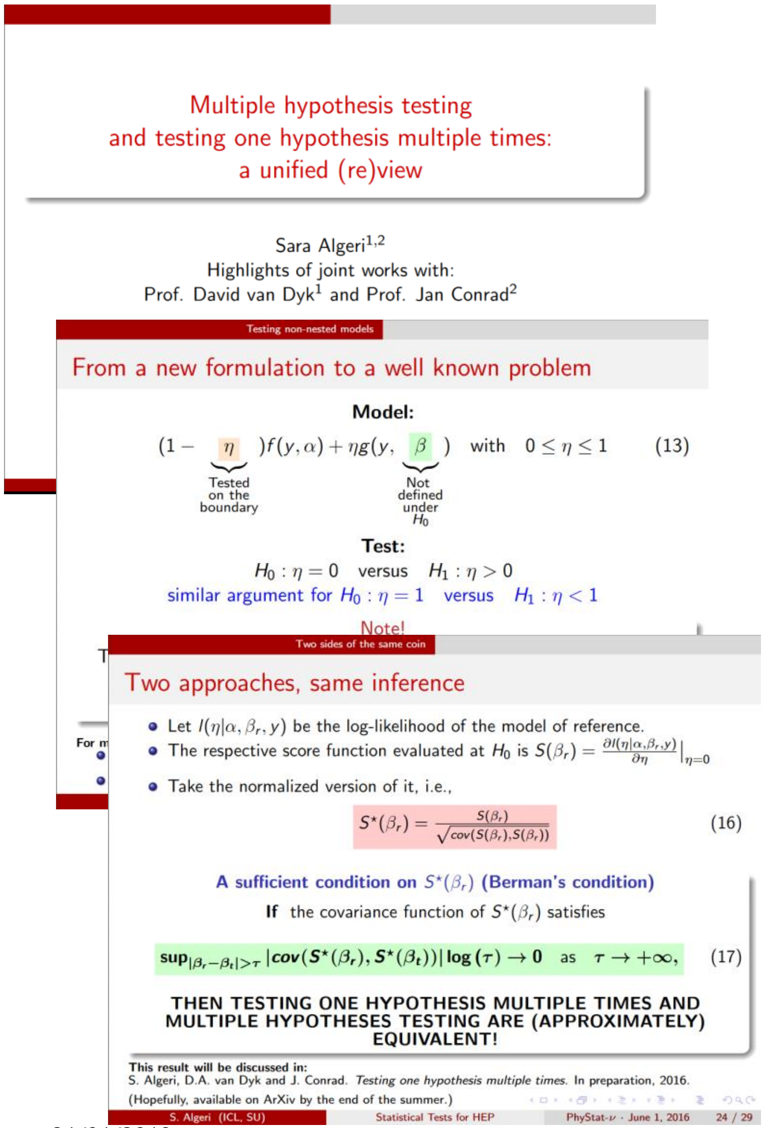
Look-elsewhere effect in neutrino oscillation searches:

Phill Litchfield, Abbey Waldron

15/01/2019

Introduction



Phystat-ν [Kashiwa] talk by S. Algeri

- How to frame discrete models (e.g. mass ordering) as a Look-Elsewhere Effect (LEE).

Followed up 2017

- Looks like we could use this approach for T2K MO [ask me over coffee break]
- Wanted to first understand this approach to LEE
- But no time to actually work on it

2018 out of the blue, Abbey contacted to ask if I had any interesting neutrino / computing projects to work on

- **Why, yes!** Yes I do.

LEE in discrete tests

If I observe a 3σ deviation from my Null Hypothesis, this means:

If	**the Null Hypothesis is true**
Then	**the probability of this occurring by chance is 0.27%**

But if I looked at 100 different data sets, how surprised should I be?

If the probability of at least one occurrence in τ trials is P_τ then:

$$P_{100} = 1 - \overline{P_{100}} = 1 - (\overline{P_1})^{100} = 1 - (1 - P_1)^{100}$$
$$\simeq 100 \times P_1$$

Z_1	Z_{100}
2σ	0.01σ
$\mathbf{3\sigma}$	$\mathbf{1.2\sigma}$
4σ	2.7σ
5σ	4.0σ

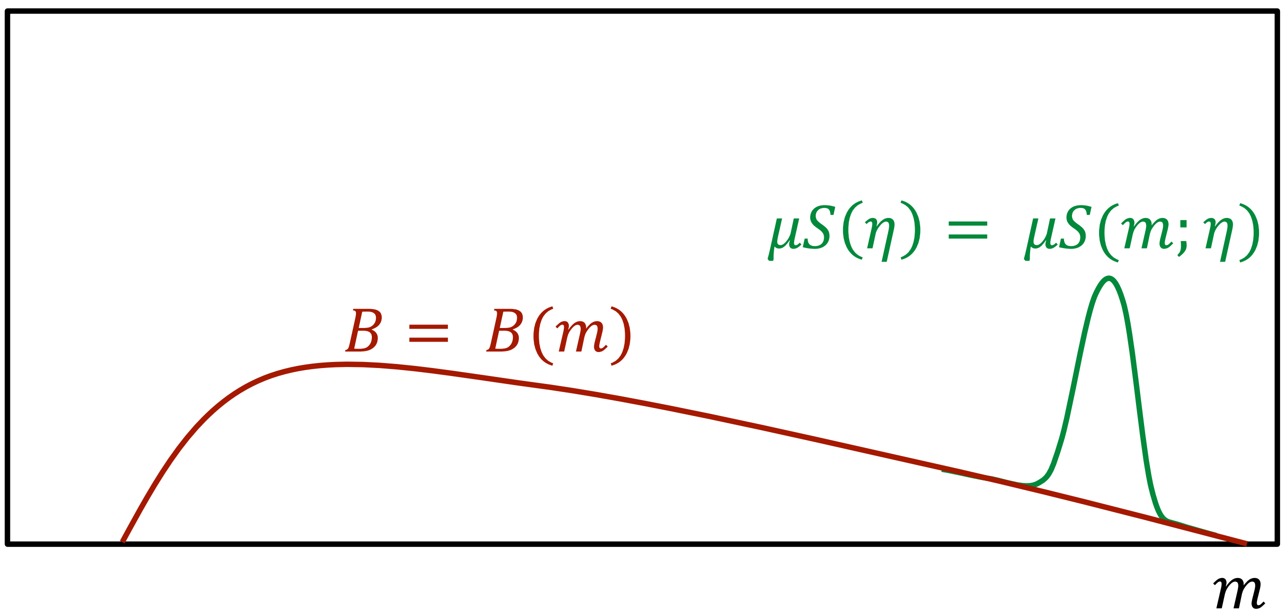
For this example:

a result with 3σ *local* significance
becomes 1.2σ *global* significance

LEE in continuous tests

Imagine a collider experiment looking for a resonance.

- There is a known **background**
- There might be a **signal** resonance *somewhere* in the search range



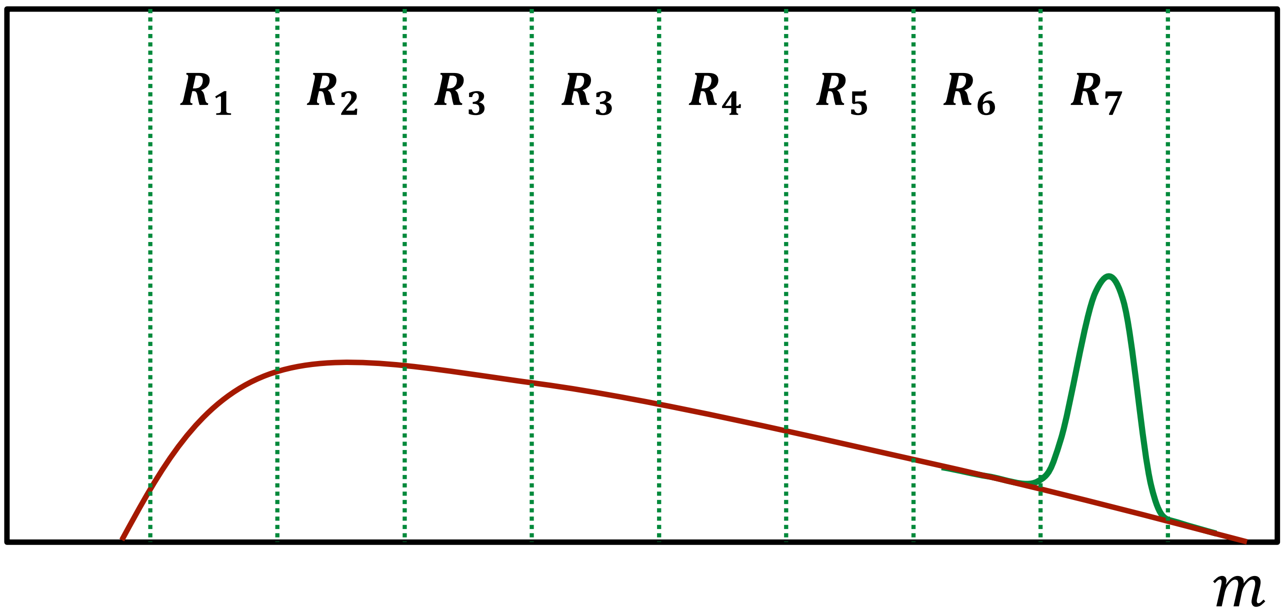
$\Rightarrow$ There is a Look-Elsewhere Effect here as well.

But how to quantify it?

The "Thumb rule"

Quasi-discrete approach: break the search range up into sub-ranges

- Wide enough that ~1 resonance can exist within each one…
- Then as in the discrete case $P_\tau = \tau \times P_1$



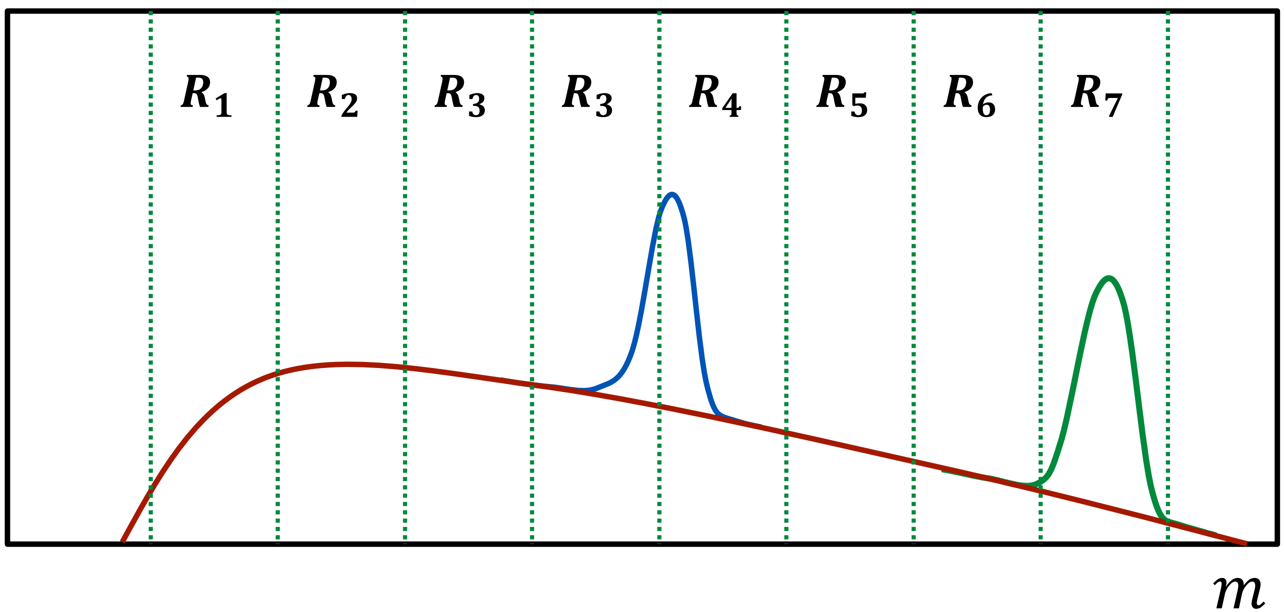
This **assumes** we are searching at only 7 specific values of η

- Still better than ignoring the issue…

The "Thumb rule"

Quasi-discrete approach: break the search range up into sub-ranges

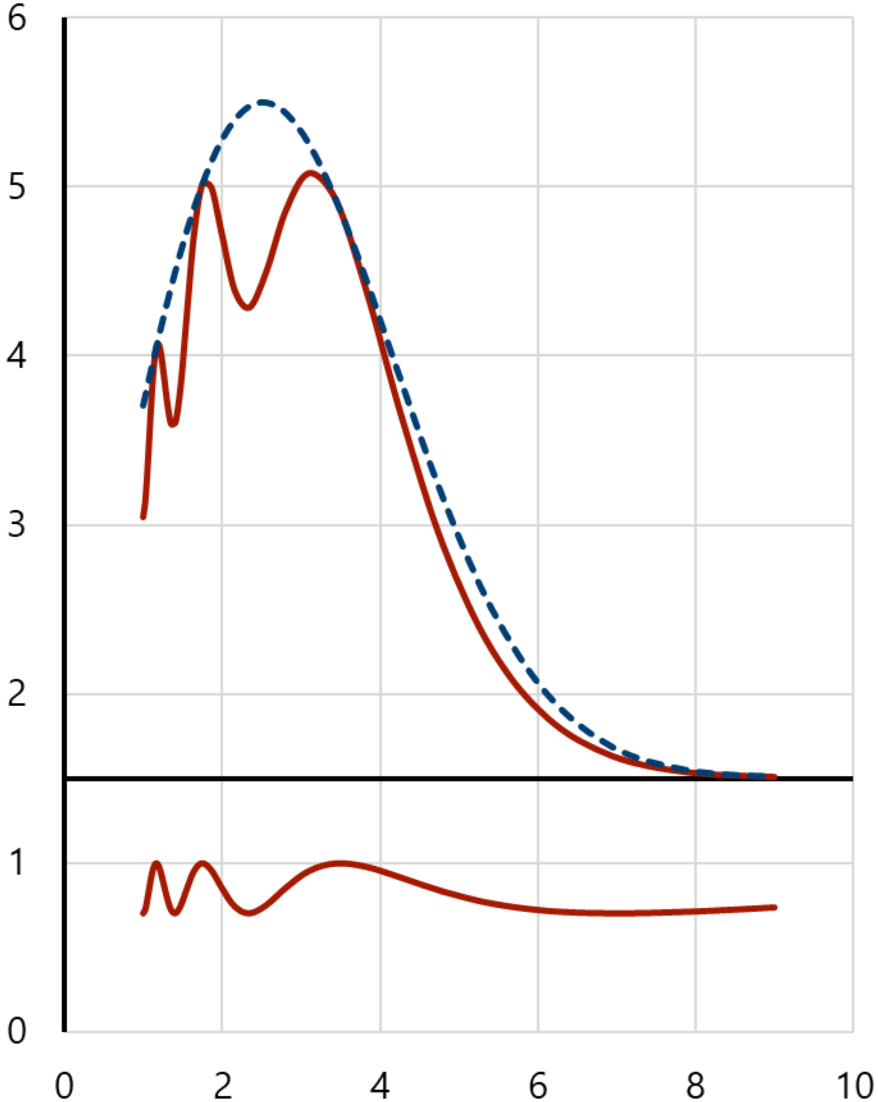
- Wide enough that ~1 resonance can exist within each one…
- Then as in the discrete case $P_\tau = \tau \times P_1$



This **assumes** we are searching at only 7 specific values of η

- Still better than ignoring the issue… **but not very realistic**

Neutrino oscillation searches



In searches for new oscillation scales we have:

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} \pm \sin^2 2\theta \sin^2 \frac{\Delta m^2 L}{4E}$$

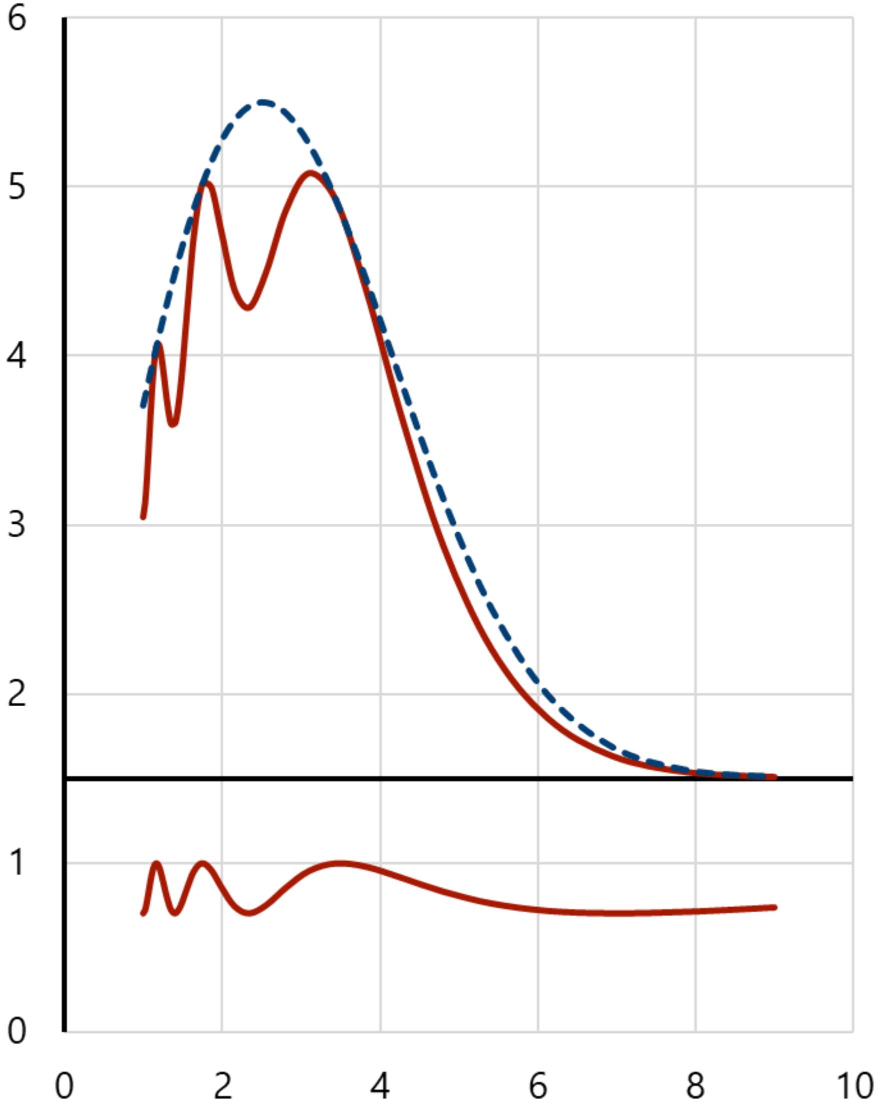
Similar to bump searches:

- Location parameter (Δm^2)
- Magnitude ($\sin^2 2\theta$)

But there is a difference:

- The signal is not localised.
- So, how many searches?

How to handle this?



In searches for new oscillation scales we have:

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} \pm \sin^2 2\theta \sin^2 \frac{\Delta m^2 L}{4E}$$

Similar to bump searches:

- Location parameter (Δm^2)
- Magnitude ($\sin^2 2\theta$)

But there is a difference:

- The signal is not localised.
- So, how many searches?

In the neutrino case:

- There's no approach equivalent to dividing the spectrum into sub-ranges.
- But the concept of a tuneable search parameter still exists —

how can we use that?

Resonance search by scanning the location parameter

Looking for a 'bump' on top of a (known) **background B**

- The bump is a localised feature, parameterised by its **location (η)**, and **magnitude (μ)**

If we already knew to search at η_0 : standard results (Wilks, Chernoff) for significance, based on log-**likelihood ratio $q(\hat{\mu}, \eta_0)$**

But if the search location $\hat{\eta}$ is determined by fitting data, these results will overestimate the significance:

- By definition $q(\hat{\mu}, \hat{\eta}) = \max_{\eta}\{q(\hat{\mu}(\eta), \eta)\} \geq \underbrace{q(\hat{\mu}(\eta_0), \eta_0)}$

$\rightarrow \chi^2$ distribution

Physicists: ***"Look-Elsewhere effect"***
Statisticians: ***"η is only present under the alternative hypothesis"***

Davies bound

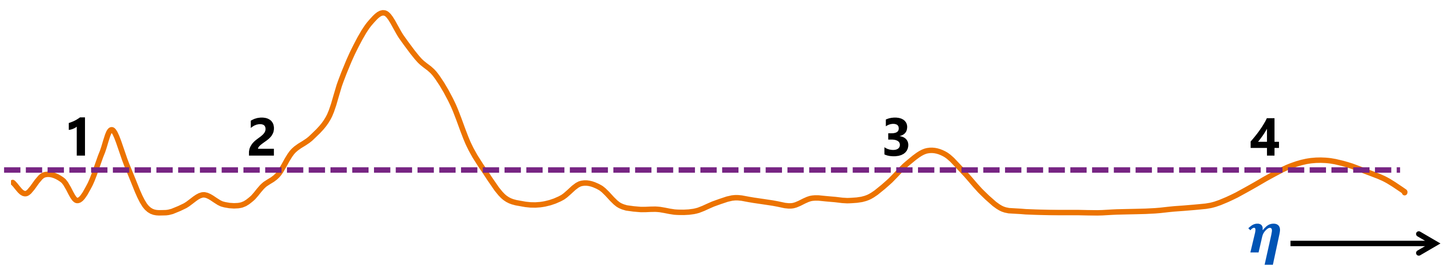
Because of the LEE:

$$P(q(\hat{\eta}) > c) > P(q_{\text{fix}} > c)$$

Davies [1977 & 1987] showed that:

$$P(q(\hat{\eta}) > c) < P(q_{\text{fix}} > c) + \langle N(c) \rangle$$

Where $\langle \boldsymbol{N}(\boldsymbol{c}) \rangle$ is the *expected* number of times $\boldsymbol{q}$ goes above the level $\boldsymbol{c}$ when scanned across $\boldsymbol{\eta}$



This is a useful result! Although it is still not exact, it bounds the significance from the other (conservative) side

Gross & Vitells extension

For a χ^2 test: $P(q_{\text{fix}} > c) = P(\chi_s^2 > c)$ and:

Davies 1987

$$\langle N(c)\rangle = \left[\frac{c^{(s-1)}e^{-c}}{\pi\, 2^s}\right]^{1/2} \frac{1}{\Gamma((s+1)/2)} \int_L^U I(\eta)\, \mathrm{d}\eta$$

Gross & Vitells [2010] point out that

1. The hard part (the integral) is independent of the threshold.
2. The expectation $\langle N(c)\rangle$ can be calculated numerically at some low threshold value (c_0) and evolved to the level of interest (c)

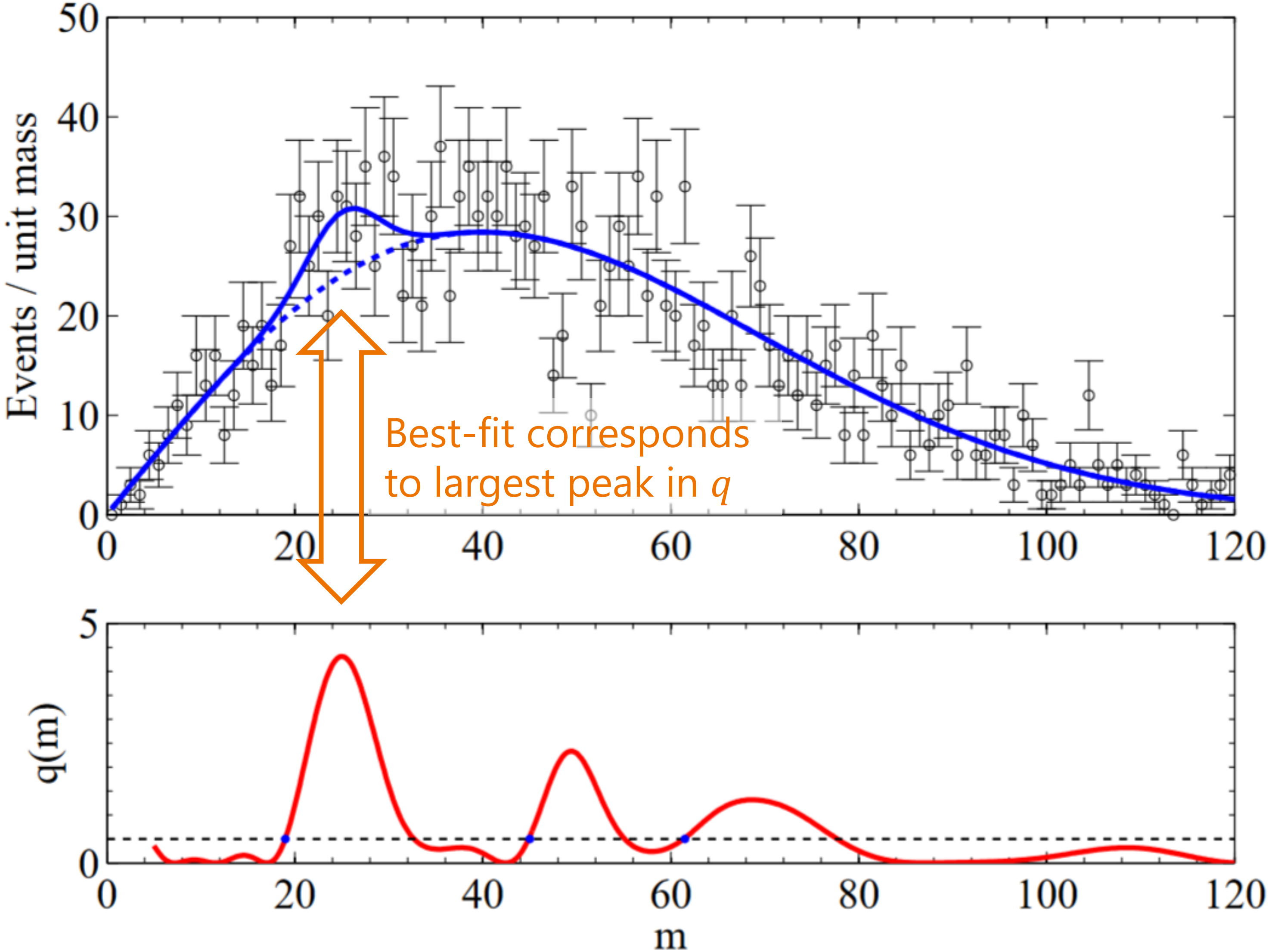
Thus:

$$P(q(\hat{\eta}) > c) < P(\chi_s^2 > c) + \underbrace{\langle N(c_0)\rangle}_{\text{Evaluate with MC}}\, \mathrm{e}^{-(c-c_0)/2} \underbrace{\left(\frac{c}{c_0}\right)^{(s-1)/2}}_{\text{s=1 here} \to \text{ignore}}$$

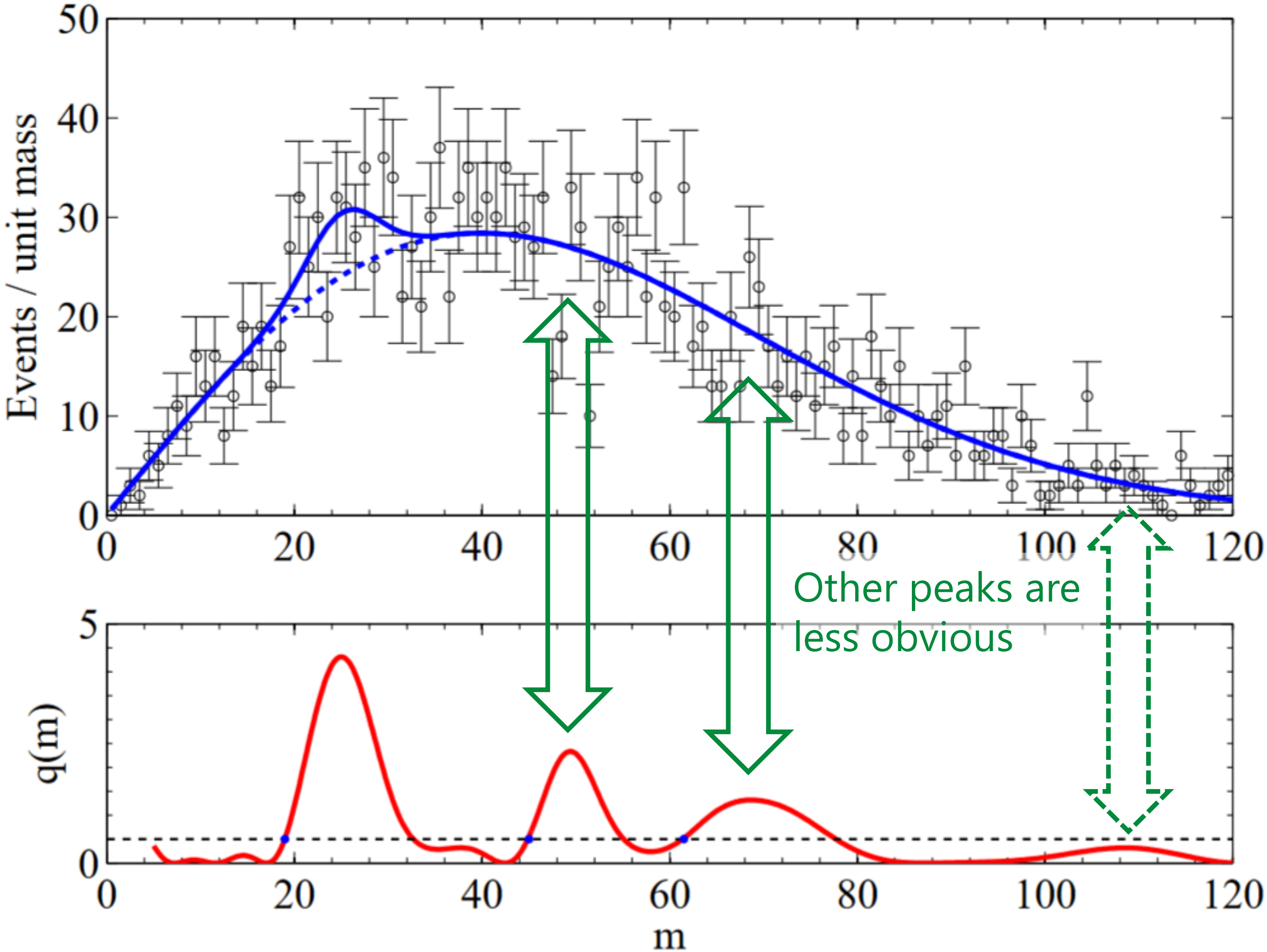
Evaluate with MC

s=1 here → ignore

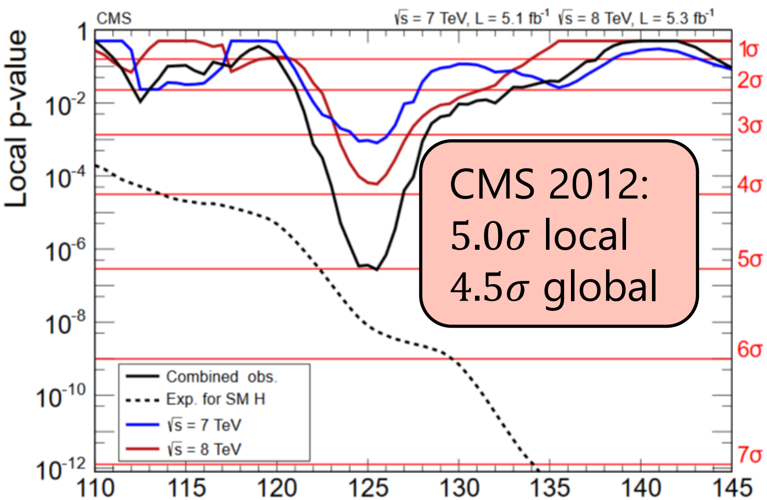
Gross–Vitells example

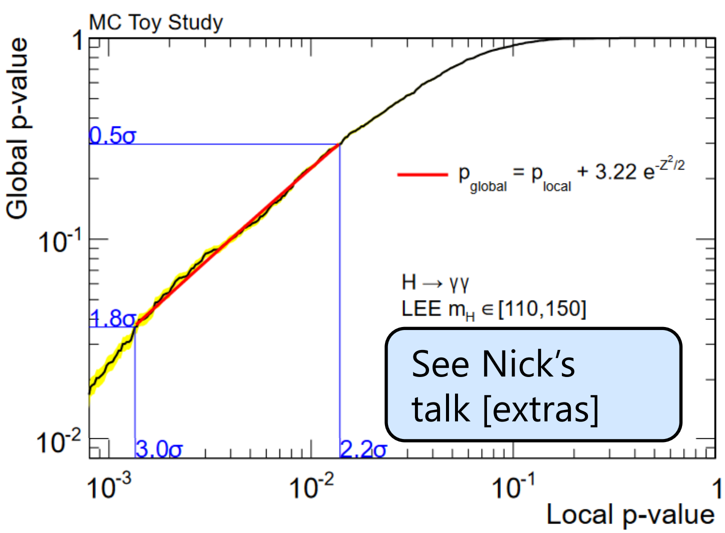


Gross–Vitells example



Use in LHC searches





Most LHC resonance searches now use this approach. Written in physicists jargon:

$$P(q(\hat{\eta}) > c) \rightarrow p_{\mathrm{global}}$$
$$P(\chi^2_s > c) \rightarrow p_{\mathrm{local}}$$
$$c = z^2 \text{ (where } z \text{ is the significance)}$$
$$s = N_{\mathrm{dof}} = 1$$

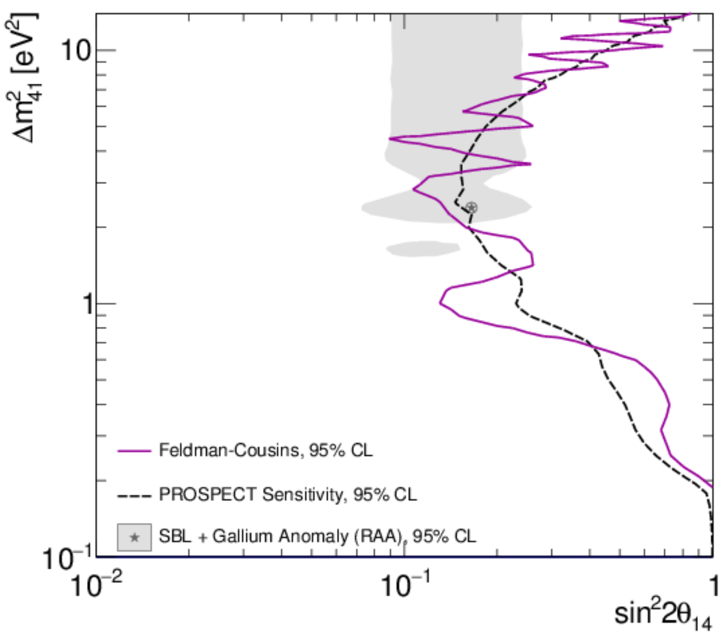
We then have:

$$p_{\mathrm{global}} = p_{\mathrm{local}} + k\, \mathrm{e}^{-z^2/2}$$

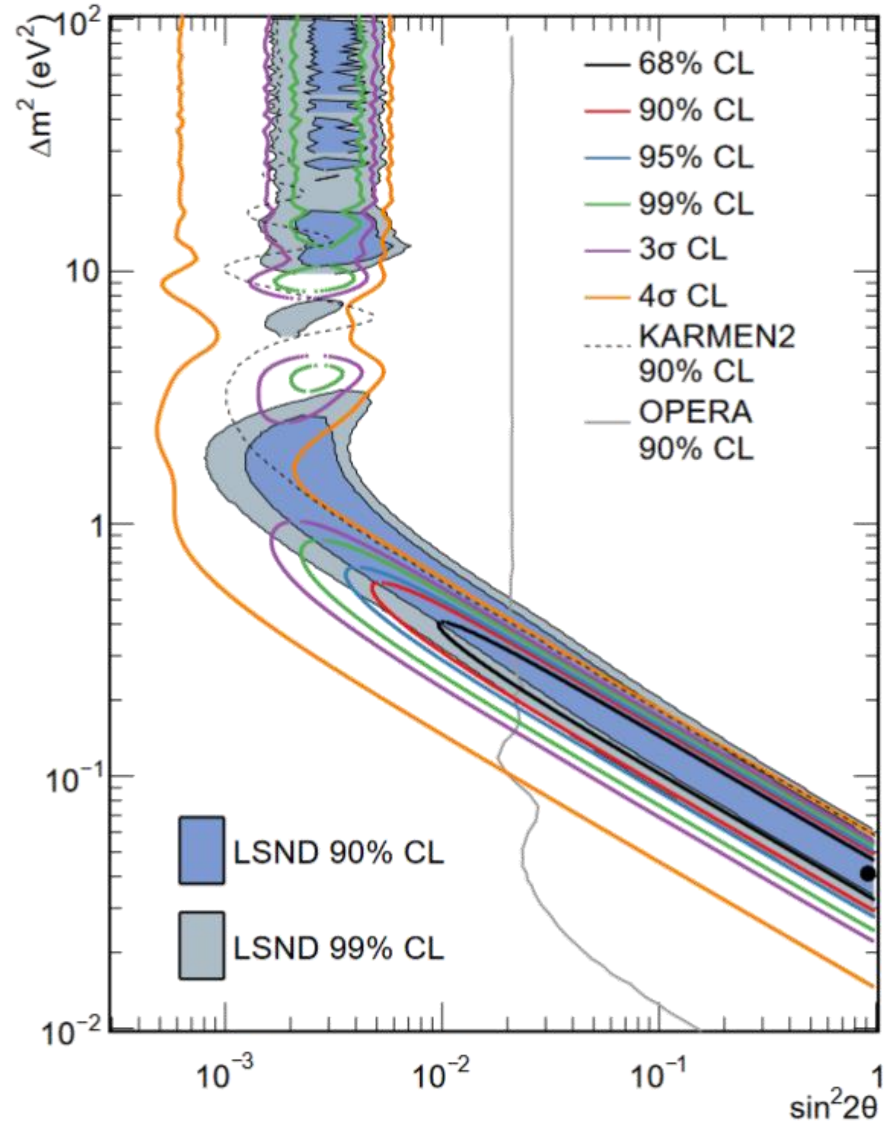
and this z dependency is used to extrapolate from lower-significance MC toys

Sterile neutrino searches

Looking at this approach for neutrino oscillation searches.



Disappearance (◀) or appearance (▶) channels



LSND/MiniBooNE

Questions to investigate

High energy regime

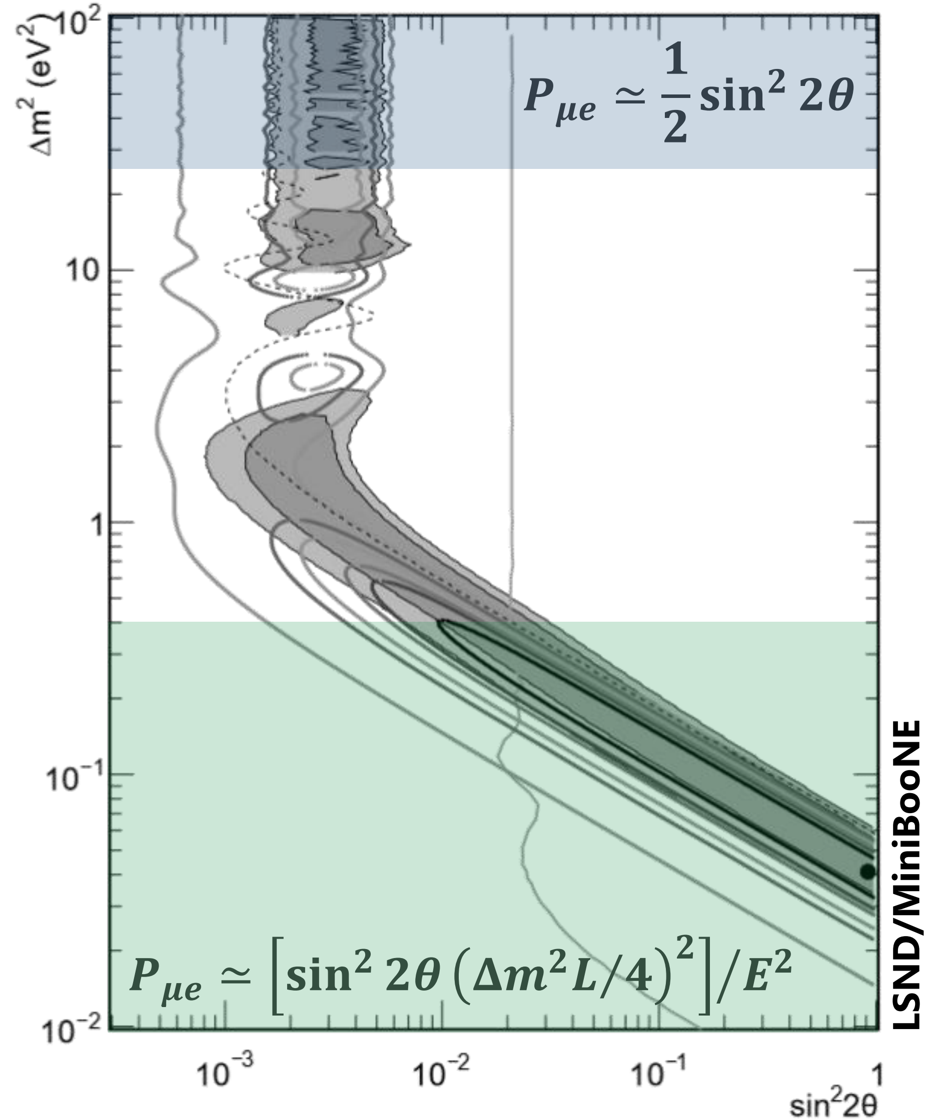
- Finite energy resolution washes out location parameter.

Low energy regime

- Location parameter becomes degenerate with the strength parameter

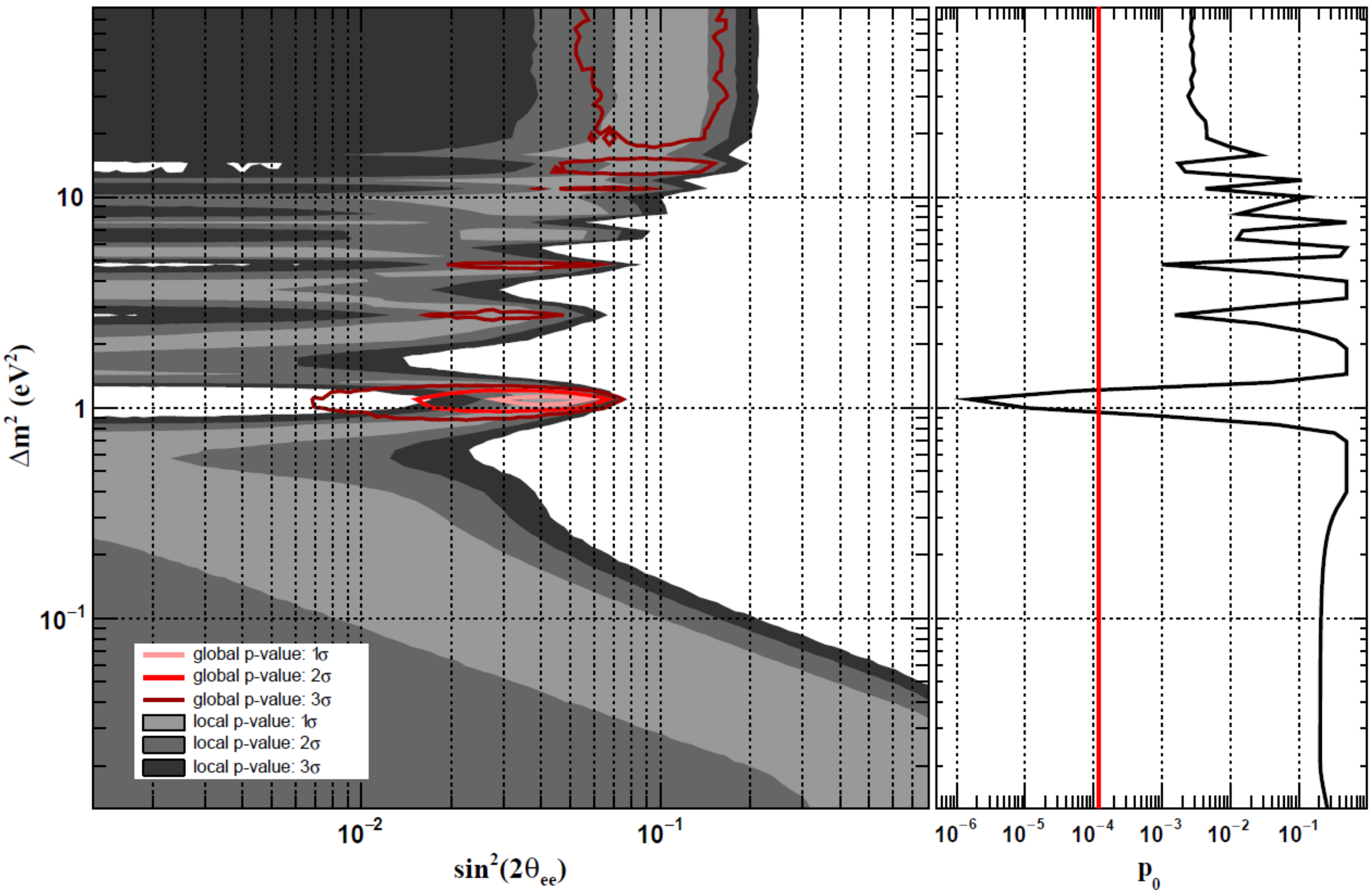
In both cases we transition back to a single degree-of-freedom;

- How will this affect the LEE correction?



-- Stop Press --

Poster “Statistical methods and issues in sterile neutrino searches“ by B.Neumair → Covers similar ideas & connection back to F&C

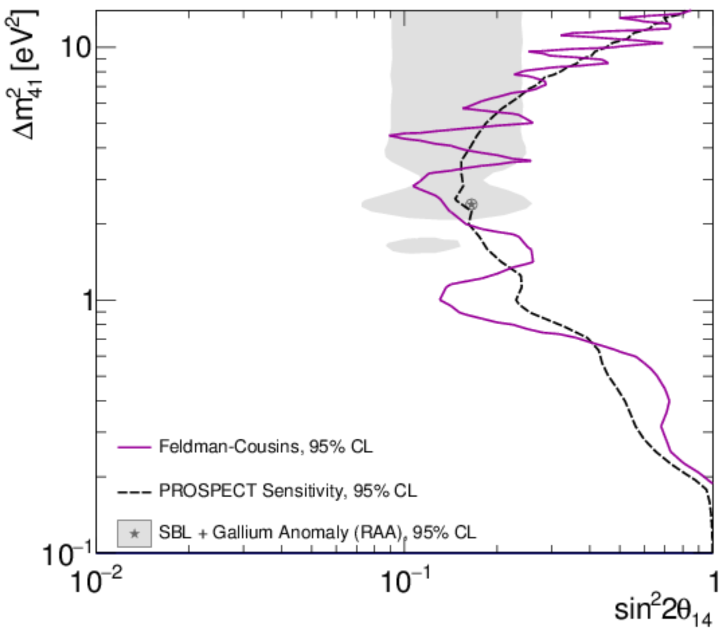


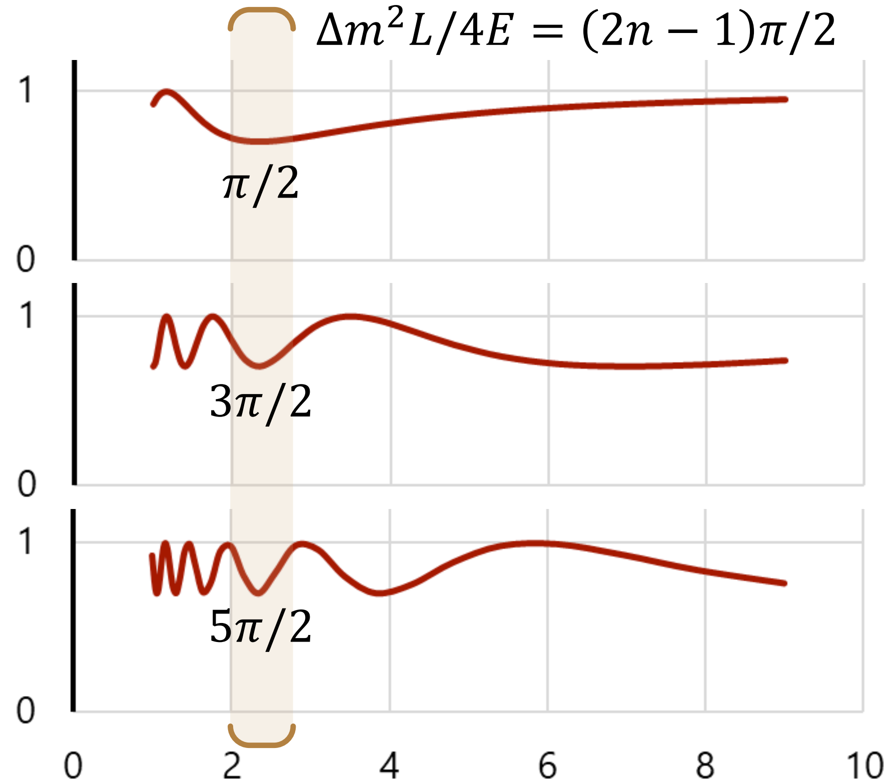
More questions to investigate

Imperial College London

Behaviour in **high-energy regime** also depends on how "normalisation" is handled.

- Dis/appearance not the (direct) cause





In the **intermediate energy regime**, possible correlation between harmonics?

Application to neutrinos

Toy study:

- Point like reactor and detector with 30m baseline

  [Reality: both reactors & detectors extended over ~m]
- Flux based on RENO 2018 data release
- Fixed 0.2 MeV energy resolution.

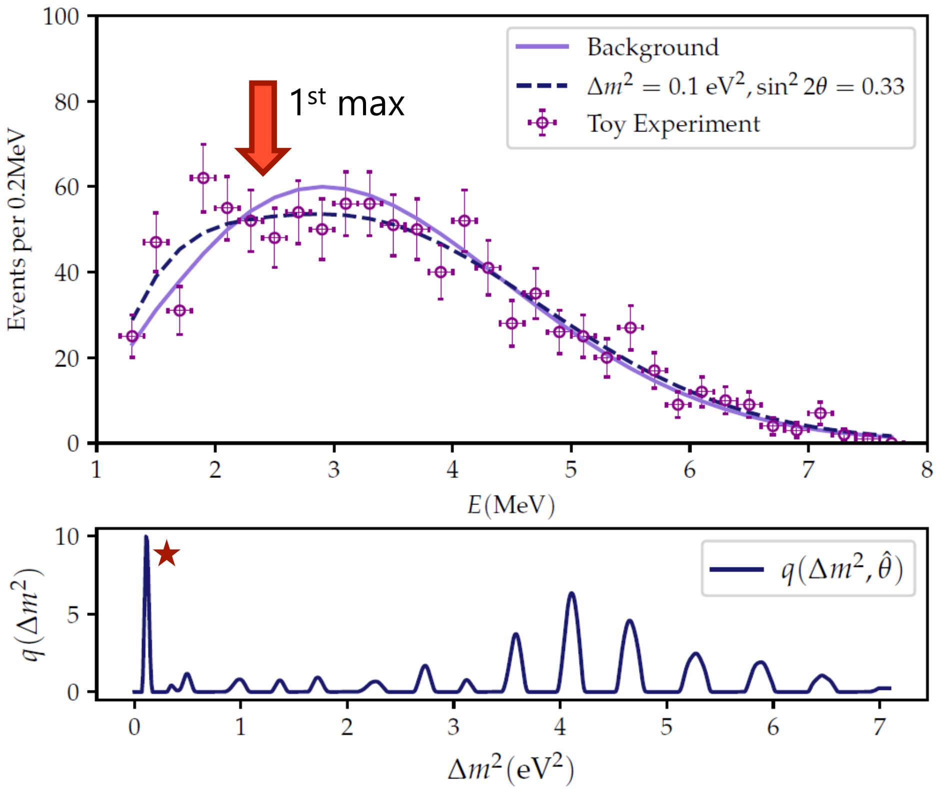
  [Typical: ~10% resolution; so worse >4MeV]

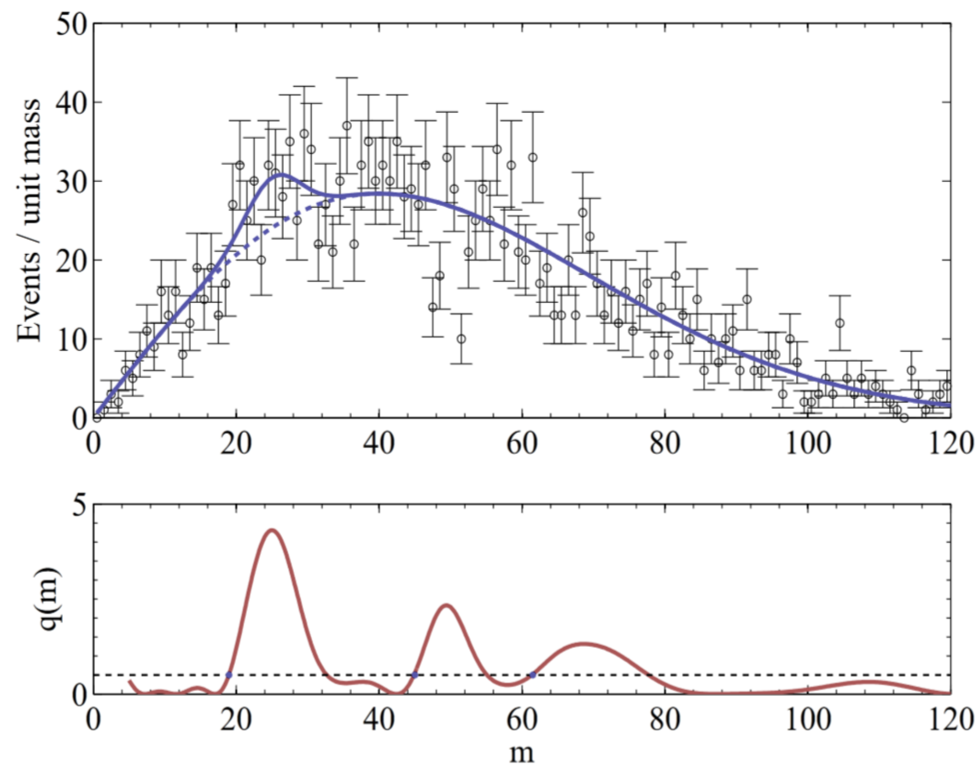
- Free normalisation (i.e. shape only-analysis)
  - Fixed or constrained normalisations have different large Δm^2 behaviour
  - Multi-baseline experiments effectively do shape-only analyses, but several (L,E) schemes in use: ratios, averaging over baselines, …

Application to neutrinos

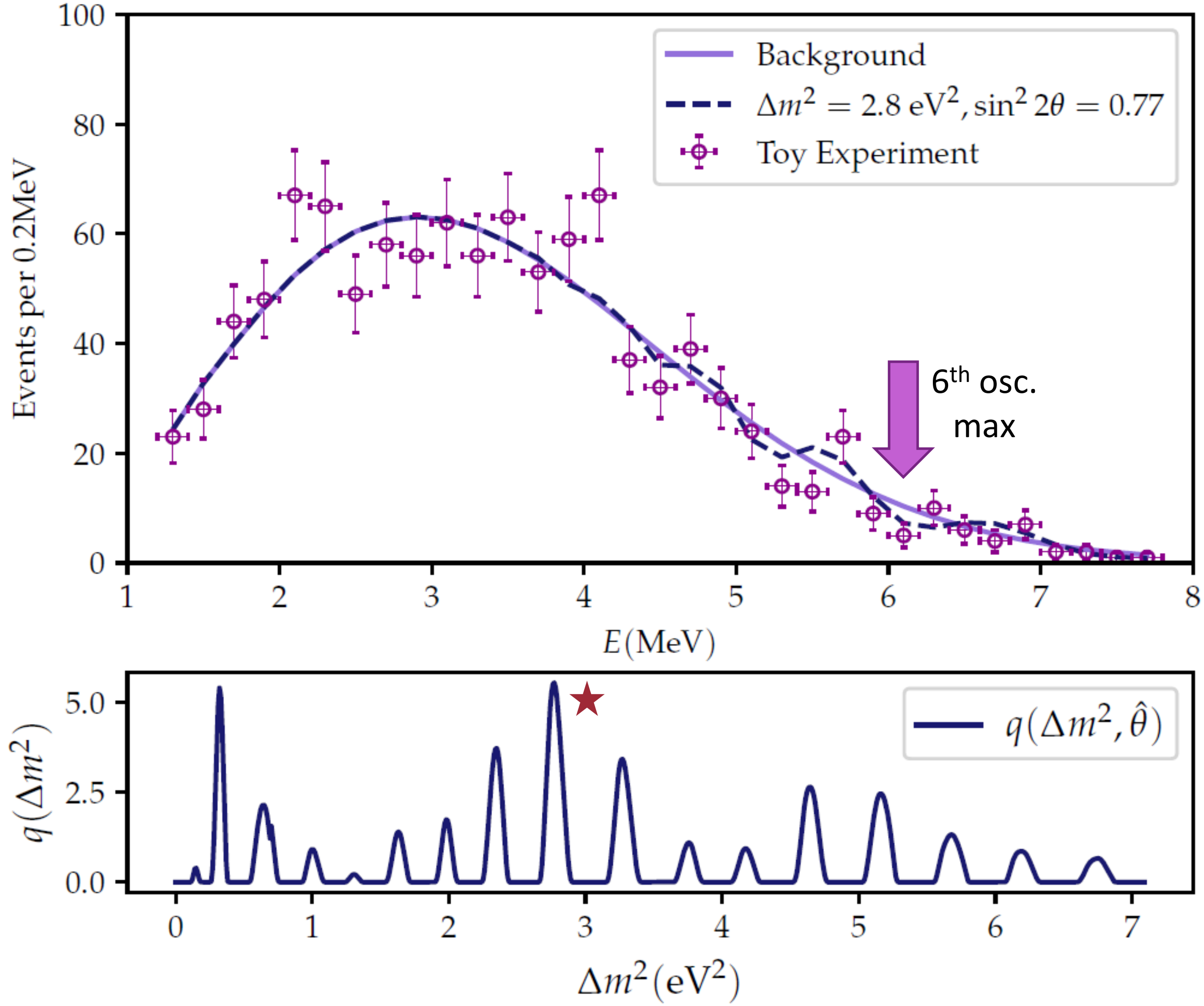
Compared to G-V toy study (right), notice:

- No visual mapping to Δm^2
- Much higher dynamic range → Better resolution → Sensitive to 'harmonic' solutions





More examples



High Δm^2 fits have tail structure – improve E resolution model.

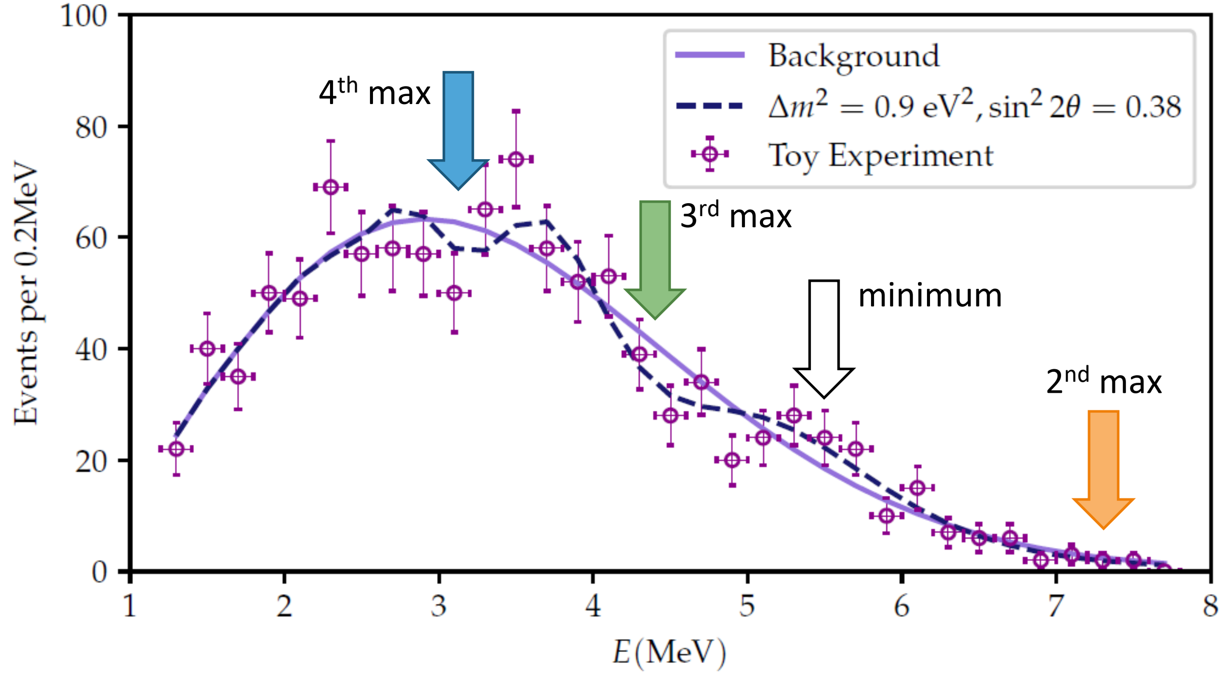
More examples

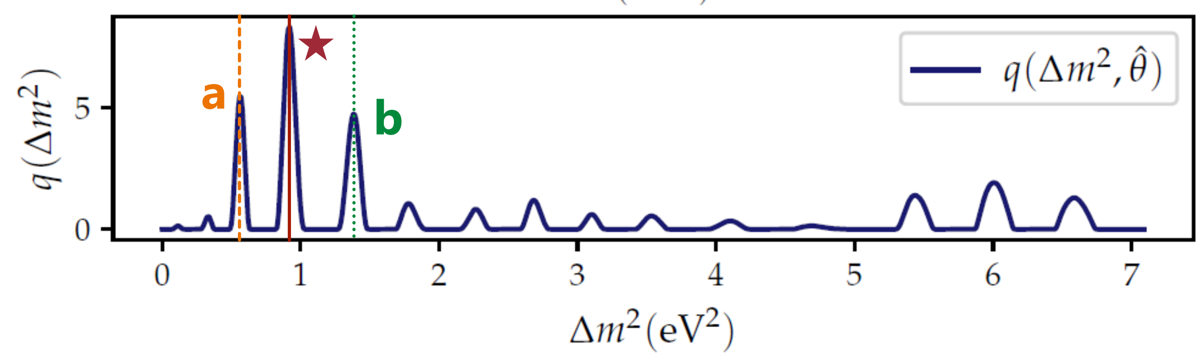
Adjacent harmonic aliases will be at $(n/n+2)\Delta m^2$

Dips ⇒ odd-n
Peaks ⇒ even-n

★/a ≃ 5/3 fits the 4.4 MeV dip.

★/b ≃ 2/3 would map a peak to a dip, so this is not an alias.





Summary

Sterile neutrino searches should account for the Look-Elsewhere Effect

- Below $\sim 2.5\sigma$, can just do B/G-only toys.
- But not so easy for 'interesting' results
- Note: Better resolution means local value is *more* wrong

Davies / Gross & Vitells method looks usable for estimating global significance for reactor searches

- But not trivial to investigate
- Normalisation, harmonic and small$-\Delta m^2$ degeneracies need to be checked & understood, but seem okay so far.

Should be equally viable for (e.g.) SBN appearance at FNAL

Extensions & references

Davis '87 result covers χ^2-like statistics. If instead use *amplitude* of oscillation as a test statistic, can use similar relationship from Davis '77

Tests of 3+*n* models simply extend this to χ^2 with *n* d.o.f. Generalised form already exists.

Also want to get back to applying this to MO…

Gross & Vitells: https://inspirehep.net/record/854732

Davies '77: Biometrika,64,247-254

Davies '87: https://inspirehep.net/record/854290

RENO flux: https://inspirehep.net/record/1676077

Imperial College London

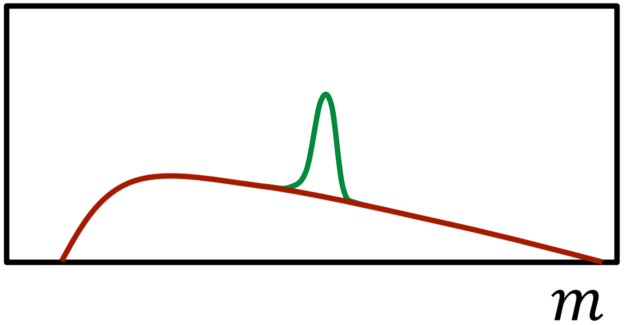
Extra slides

20/01/2019

Letting η be continuous

Could test only at wide-spaced values of η, but this is very limiting.

Instead, consider a test statistic $T(\mu, \eta)$.



Most commonly: $$T = -2 \ln \frac{\mathcal{L}(B)}{\mathcal{L}(\mu S(\eta) + B)}$$

For an *assumed* value η_0

we can find the best fit value $\hat{\mu}(\eta_0)$ that maximises T:

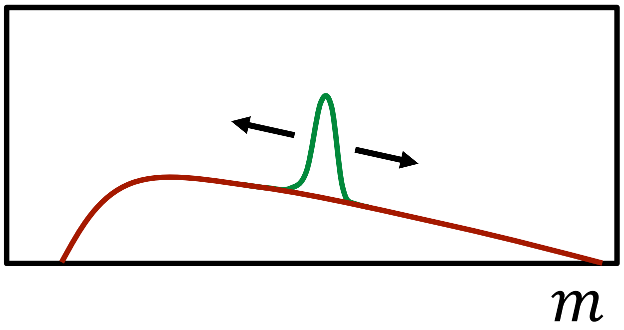
✓ Obeys Wilks' theorem → significance can be estimated from χ^2

✘ Significance only meaningful if the assumption $\eta = \eta_0$ is true.

Fitting for η

Could test only at wide-spaced values of η, but this is very limiting.

Instead, consider a test statistic $T(\mu, \eta)$.



Most commonly: $$T = -2\ln\frac{\mathcal{L}(B)}{\mathcal{L}(\mu S(\eta) + B)}$$

Instead of assuming η_0

we can find the 2D best fit value $(\hat{\mu}, \hat{\eta})$. But by definition:

$$T(\hat{\mu}, \hat{\eta}) = \max_{\eta}\{T(\hat{\mu}(\eta), \eta)\} \geq T(\hat{\mu}(\eta_0), \eta_0)$$

The RHS obeys Wilks' theorem, so the LHS must not!

Trial factors

In the discrete case we had that $P_\tau \simeq \tau P_1$, (provided $P_1 \ll 1$)

The factor τ is generalised to a non-integer **trial factor**:

$$\tau = \frac{P(T(\hat{\eta}) > c)}{P(T(\eta_0) > c)} = \frac{P(T(\hat{\eta}) > c)}{P(\chi_s^2 > c)} \underset{s=1}{\Longrightarrow} \frac{P(T(\hat{\eta}) > c)}{\sqrt{c}}$$

With which one can convert a "local" significance to a "global" one.

- Or if you prefer, an **effective number of search regions**

The expression derived for this is:

$$\tau_{s=1} \simeq 1 + \frac{\sqrt{c}}{2} \int_L^U I(\eta)\, \mathrm{d}\eta$$

[The integral is deducible from the small threshold $\langle N(c_0) \rangle$]

Use in matter effect

Blennow M. et al., JHEP, 2014 - Assumptions

...BUT

To make it work we need:

- The $x_1, \dots, x_B$ are independent and approximately Gaussian. (If not independent, the variance is no longer as simple as $4T_0$.)
- We need assumptions similar to those required by Wilks' theorem.

PLUS one (or more) of the following:

1. There are no free parameters α and β.
2. OR The hyperplanes of the two hypotheses at their minima are parallel, i.e.,

$$\left.\frac{\partial f(y_i,\alpha)}{\partial \alpha}\right|_{\alpha=\alpha_{\min}} = \left.\frac{\partial g(y_i,\beta)}{\partial \beta}\right|_{\beta_{\min}}.$$

This is very restrictive!

E.g.: Let $f(\alpha)=\alpha$, $\alpha \in [0;10]$, $g(\beta)=\beta^2$, $\beta \in [15;20]$ and $\beta_{\min}=1$

$$1 = \left.\frac{\partial f(\alpha)}{\partial \alpha}\right|_{\forall\alpha\in[0;10]} \neq \left.\frac{\partial g(\beta)}{\partial \beta}\right|_{\beta=1} = 2 \quad \Rightarrow \textbf{NOT OK.}$$

3. OR if T_0 much greater than the number of parameters in the hypothesis.

This condition Is broken?

$$\left.\frac{dN}{d\delta_{CP}}\right|_{NO} \neq \left.\frac{dN}{d\delta_{CP}}\right|_{IO}$$