Bayesian and frequentist approaches to resonance searches

arXiv:1902.03243

Andrew Fowlie

July 2, 2019

Nanjing Normal University

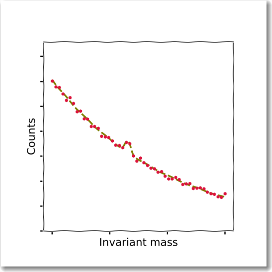
Table of contents

Background

What is that?

A new particle? or just a fluctuation?

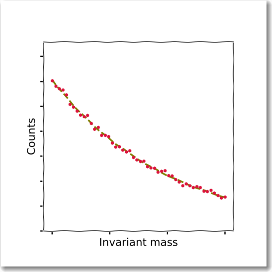
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

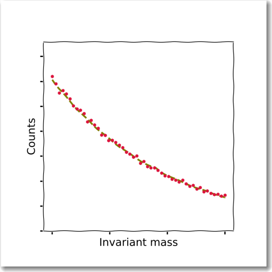
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

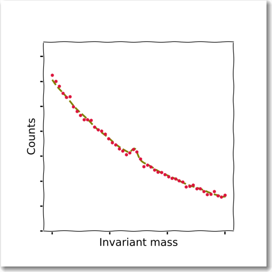
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

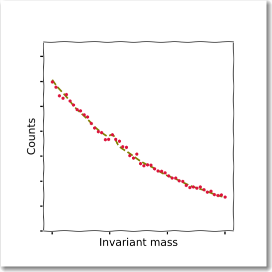
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

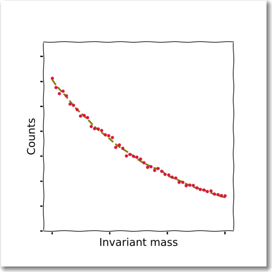
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

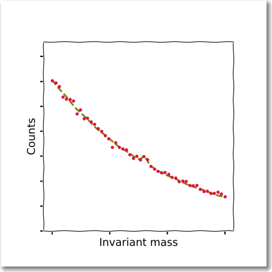
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

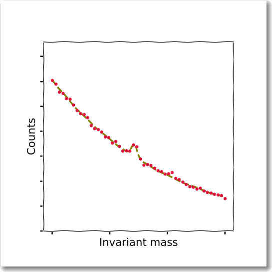
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

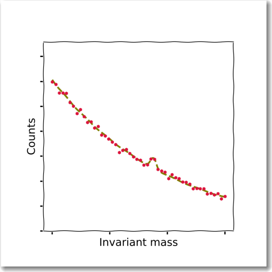
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

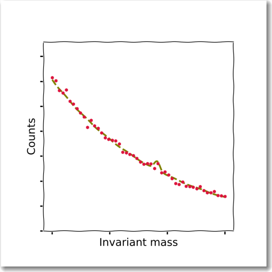
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

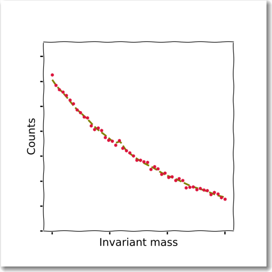
How can we characterise our uncertainty?



What is that?

A new particle? or just a fluctuation?

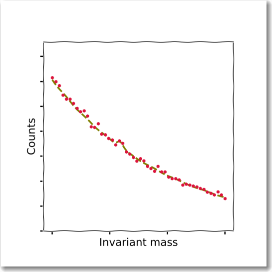
How can we characterise our uncertainty?



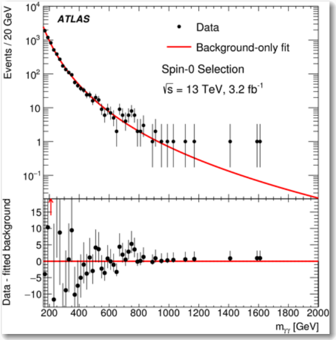
What is that?

A new particle? or just a fluctuation?

How can we characterise our uncertainty?



Another 750 GeV?



or something real? Should you write a paper about it? Announce a press conference? Start writing your Nobel prize speech?

Interpretations

Frequentist: what is probability?

Probabilities are not degrees of certainty or belief.

Probabilities are frequencies at which events occur in identical repeat experiments.

$$P(A) = \lim_{N \to \infty} \frac{n_A}{N}$$

Frequentist: what can we do?

We cannot quantify our uncertainty about the resonance.

We can attempt to control the frequency at which we would make a type-1 error.

Type-1 error: Reject null hypothesis when it is true.

We must specify a null hypothesis, H_0, and a desired type-1 error rate, α. We reject H_0 at a pre-chosen significance α or we do not.

The rate α (implicitly) chosen to be about 10^{-7} (5σ) in particle physics.

We construct a test-statistic that measures discrepancies between data and the null hypothesis, e.g. the log-likelihood ratio,

$$q \equiv -2 \ln \frac{\max_{\boldsymbol{\theta}_1} P\left(D \,|\, M_1, \boldsymbol{\theta}_1\right)}{\max_{\boldsymbol{\theta}_2} P\left(D \,|\, M_0, \boldsymbol{\theta}_2\right)}$$

This involves numerical optimisation of the likelihood function.

We calculate the p-value.

> p-value: probability of obtaining a test-statistic at least as extreme as the one we saw, if the null hypothesis was true.

Frequentist: how do we do it? ii

The observed p-value is not a continuous measure of our confidence in H_0. The p-value was a means to controlling the type-1 error rate.

It is common nevertheless to interpret p as a measure of our confidence in H_0.

Frequentist: global or local?

If the data had been different, we would have constructed a resonance model with a different mass to match the different data.

We would have looked elsewhere.

Global p-values account for this look-elsewhere effect.

We calculated global p-values with Gross-Vitells [1] and Monte-Carlo simulations.

Bayesian: what is probability?

Probabilities are degrees of belief about any proposition.

There is a unique rule for updating them in light of information — Bayes' theorem.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Bayesian statistics $\Leftrightarrow$ probability theory

Bayesian: what can we do?

We can simply update our belief in the signal + background model relative to the background only model.

The factor that updates our belief is a Bayes factor.

$$\text{Bayes factor} = \frac{\text{Relative belief after data}}{\text{Relative belief before data}}$$

$$= \frac{P\left(D \,|\, M_1\right)}{P\left(D \,|\, M_2\right)}$$

Bayesian: how do we do it?

The numerator and denominator are so-called Bayesian evidences. For a model with parameters $\boldsymbol{\theta}$,

$$P(D \,|\, M) = \int P(D \,|\, M, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \,|\, M)\, d\boldsymbol{\theta}$$

To compare with the p-value, we calculate the posterior of the background model, assuming equal prior odds,

$$P(M_0 \,|\, D) = \frac{1}{1 + B}$$

This is the plausibility of the background model in light of data.

Likelihood function

A component of Bayesian and frequentist analysis. The probability of obtaining data given a particular model and parameters.

Our data is binned. The likelihood is a product of Poissons, one for each bin.

$$P(D|M,\boldsymbol{\theta}) = \prod_i \frac{e^{-\lambda_i}\lambda_i^{o_i}}{o_i!},$$

where the expected number of events depends on the model parameters, $\lambda = \lambda(\boldsymbol{\theta})$.

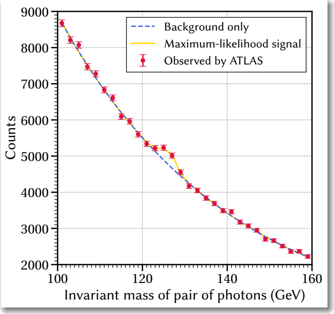
Results from toy Higgs search

From quantum mechanics, we learned an antidote to disputes about interpretations.

Shut up and calculate.

Toy problem

To make calculations, let's pick a toy problem to study. The search for the Higgs in the diphoton channel by ATLAS with 25/fb [2].

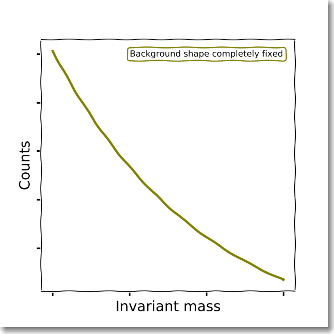


An important search for the discovery of the Higgs.

Background model

There is a monotonically falling background.

We could describe it by a basis of polynomials (e.g. Bernstein) but so that we can perform many calculations, we just use a fixed background and neglect parametric uncertainties in it.



Signal model i

We model the signal predicted by a Higgs as a Gaussian centred at m_h.

The width was the experimental resolution of about 1.5 GeV.

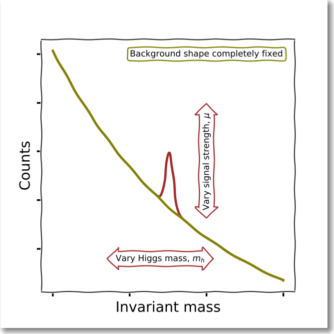
We specified the strength relative to the Standard Model prediction (at 125 GeV),

$$\mu = \frac{\text{efficiency} \times \text{cross section}}{(\text{efficiency} \times \text{cross section})_{\text{SM @ 125 GeV}}}$$

This is an approximation as we did not model dependence of efficiency or cross section as functions of Higgs mass.

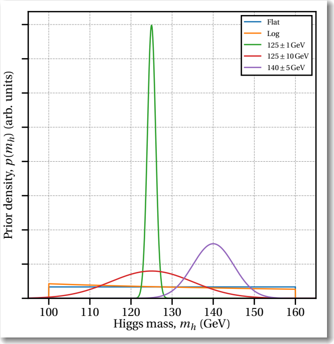
Signal model ii

There were thus two unknown parameters describing the location and strength of the resonance, m_h and μ.



Priors

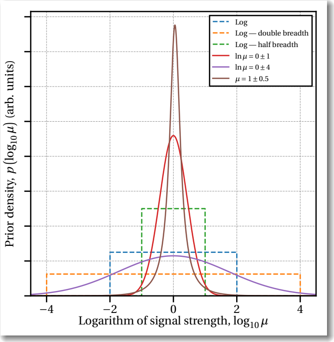
For our Bayesian calculations, we must place priors on m_h and μ. We experiment with several choices.



Broad priors (log and flat) and narrow ones representing specific prior knowledge.

Priors

For our Bayesian calculations, we must place priors on m_h and μ. We experiment with several choices.



We vary the breadth of the log prior for the signal strength, and the shape of the prior.

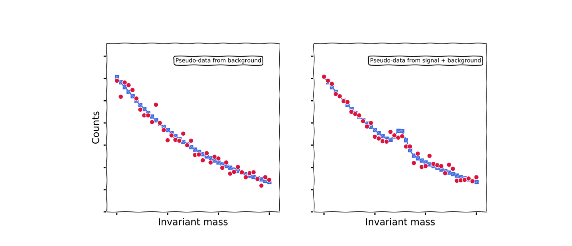
Data and pseudo-data

We use the real 25/fb collected by ATLAS [2].

We sample our own pseudo-data from the background model and the signal + background model with $\mu = 1$, $m_h = 125\,\text{GeV}$.

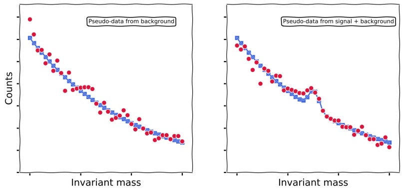
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



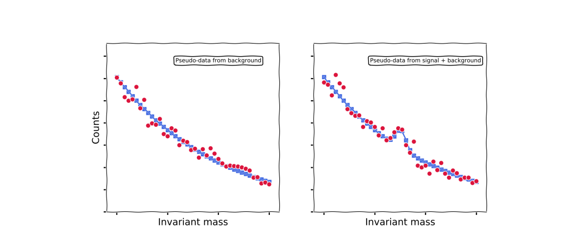
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



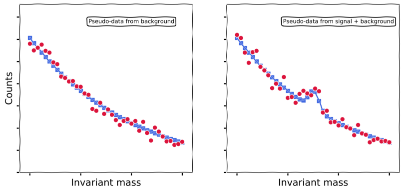
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



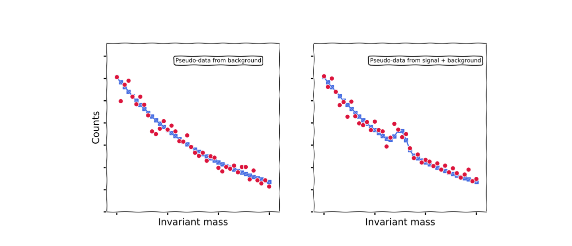
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



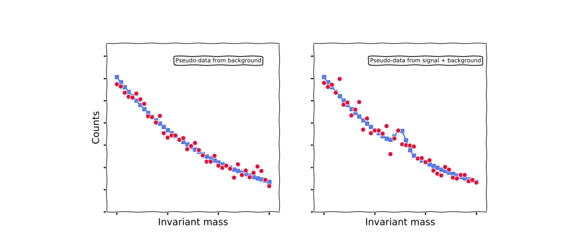
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



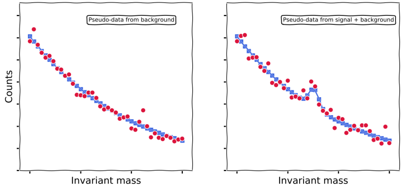
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



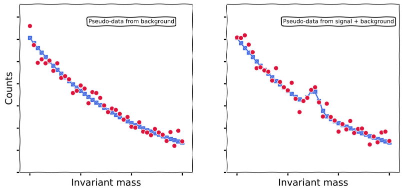
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



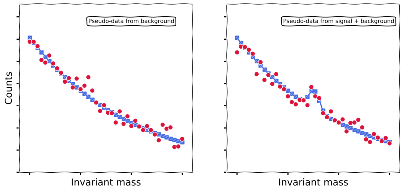
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



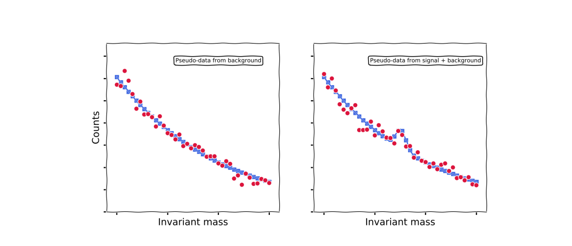
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



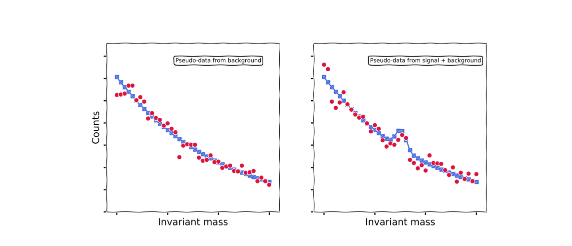
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.



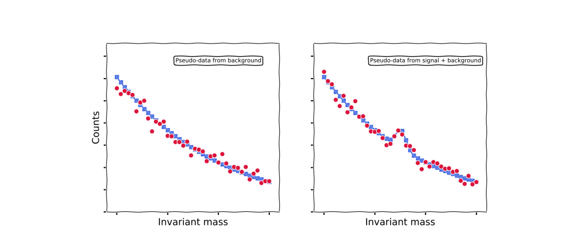
The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.

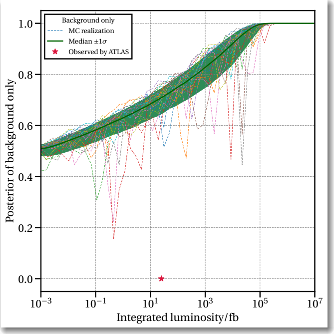


The models tell us the expected number of counts in each bin for a particular integrated luminosity.

We sample pseudo-data at many integrated luminosities by drawing counts from Poisson distributions in each bin.

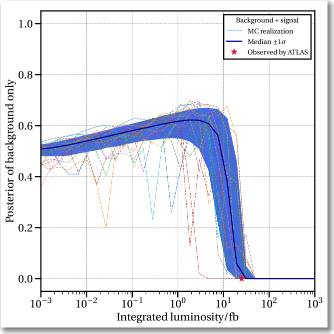


Evolution of p-value and posterior as we collect data



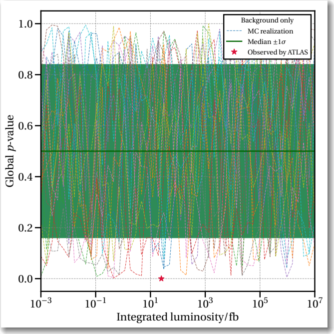
The posterior slowly approaches 1 when the background model is correct

Evolution of p-value and posterior as we collect data



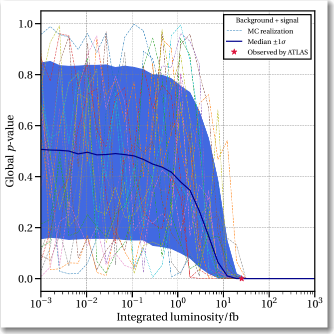
and zero when the signal model is correct, though in this case there is an extremely mild preference for the background model until about 10/fb.

Evolution of p-value and posterior as we collect data

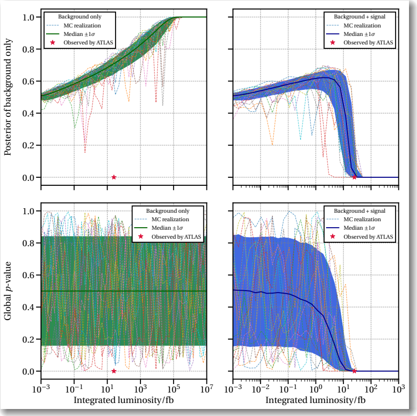


The p-value makes a random walk between 0 and 1 when the background model is correct

Evolution of p-value and posterior as we collect data



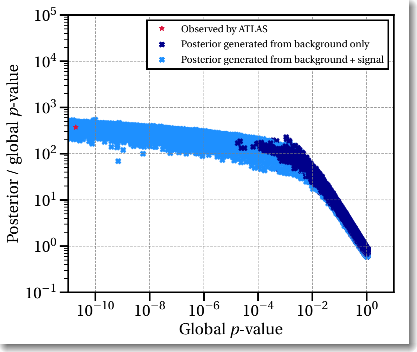
and when the signal model is correct, it makes a (noisy) walk towards zero.



Bayesian (top)/frequentist (bottom). Background model true (left)/signal model true (right).

Comparison between p-value and posterior

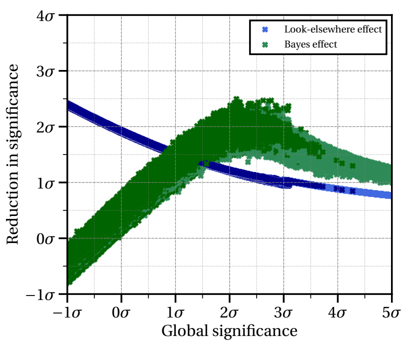
We performed about a million pseudo-experiments.



The posterior of the background model about 10^2 – 10^3 times greater than global p-value!

The Bayes effect

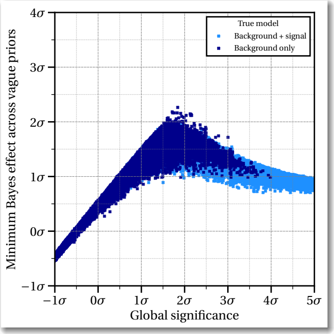
The magnitude of the effect greater than the well-known look-elsewhere effect.



Global significances reduced by $1 - 2\sigma$.

Prior dependence

We checked many priors. The effect could be reduced but remained important.



See paper [3] for full discussion about prior dependence of this effect.

Conclusions

1. First detailed comparison of Bayesian and frequentist methods in resonance searches
2. Posterior ultimately converged to 0 or 1; p-value makes random walk if H_0 correct
3. p-values overstate evidence against the null! p-value $\lll$ posterior of background model
4. Checked that the effect was robust with respect to several choices of prior
5. When looking at an anomaly, we must remember the look-elsewhere effect and the Bayes effect

[1] E. Gross and O. Vitells, “Trial factors for the look elsewhere effect in high energy physics,” Eur. Phys. J. **C70**, 525–530 (2010), arXiv:1005.1891.

[2] G. Aad et al., “Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC,” Phys. Lett. **B726**, [Erratum: Phys. Lett.B734,406(2014)], 88–119 (2013), arXiv:1307.1427.

[3] A. Fowlie, “Bayesian and frequentist approaches to resonance searches,” (2019), arXiv:1902.03243.