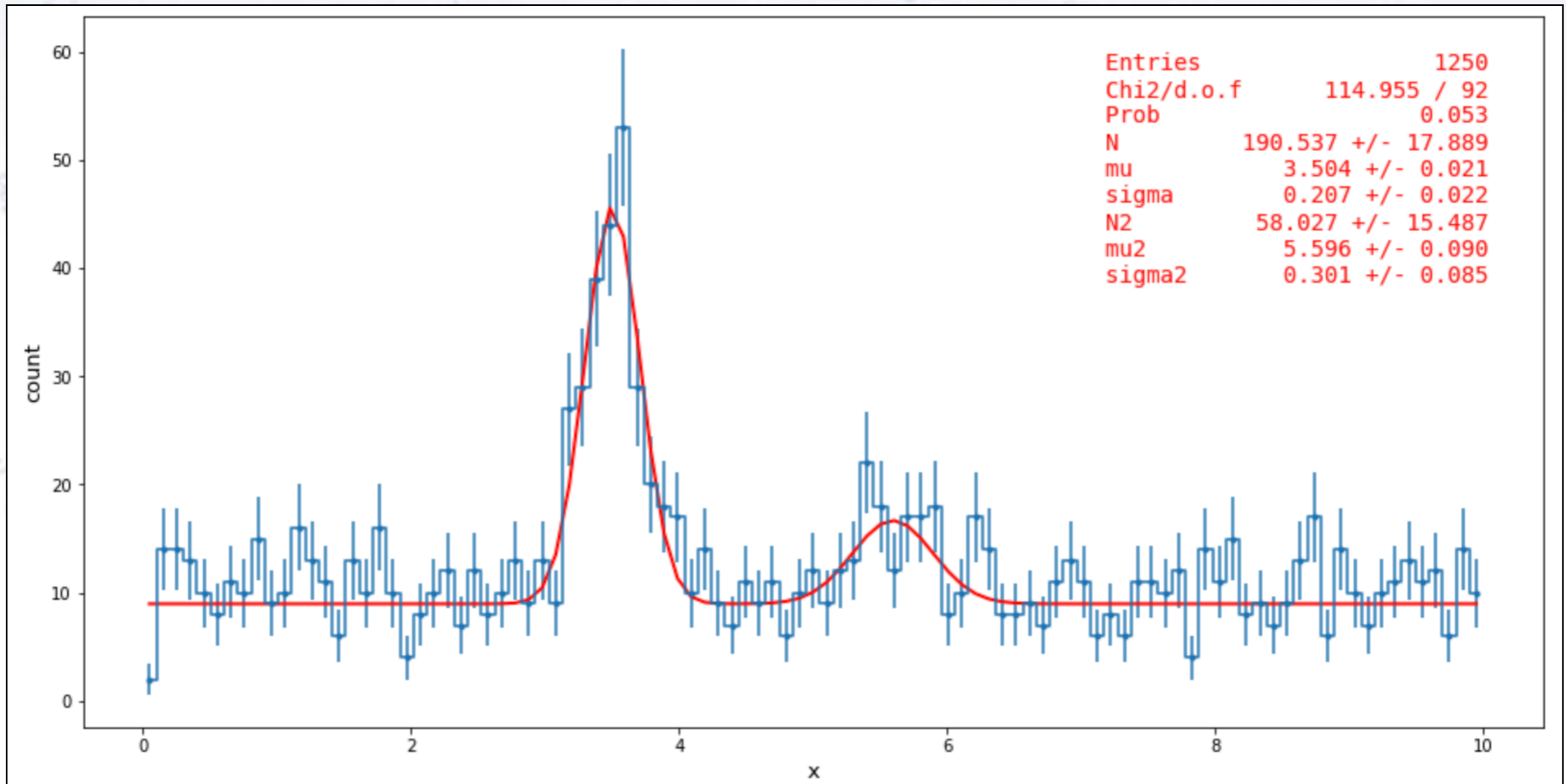
But what to do if one is searching for peaks?

How can one evaluate, if the peaks are real or just fluctuations of noise?

And can a combined / simultaneous fit to several peaks improve the result?

# Peaks on a background

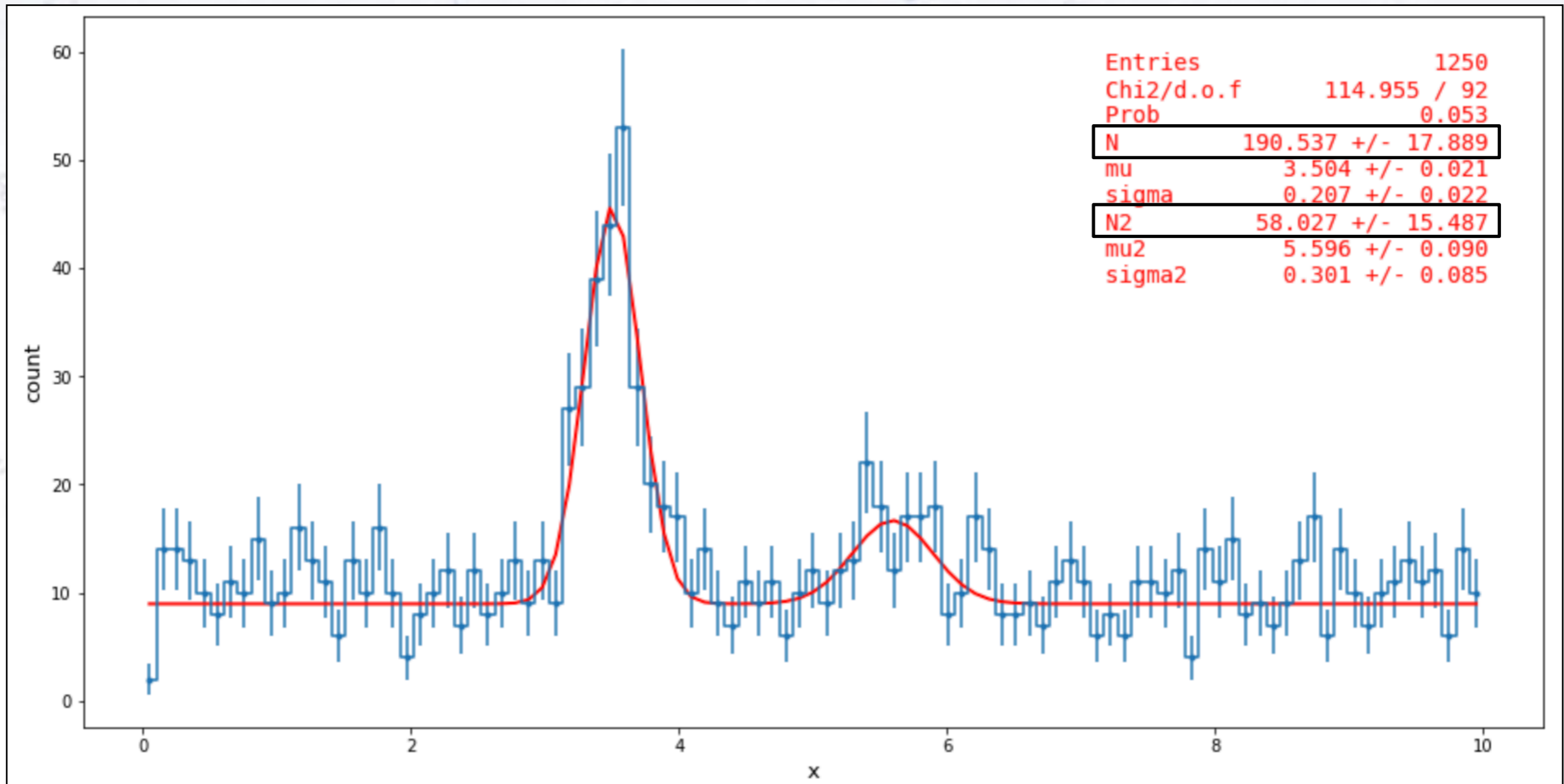
The (local) significance of a peak can be determined by considering how far away from zero the normalisation is. In the below case, the two peaks have significance ?





# Peaks on a background

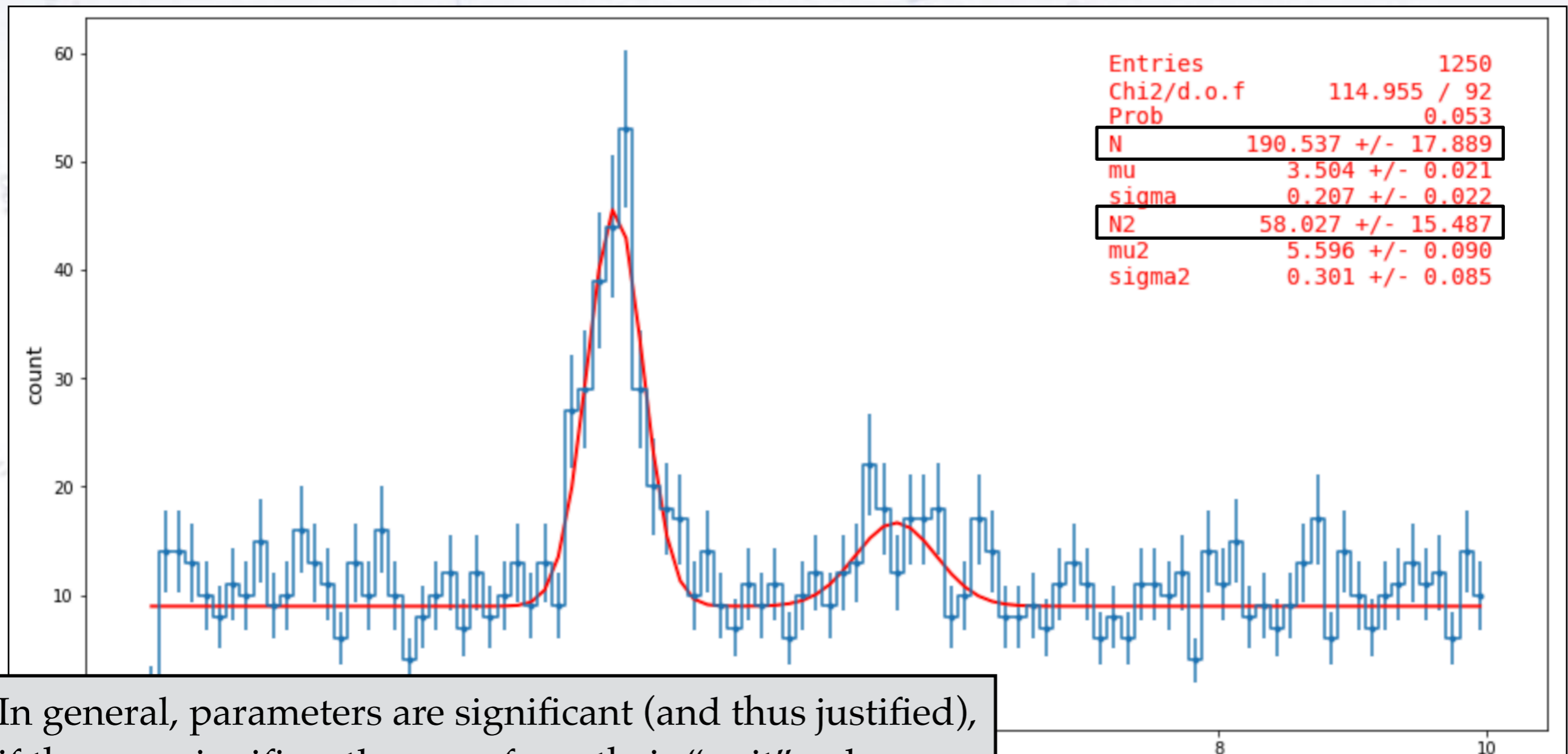
The (local) significance of a peak can be determined by considering how far away from zero the normalisation is. In the below case, the two peaks have significance:  
Peak 1:  $190.5 / 17.9 = 10.6 \sigma$  (Certain!). Peak 2:  $58.0 / 15.5 = 3.7\sigma$  (likely).



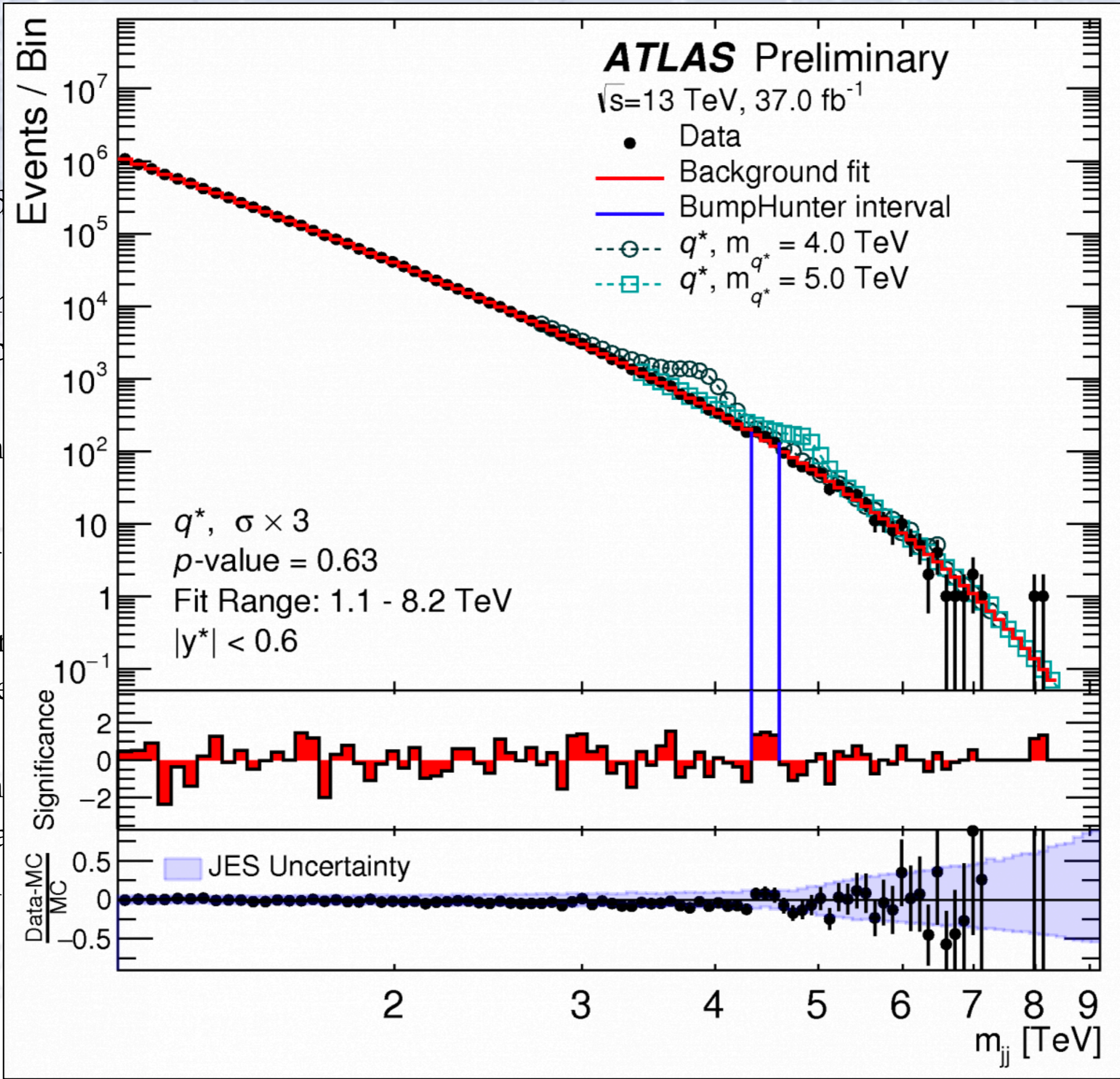
# Peaks on a background

The (local) significance of a peak can be determined by considering how far away from zero the normalisation is. In the below case, the two peaks have significance:

Peak 1:  $190.5 / 17.9 = 10.6 \sigma$  (Certain!). Peak 2:  $58.0 / 15.5 = 3.7\sigma$  (likely).



In general, parameters are significant (and thus justified), if they are significantly away from their “unit” value.



Often o

Though  
 simplic

In such  
 Due to  
**Gaussi**

Of cour  
 find the

But wh  
 How ca  
 And ca

n.

h seeks

ak.

or more

n't

# Three attempts at answers

**a) What to do if one is searching for peaks?**

First of all, establish a good (non-peaking) background model (crucial!).

Secondly, determine the expected signal resolution.

Only then, start fitting the background with a potential peak with **fixed width**.

**b) How can one evaluate, if the peaks are real or just fluctuations of noise?**

The significance of the (floating) amplitude tells you the local significance.

As you might be searching in many places, this reduces your certainty to the global significance:

$$p_{\text{global}} = 1 - (1 - p_{\text{local}})^N \approx N p_{\text{local}}$$

**c) And can a combined/simultaneous fit to several peaks improve the result?**

Yes! If you are fitting several peaks, which you expect to have overlapping parameters (e.g. same width), then fitting with a reduced number of parameters makes the fit more constrained, accurate, and convergent.

Note, that in the last case, this is why a “calibration channel” is great to have.

# How many parameters to use?

*Numquam ponenda est pluralitas sine necessitate.*

Plurality must never be posited without necessity.

[William of Ockham, 1330ies]

*Entia non sunt multiplicanda praeter necessitatem.*

Entities must not be multiplied beyond necessity.

[John Punch, 1650ies]

# How many parameters to use?

*Numquam ponenda est pluralitas sine necessitate.*

Plurality must never be posited without necessity.

[William of Ockham, 1330ies]

*Entia non sunt multiplicanda praeter necessitatem.*

Entities must not be multiplied beyond necessity.

[John Punch, 1650ies]

Do not put in more parameters than necessary and explainable! What does this mean?

With too few, you might have a poor model, where some parameters try to “make up for” the missing degrees of freedom.

With too many, you might “dilute” your model strength, and render the crucial parameters insignificant despite being so.