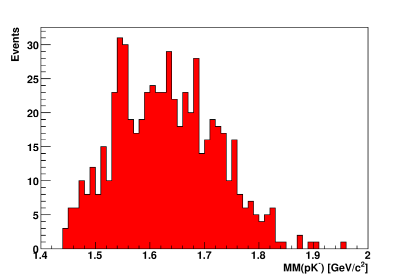
Bayesian Analysis Methods for Hadronic Spectroscopy

D.G. Ireland

Exotic Hadron Spectroscopy Workshop,
September 26-27, 2016

Statistical Evidence from a Spectrum

Given a spectrum with a feature, what is the evidence that there is a signal?

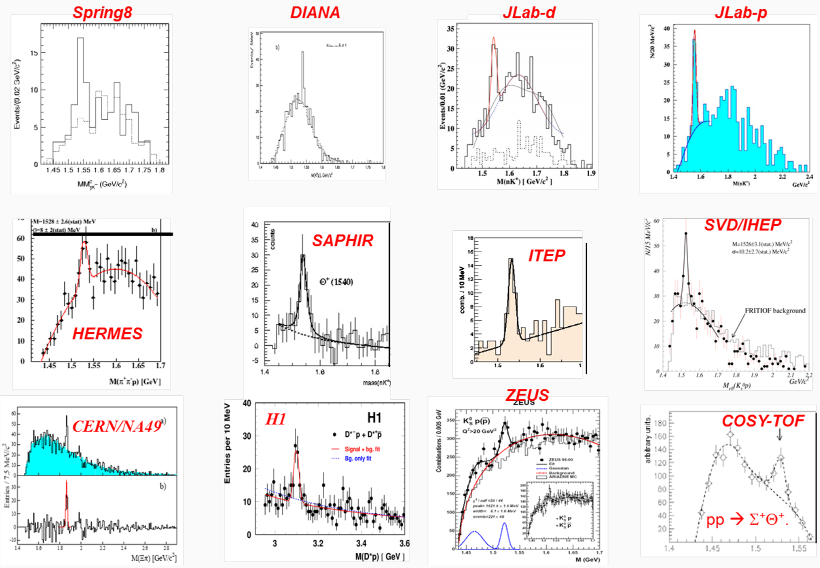


- Do you know where the peak should be?
- Can you model the background?
- Are you prepared to make a bet?

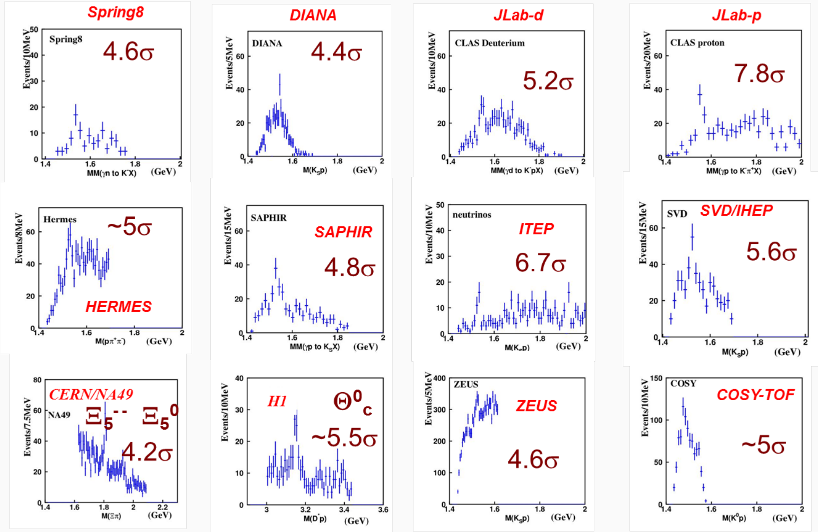
Pentaquark Discovery?

Two years later...

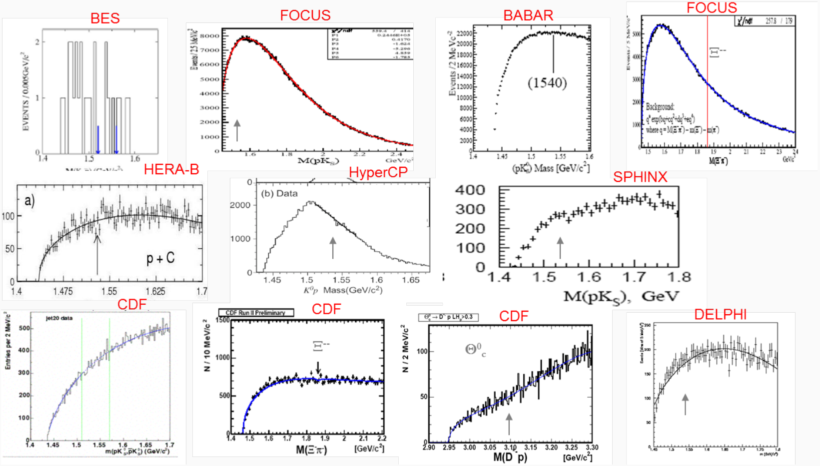
Positive Evidence



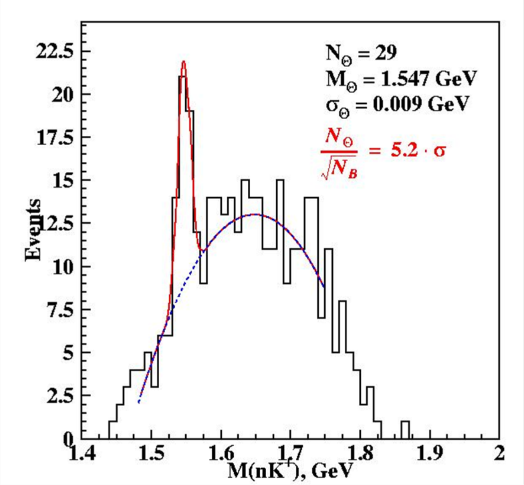
Data Undressed

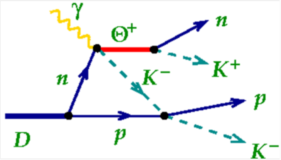


Non-evidence

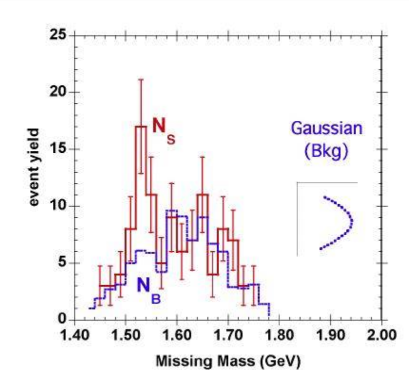


Evidence from CLAS





Statistical Significance



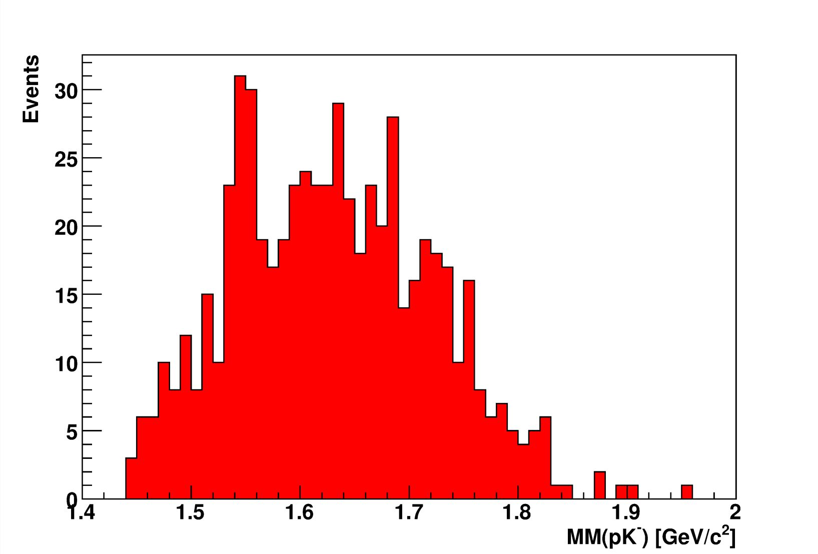
A simple statistic is

$$\frac{N_s}{\sqrt{N_b}}$$

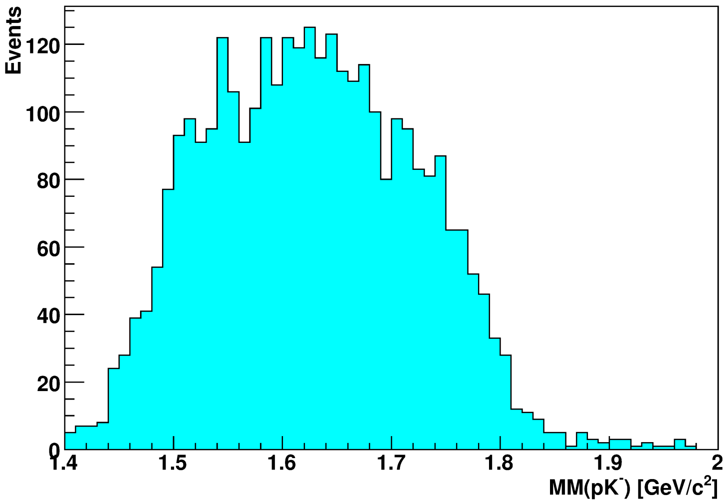
If this is > 5 then the probability of feature being a fluctuation is 10^{-6}

Do you want to bet £1M against £1?

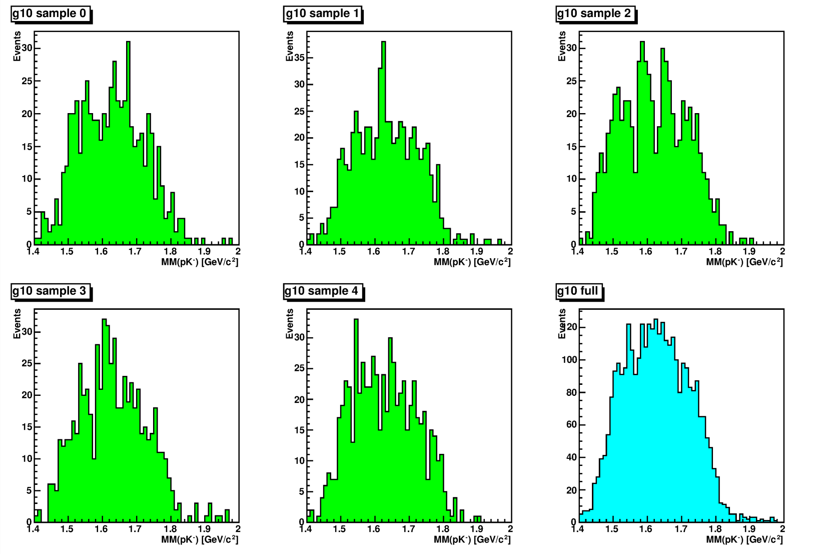
First data



Higher Statistics Run



Samples from Higher Statistics Run



Probability Distribution: Peak Height

Set up a data model M_0 for no signal:

$$S_0(x_i) = \beta f_b(x_i)$$

and a model M_p for a peak and background:

$$S_P(x_i) = \beta \left(f_b(x_i) + \rho f_s(x_i) \right),$$

where $f_b(x_i)$ and $f_s(x_i)$ are functions describing the shape of background and signal distributions. x_i are the histogram bin centres.

f_b depends on parameters c_j, and if f_s is modelled by a gaussian, it has parameters μ and σ.

The spectrum is a histogram with counts n_i in each bin i.

The likelihood of obtaining a spectrum $\{n_i\}$, given a set of parameters can be written:

$$P\left(\{n_i\}|\rho, \beta, \mu, \sigma, c_j, M_P\right) = \prod_i \frac{S_i^{n_i} \exp(-S_i)}{n_i!}$$

Probability Distribution: Peak Height

Bayes Theorem allows the definition of a joint (posterior) probability:

$$P\left(\rho, \beta, \mu, \sigma, c_j | \{n_i\}, M_P\right) \propto P\left(\{n_i\} | \rho, \beta, \mu, \sigma, c_j, M_P\right) P\left(\rho, \beta, \mu, \sigma, c_j | M_P\right),$$

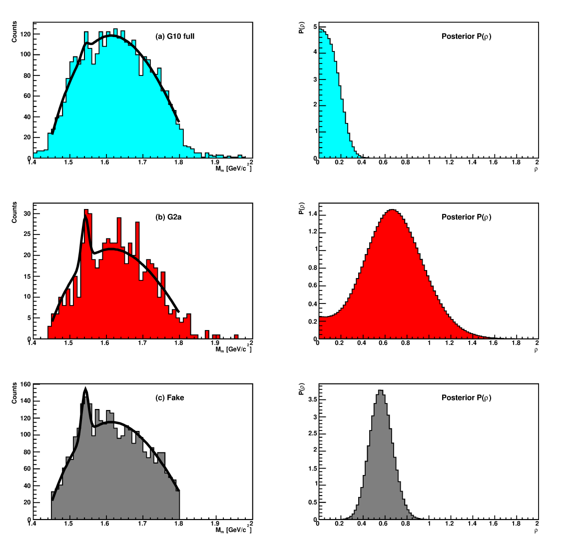
where the last factor is the prior.

Marginalisation allows the integration over parameters that are not of interest:

$$\begin{aligned} P\left(\rho | \{n_i\}, M_P\right) &= \int P\left(\rho, \beta, \mu, \sigma, c_j | \{n_i\}, M_P\right) d\beta d\mu d\sigma dc_j \\ &\propto \int P\left(\{n_i\} | \rho, \beta, \mu, \sigma, c_j, M_P\right) P\left(\rho, \beta, \mu, \sigma, c_j | M_P\right) d\beta d\mu d\sigma dc_j, \end{aligned}$$

Hence, evidence for a peak will be apparent in the distribution $P\left(\rho | \{n_i\}, M_P\right)$ showing a maximum at a non-zero value of ρ.

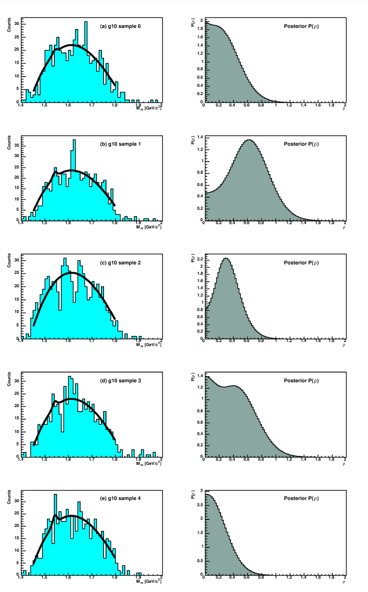
Probability Distribution: Peak Height



$P\left(\rho|\{n_i\}, M_P\right)$ distributions for

1. High stats (original $\times 5$) run
2. Original run
3. Fake spectrum (original $\times 5$ stats; keep original signal to background)

Peak Height

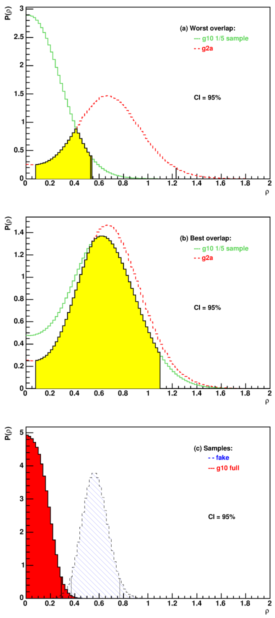


- Evaluate $P\left(\rho|\{n_i\}, M_P\right)$ distributions for the subsamples
- At least two show suggestion of non-zero peak heights.

Peak Height

Plots show overlaps of $P\left(\rho|\{n_i\}, M_P\right)$ distributions for combinations:

1. Original run and subsample with no peak
2. Original run and subsample with peak
3. High stats run and fake



Final thoughts...

- Inference based on information obtained from joint posterior probability, given a data model.
- Modern computing allows the use of Markov Chain Monte Carlo methods.
- Marginalisation over peak position $\equiv$ look elsewhere effect.
- Can use Bayes factors

$$\frac{P(M_P|\{n_i\})}{P(M_0|\{n_i\})}$$

- Alternatively, mixture models

$$S_M = \alpha S_0 + (1-\alpha) S_P$$

  can show which hypothesis is more likely
- Both Bayesian and frequentist methods should be deployed.