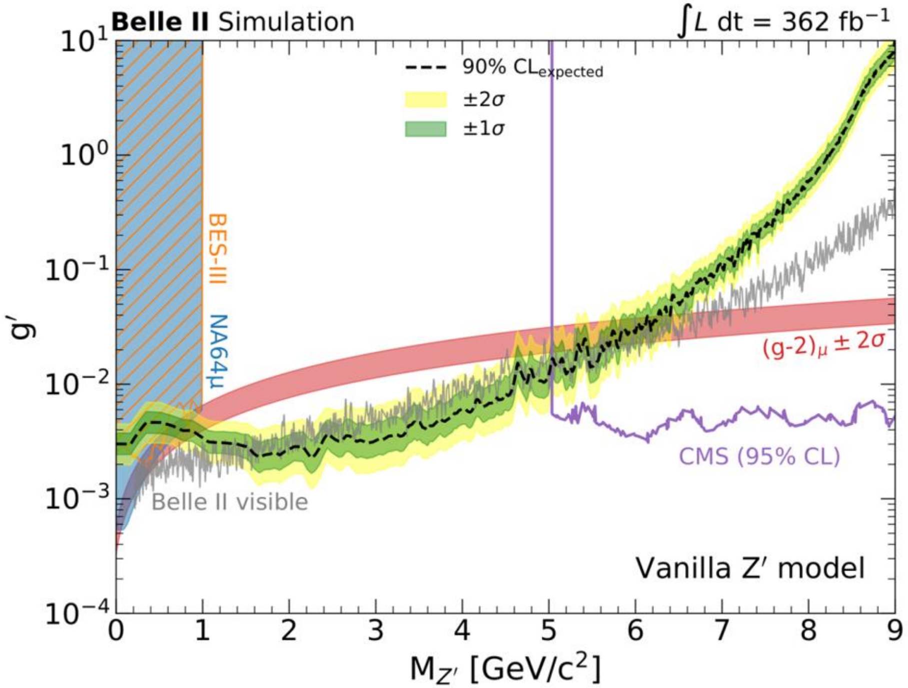
Limit setting: a how to

Enrico Graziani
INFN – Roma 3

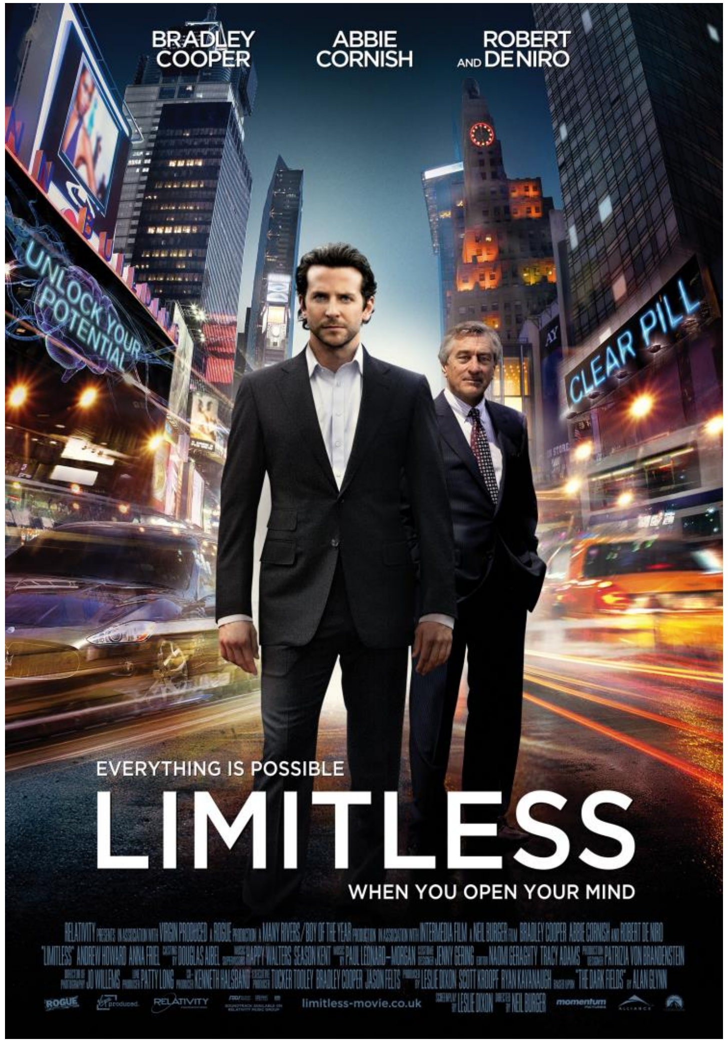


Belle II Physics Week

Kek, 14-18 October 2024

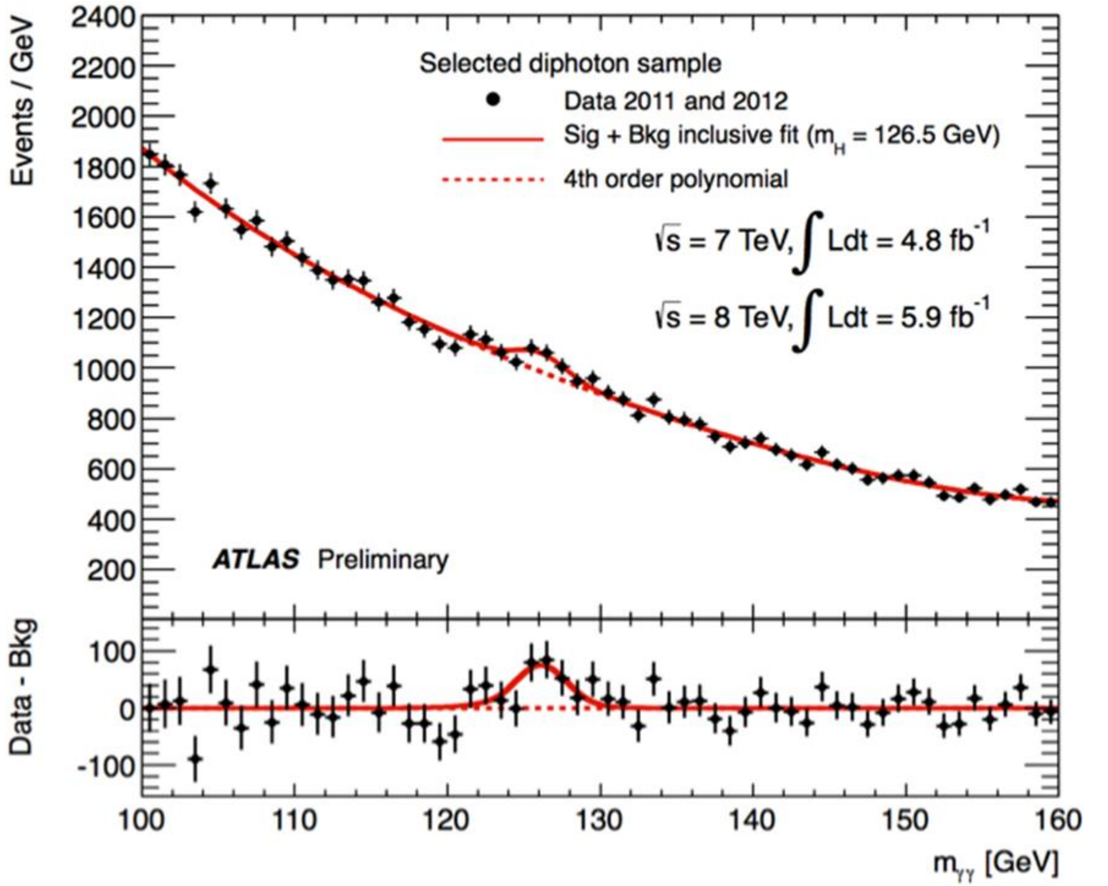
Limits in the real world

Quick search on the web for 'limits'



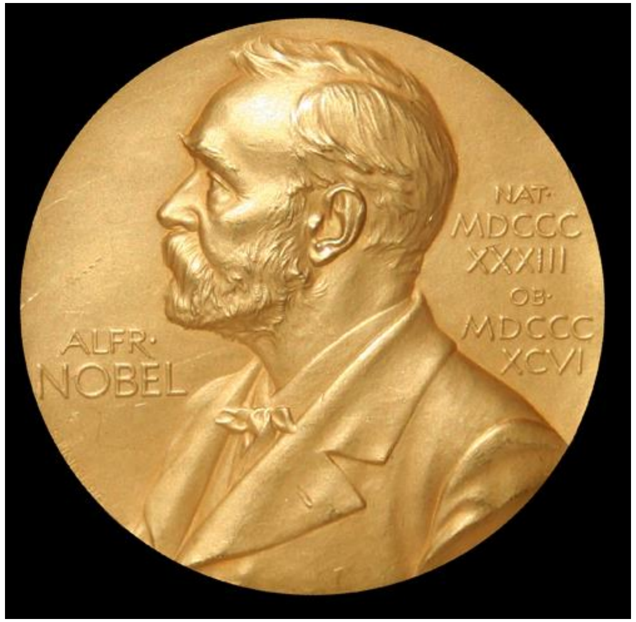
Limits in our real HEP world

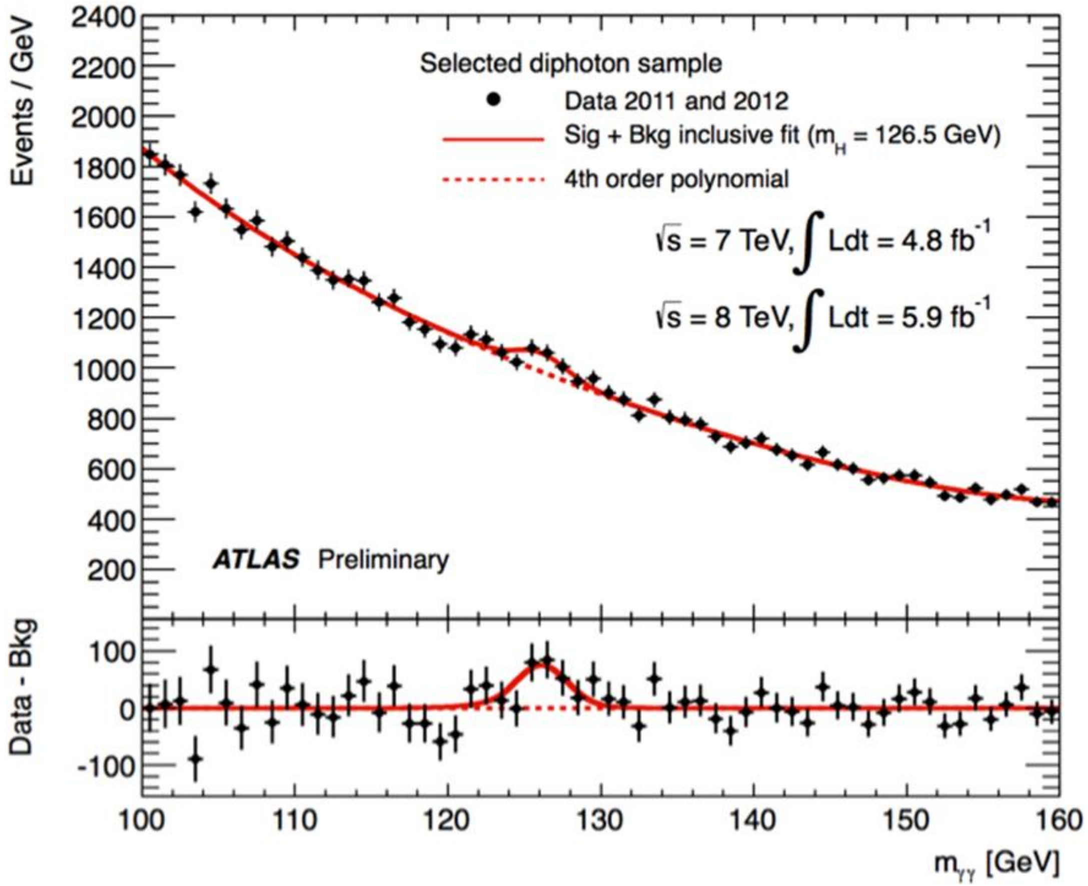
They are the other side of the nobler, but far more rare, medal side of discovery.



Limits in our real HEP world

They are the other side of the nobler, but far more rare, medal side of discovery.





Limits in our real HEP world

They are the other side of the nobler, but far more rare, medal side of discovery.

We perform a search and find just nothing. If we can’t go to Stockholm, let’s at least try to go on PRL!

PHYSICAL REVIEW LETTERS

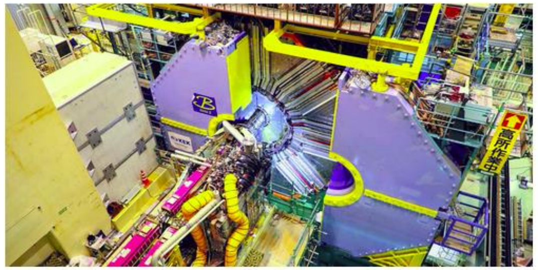
Highlights | Recent | Accepted | Collections | Authors | Referees | Search | Press | About | Editorial Team

Featured in Physics | Editors' Suggestion | Open Access

Search for an Invisibly Decaying Z' Boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ Plus Missing Energy Final States

I. Adachi *et al.* (Belle II Collaboration)
Phys. Rev. Lett. **124**, 141801 – Published 6 April 2020

Physics SYNOPSIS



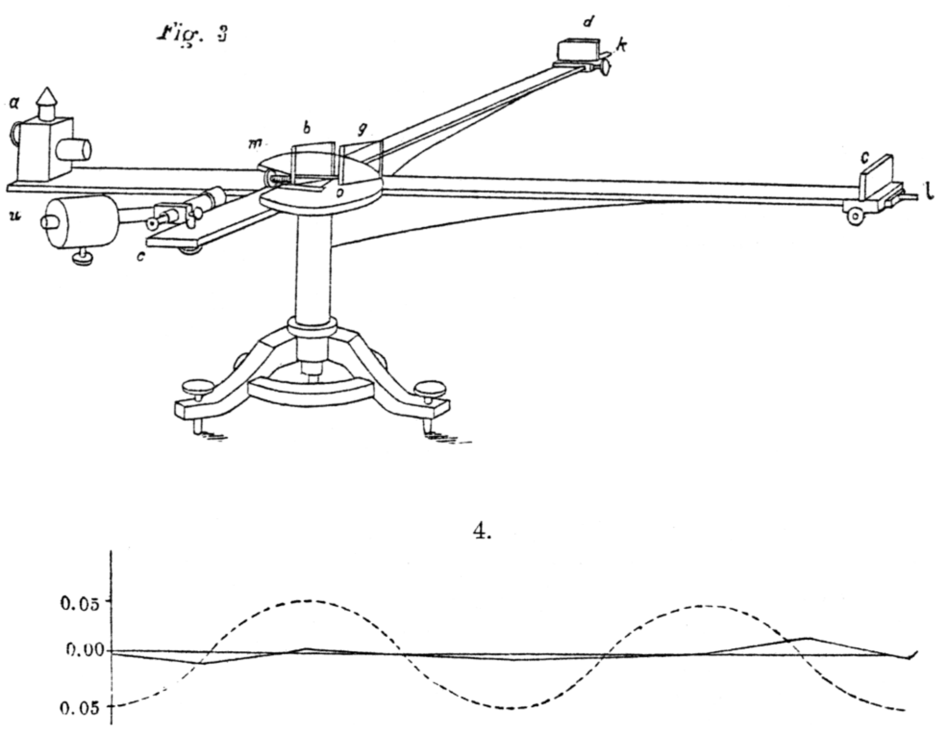
Closing in on the Z' Boson
Published 6 April 2020

The Belle II experiment finds no Z' boson in its first results, but it does constrain how strongly the particle might interact with standard model particles.

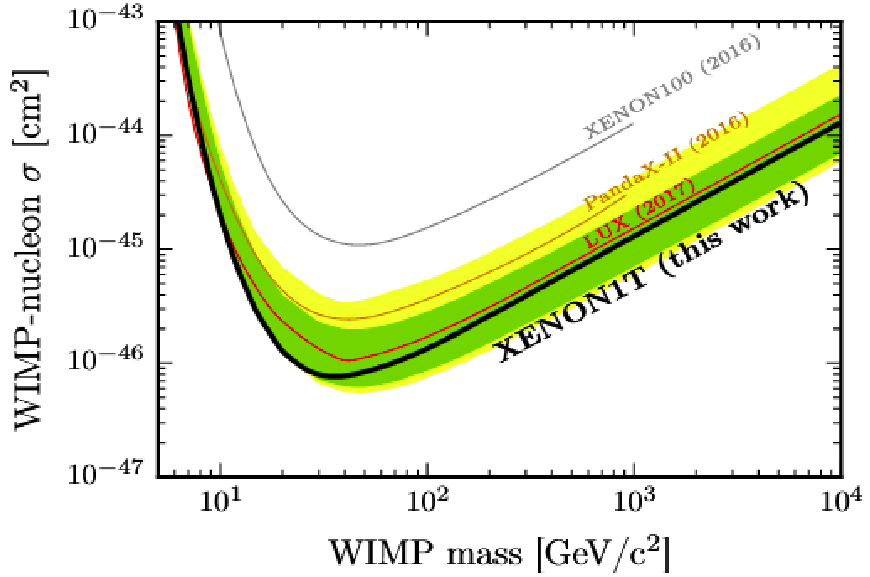
Limits

Beside healing powers (and jokes) the history of physics is full of null results that turned out to suggest new directions (think of Michelson-Morley), to stop wrong ones (supersymmetry? ☺ [half] joking again), or to shape the path towards new searches (LEP vs LHC, light dark matter, ...) → **Limits are important.**

Michelson-Morley



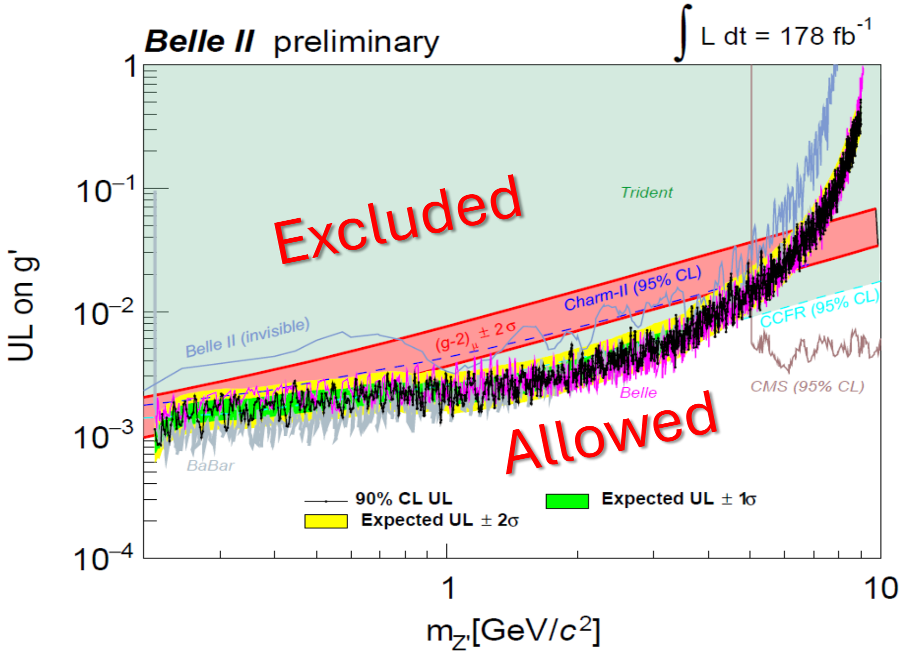
Xenon – DM search



Limits

Let's fully restrict here to our case: dark-sector physics (new BSM particles) or τ LFV/BNV decays.
We search for a positive signal over a background that can (and often is) or cannot be close to zero.

→ Limits are always upper limits.



We want to make statements about the largest possible value of a signal for a certain target probability.

Limit setting: a how to

This is an introductory lecture to limit setting

Not exhaustive

$\rightarrow$ Important methods are not covered (Feldman-Cousins, PCL, ...)

$\rightarrow$ Others are only sketched (Bayesian priors, ...)

The main goal is to give you a starting point

$\rightarrow$ To be used for further deepening and exploration of the topic, if needed

$\rightarrow$ To handle the many available tools in a responsible way, and not as black box

Complemented by the Thursday tutorial

Interpretation/definition of probability

Frequentist

Limiting relative frequency

$P(A)$ fraction of times A occurs in the limit of infinite-times measurements

Very intuitive

Well suited for intrinsic repeatable experiments (eg collisions)

Problematic for unique phenomena (eg Big Bang)

Bayesian

Subjective probability

$P(\mathrm{A})$ degree of belief that A is true. Probability associated with a hypothesis as a measure of degree of belief

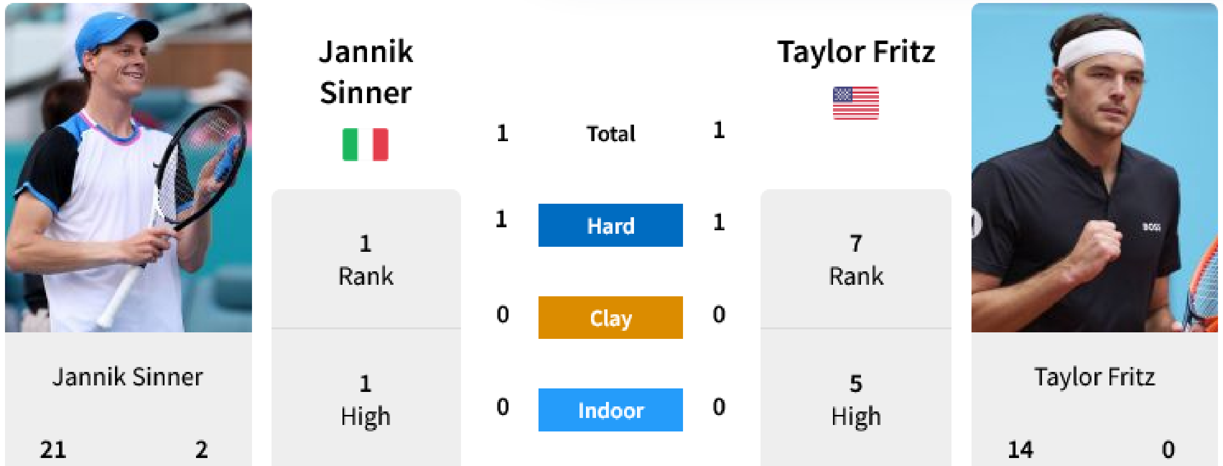
Very intuitive as well

It includes the relative frequency interpretation (considering as a hypothesis the statement that an experiment will give a specific outcome a certain fraction of times)

Interpretation/definition of probability

Frequentist

Bayesian



Win probabilities

78% *22%*

More on the frequentist-Bayesian duality

Frequentist, aka ‘classic’

- Probability that an experiment will yield a particular result

- Frequentist has to to do with P**(data|theory)** : probability to observe the data under the assumption that the theory is true.

- Probability of observations are quoted as a function of the theory parameters. These are NOT probabilities of the theory. Think of the probability to **observe** the Higgs boson, if it exists and if it has SM properties.

R.Fisher

J.Neyman

E.Pearson

More on the frequentist-Bayesian duality

Bayesian

- Bayes also start from P(data|theory), but then uses the Bayes' theorem: ***P*****(theory|data) ∝ *P(*data|theory) x *P*(theory)**.

- Posterior probability that the theory is correct after looking at data: degree of belief in the theory. Think of the probability that the SM Higgs boson **is true**, once you observed the mass peak.

- Power of Bayes' theorem: it relates the quantity of interest, the probability that the hypotheis is true given the data, to the more accessible term, the probability that we would have observed the measured data if the hypothesis was true.

- The price to pay is the introduction of ***P*****(theory)**: prior probability that the theory is true. No golden rules for that.

Bayes' theorem relates $P(A|B)$ and $P(B|A)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

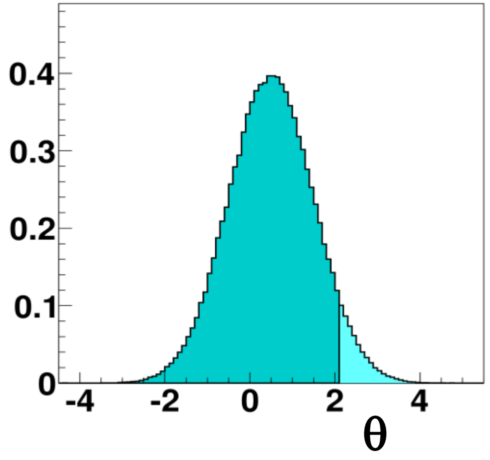
Rev. Bayes

P.S.Laplace

Even though Bayes was (posthumously) credited of the theorem, it is generally recognized that was Laplace to rediscover and use it 'seriously' (celestial mechanics)

Back on limits

Excluding a signal hypothesis implies (much) milder requests than claiming its discovery.
Typical values are **5-10% for exclusion (α) and $\sim 10^{-7}$ for discovery (p-value).**



Limits are one-sided intervals with an associated probability content.

Typical statement: $s < s_{up}$ @ 90(95)% CL

→ Confidence level → frequentist

→ Credibility level → Bayesian

Limits are measurements, but not high precision measurements:

- You can *indicatively* think at ~15% precision on s_{up}
- Proficiently used to make approximations

Duality at work

data

model parameters

$L(x \mid \theta) = P$(**data|theory**) is the likelihood.

Bayesians see the likelihood as the only way experimental data affects inference, providing a direct connection between prior and posterior. **Data are always observed data, what they are, whatever they are.**

In contrast, most of the frequentist constructions require not only the likelihood for the actual data, **but also for all possible data that *might* have been observed.**

Bayesian limits tell about the model probability. **Probability / degree-of-belief that the true value of the parameter is outside a fixed interval set by s_{up}.**

Frequentist limits tell about the probability of repeated (real, gedanken, toy, …) experiments assuming the model. **The confidence interval set by s_{up} is a random variable and fluctuates experiment by experiment.**

Nuisance parameters

Parameters that enter in our inference, but are not the goal of the measurement.

We distinguish **parameter(s) of interest (PoI)** from **nuisance parameters**.

Example: we search for signal (typically a peak in a mass distribution) over the background. We estimate the background as a necessary step of the search, often providing useful informations (MC accuracy, …), but we don't care about it per se. The background is a nuisance parameter.

The list is of course much longer: detector resolutions, shape modeling, systematic uncertainties, …

How do we treat them?

Nuisance parameters

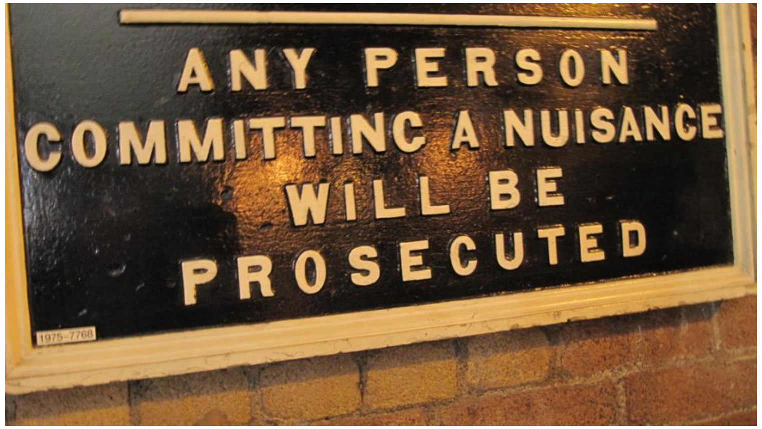
Parameters that enter in our inference, but are not the goal of the measurement.

We distringuish **parameter(s) of interest (PoI)** from **nuisance parameters**.

Example: we search for signal (typically a peak in a mass distribution) over the background. We estimate the background as a necessary step of the search, often providing useful informations (MC accuracy, ...), but we don't care about it per se. The background is a nuisance parameter.

The list is of course much longer: detector resolutions, shape modeling, systematic uncertainties, ...

Some treatment must be **absolutely** devised



Limit setting in practice: a simple example

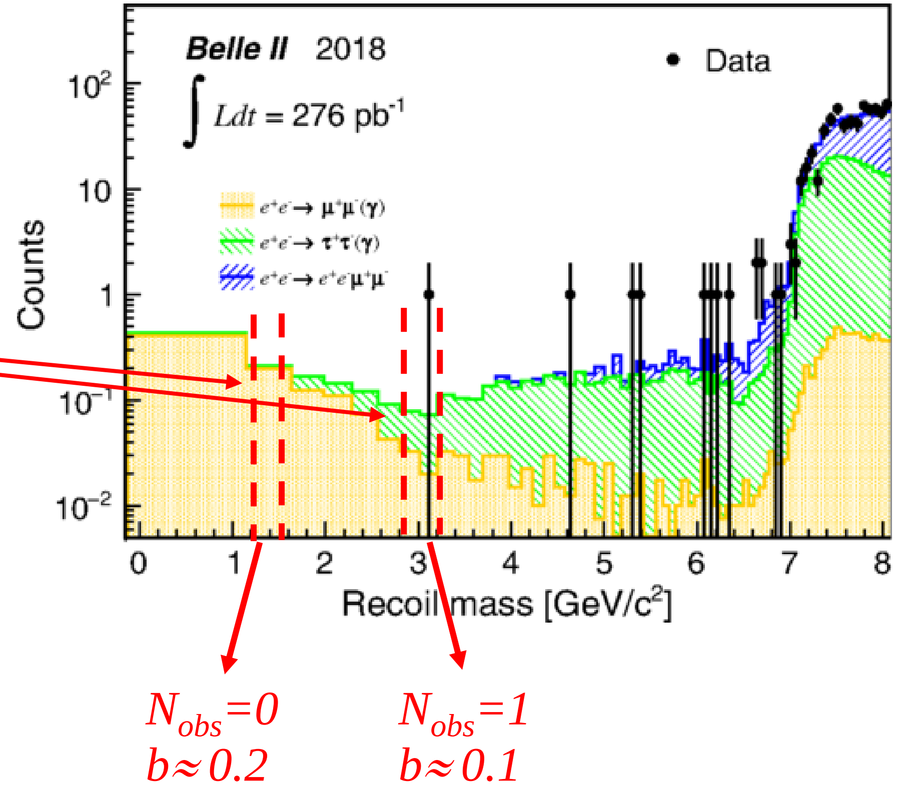
Poissonian event counting

We search for a signal process on top of a background process.
This is often a counting experiment with small number of events.

No full signal/background discrimination → we measure the total number of events only.

Signal, background and signal+background counts follow Poissonian distributions with expected values s, b and $s+b$.

We observe N_{obs} events.
Let's suppose for now b is known with no uncertainty.



Counts in mass windows

If N_{obs} roughly compatible with b, our goal is to find an upper limit s_{up} on s at, say, 95% CL.

Limit setting in practice: a simple example

Poissonian event counting: **frequentist**

s+b hypothesis → $$P(n/s+b) = \frac{(s+b)^n e^{-(s+b)}}{n!}$$

Find the value s_{up} such that the probability to observe N_{obs} events or less is α

Does it remind something to you?

→ Inversion of the p-value test for the s+b hypothesis

$$\alpha = P(n \leq N_{obs}) = \sum_{n=0}^{N_{obs}} \frac{(s_{up}+b)^n e^{-(s_{up}+b)}}{n!}$$

α = 1-CL=0.05

Solve numerically (or analytically!) for s_{up}

Limit setting in practice: a simple example

Poissonian event counting: **frequentist**

A semi-analytically solution is actually available

$$s_{up} = \frac{1}{2} F^{-1}_{\chi^2}(1-\alpha; 2(N_{obs}+1)) - b$$

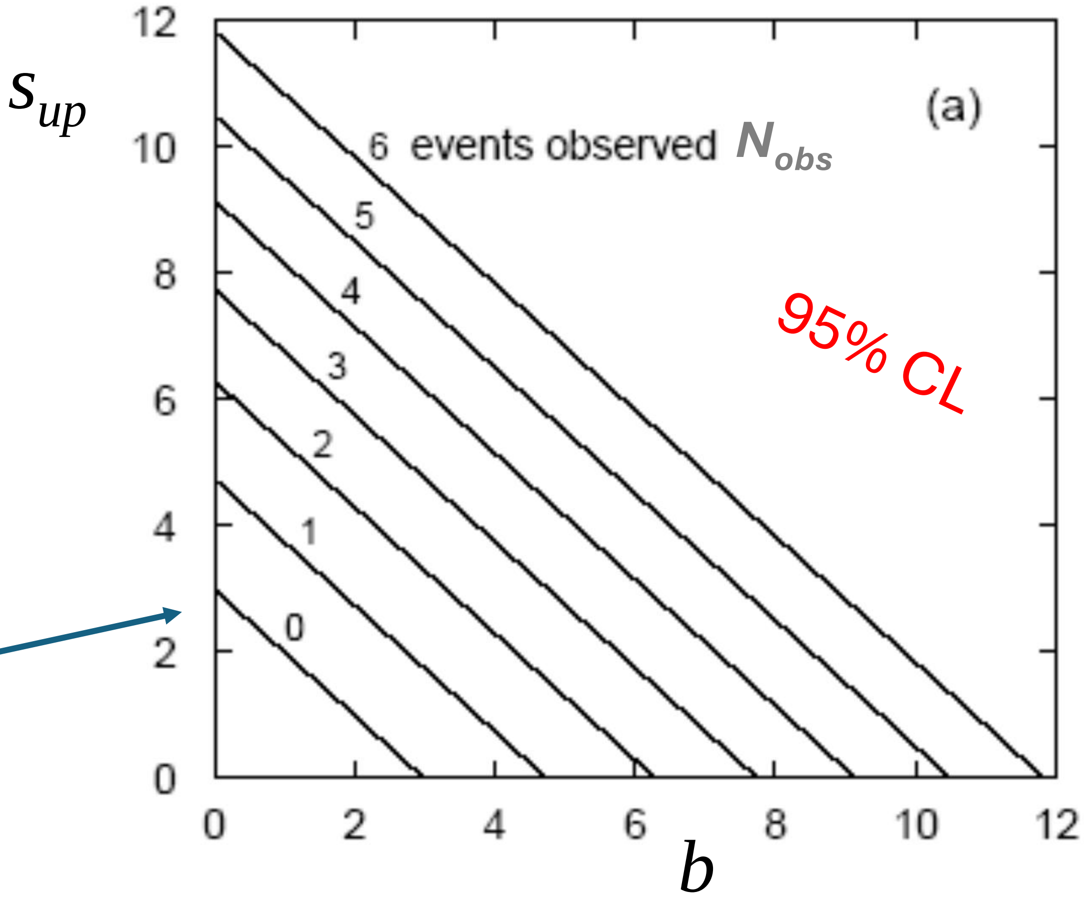
Quantile of χ^2 distribution

Notable case: $b=0,\ N_{obs}=0$

$$\alpha = P(n \leq N_{obs}) = e^{-s_{up}}$$

$$s_{up} = -\log \alpha \approx 3.0$$

More generally, for $b \neq 0,\ N_{obs} \neq 0$



Limit setting in practice: a simple example

s, s_{up} are (expected) numbers of signal event

$N = \varepsilon\, \sigma\, L_{int}$ connects number of events with cross section σ, efficiency ε and integrated luminosity L_{int}

$$\sigma_{up} = \frac{s_{up}}{\epsilon \times L_{int}}$$

← Neglecting smaller effects (correlations, systematics, …)

is the upper limit on the cross section, which defines in turn a limit on the coupling: usually $\sigma \propto g^2$

Sometimes **a strength factor** $\mu = \sigma / \sigma_{ref}$ is introduced and used to quote limits, when there is a special reference cross section to which to compare (ie σ_{SM} for the SM Higgs boson). Rarely the case in Belle II, but we can always pick up a specific coupling/cross section and refer to it.

More on limits: coverage

Coverage is the fraction of intervals (probability) that the exclusion range set by the upper limit does not contain (cover) the parameter when the model is true.

Ideally, one expects this to be CL.

As such, it is a frequentist concept, and automatically plugged in the frequentist construction.
Due to approximations and/or intentional departures from the pure frequentism, there can be cases of over or under coverage.

In the Bayesian case, coverage is anyway an auspicable feature, and needs to be checked.

Limit setting in practice: a simple example

Poissonian event counting: **frequentist**

For underfluctuations of N_{obs} , s_{up} can result negative: confidence interval is empty

$N_{obs} = 0$, $b = 3.2$ $\alpha = 0.05$ (95% CL) $\rightarrow$ $\mathbf{s_{up} \approx -0.2}$

Not an uncommon case for limits near a physical boundary.

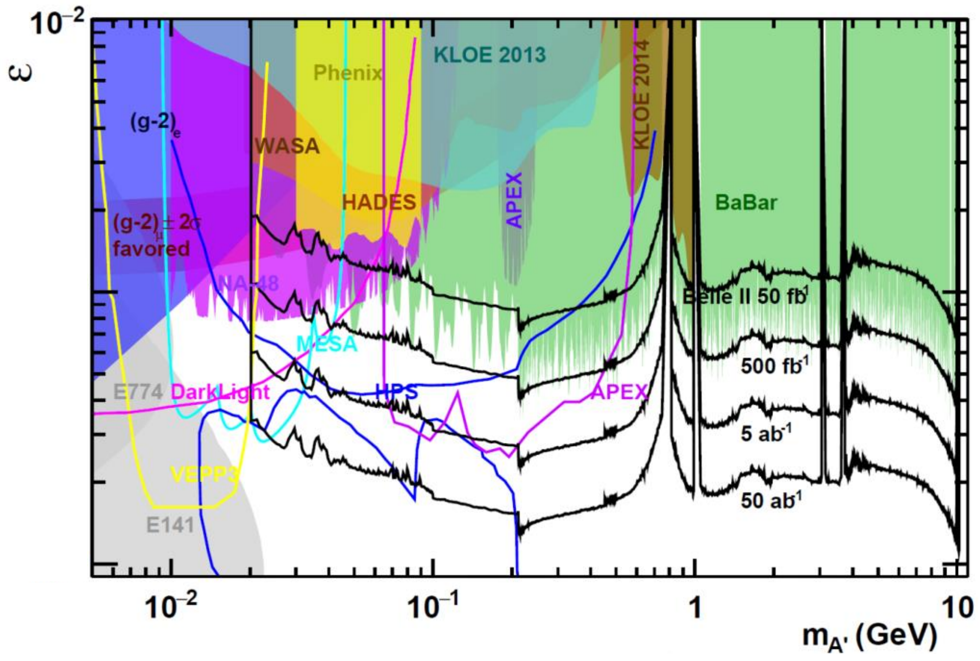
The interval is designed to cover the true value 95% of the times, and this case belongs to the remaining 5%.

Math is ok, but the result is unphysical.

Bayesian avoids this, because the prior of s, however chosen, will be 0 for $s<0$.

Limit setting in practice: spurious exclusions

Poissonian event counting: **frequentist**

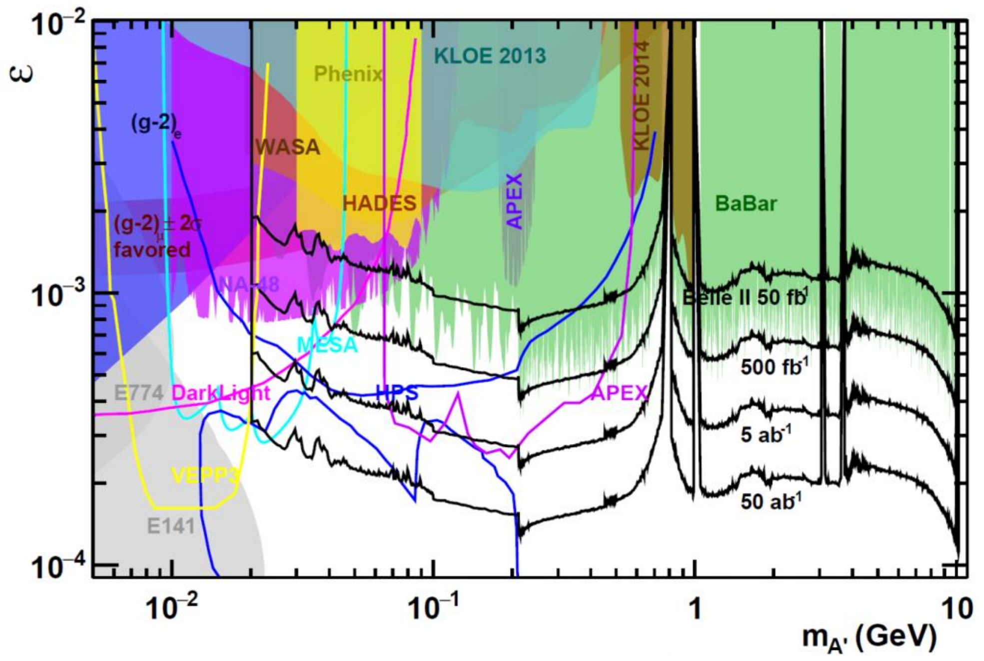


Imagine we get a limit on $\varepsilon < 0$ for some $m_{A'}$.
Let's close an eye about the *unphysicality*.

- Should we be happy that the constraint is so tight?
- Would this mean that games are over for that $m_{A'}$?
- Would really SM ($\varepsilon = 0$) be **excluded** in that region??
- Should we believe at all in this result?

Limit setting in practice: spurious exclusions

Poissonian event counting: **frequentist**



Imagine we get a limit on $\varepsilon < 0$ for some $m_{A'}$.
Let's close an eye about the *unphysicality*.

- Should we be happy that the constraint is so tight?
- Would this mean that games are over for that $m_{A'}$?
- Would really SM ($\varepsilon = 0$) be **excluded** in that region??
- Should we believe at all in this result?

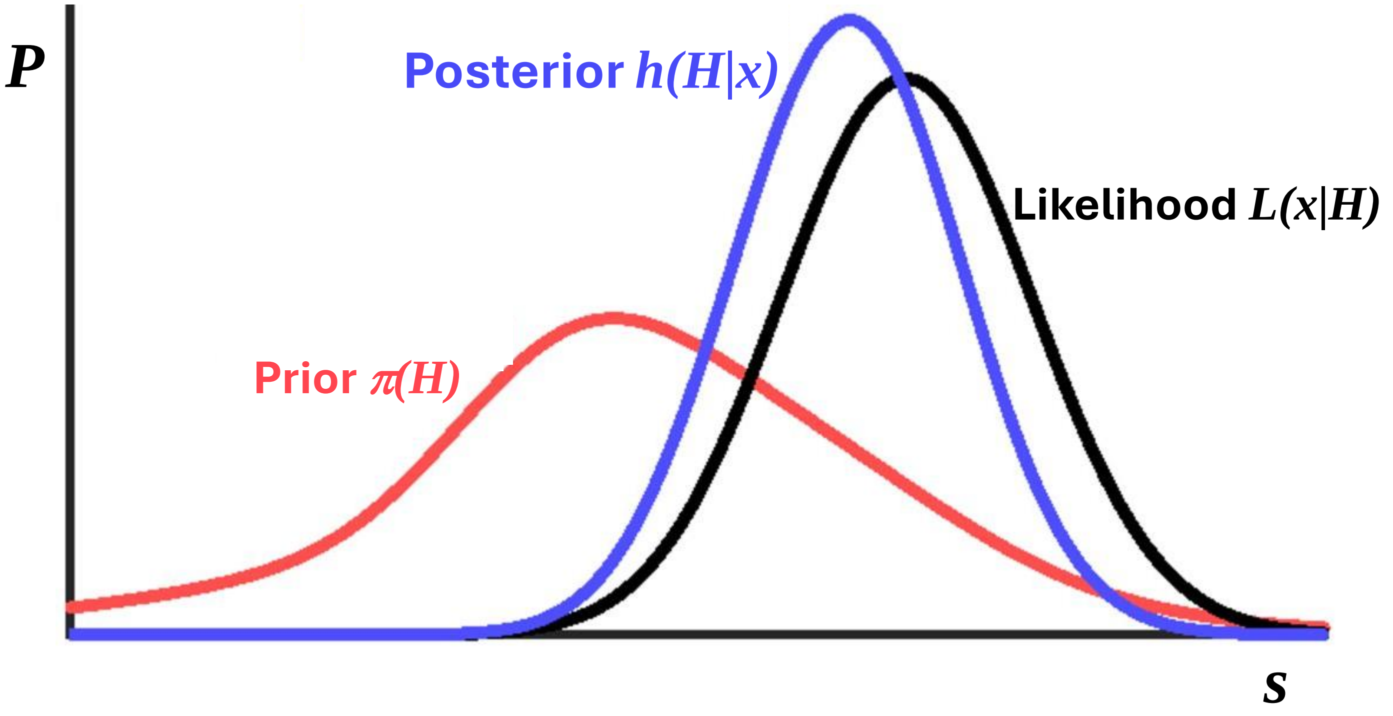
The answers are of course **NO**.
This is a real problem that we want to solve (we will ☺)

Bayesian limits

$$h(H|x) = \frac{L(x|H)\pi(H)}{\int L(x|H)\pi(H)dH}$$

- Hypothesis (model) $H=H(s)$
- Prior model probability $\pi(H)$
- Likelihood $L(x|H)$

→ Posterior probability $h(H|x)$

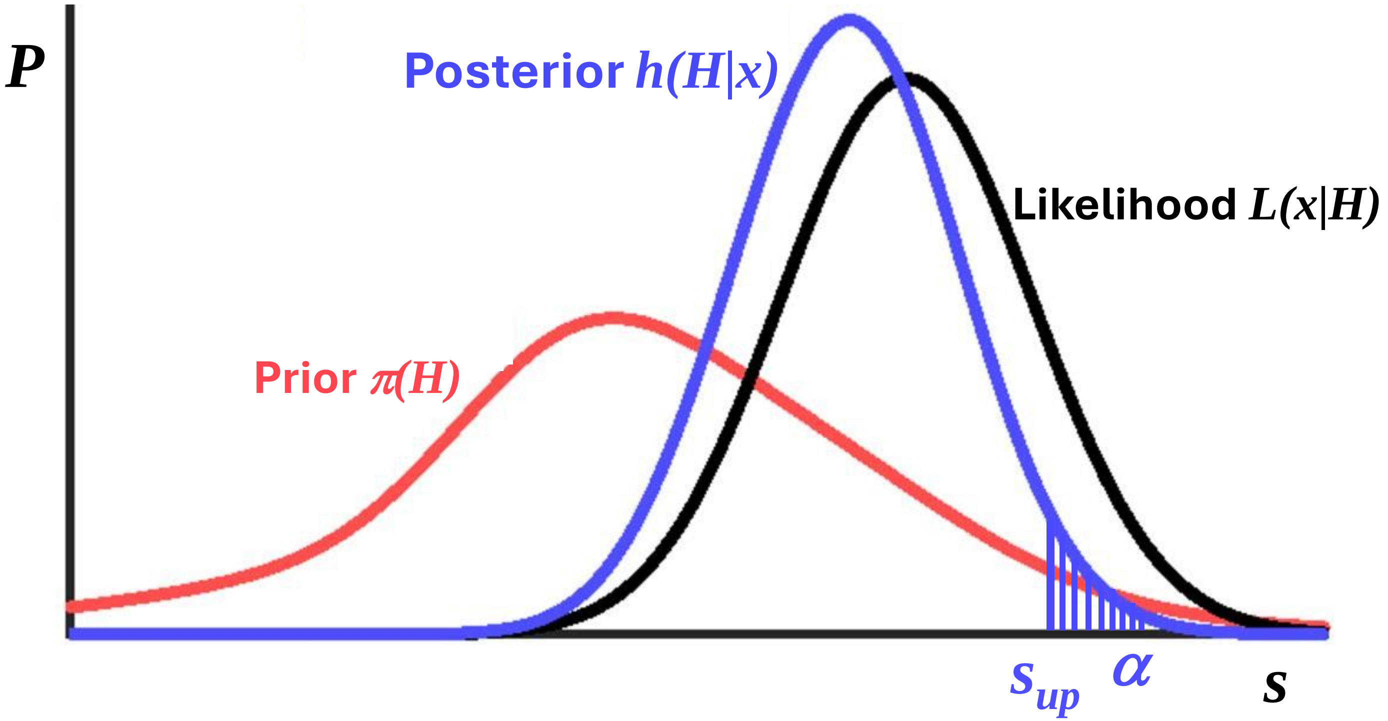


Bayesian limits

$$h(H|x) = \frac{L(x|H)\pi(H)}{\int L(x|H)\pi(H)dH}$$

- Hypothesis (model) $H=H(s)$
- Prior model probability $\pi(H)$
- Likelihood $L(x|H)$

→ Posterior probability $h(H|x)$



Upper limit on s by integrating over models, with fixed data, such that the posterior probability of the excluded theories is $CL=1-\alpha$

Limit setting in practice: a simple example

Poissonian event counting: **Bayesian**

Remember: b is known with no uncertainty (for now)

$$h(s|N_{obs}) = \frac{\frac{(s+b)^{N_{obs}} e^{-(s+b)}}{N_{obs}!}\pi(s)}{\int_0^{\infty}\frac{(s'+b)^{N_{obs}} e^{-(s'+b)}}{N_{obs}!}\pi(s')ds'}$$

→ Posterior probability with Poisson likelihood

The limit on s is found by integrating with fixed N_{obs}, such that the posterior probability of the excluded theories is *CL=1-α*

$$1-\alpha = \frac{\int_0^{s_{up}} h(s')ds'}{\int_0^{\infty} h(s')ds'}$$

and then solve for s_{up}

Limit setting in practice: a simple example

Poissonian event counting: **Bayesian**

We have to define a prior $\pi(s)$ for the signal

$$\pi(s) = \begin{cases} 1 & s \geq 0 \\ 0 & s < 0 \end{cases}$$

flat prior

$$h(H|x) = \frac{L(x|H)\pi(H)}{\int L(x|H)\pi(H)dH}$$

Posterior $\propto$ likelihood $\rightarrow$ posterior peak coincides with maximum likelihood estimators.

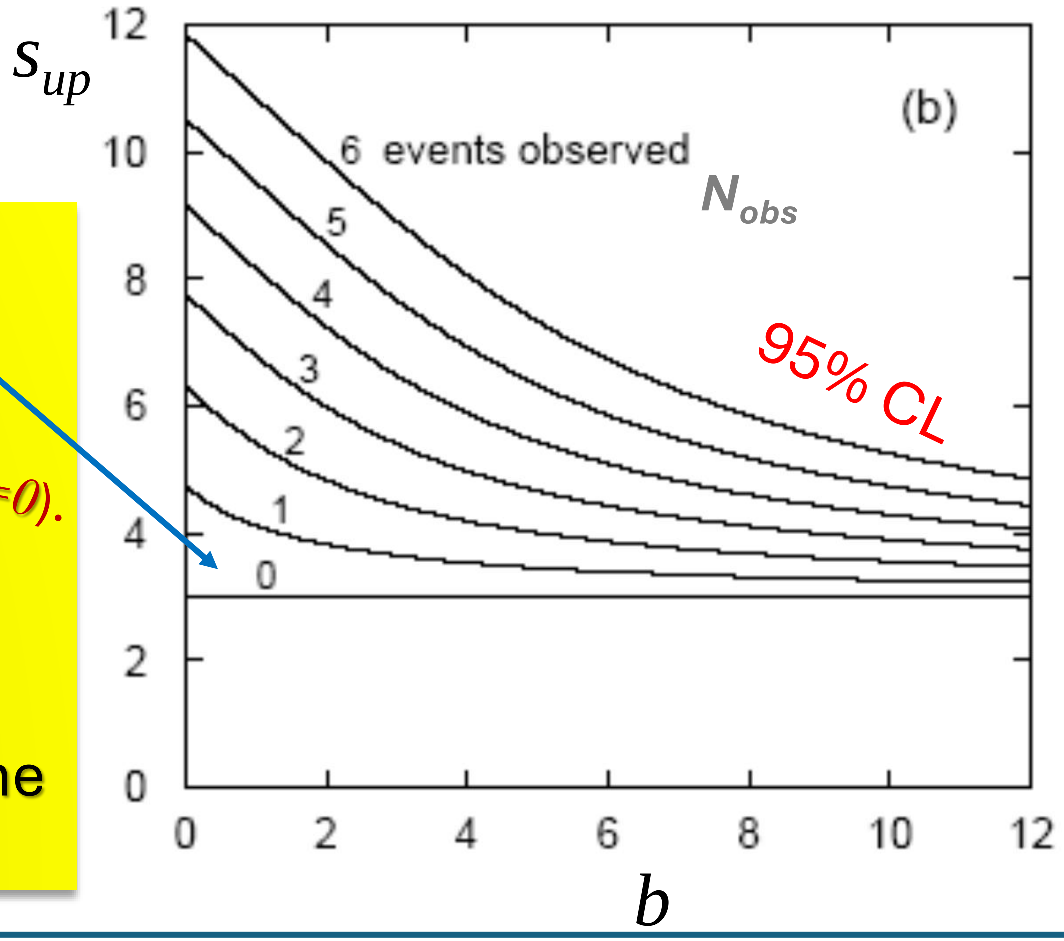
- Not normalized ('improper prior'), but the Poissonian likelihood penalizes high s because of N_{obs}.
- Reasonable, since it reflects our degree of belief in the signal (non-negative), expresses ignorance about the rest, and is widely used as a reference for counting experiments.
- Criticized, because it does not represent a degree of belief and the probability of having s in any finite interval approaches zero.

Limit setting in practice: a simple example

Poissonian event counting: **Bayesian**

Flat prior $\pi(s)$ ⟶ $$\alpha = e^{-s_{up}} \frac{\sum_{m=0}^{N_{obs}} \frac{(s^{up}+b)^m}{m!}}{\sum_{m=0}^{N_{obs}} \frac{b^m}{m!}}$$

More generally, for $b \neq 0$, $N_{obs} \neq 0$



Notable case: $N_{obs}=0$ → $s_{up} = -\log \alpha \approx 3.0$

does not depend on b

Same as for the frequentist ($b=0$). The interpretation is different!

$$\sigma_{up} = \frac{s_{up}}{\epsilon \times L_{int}}$$

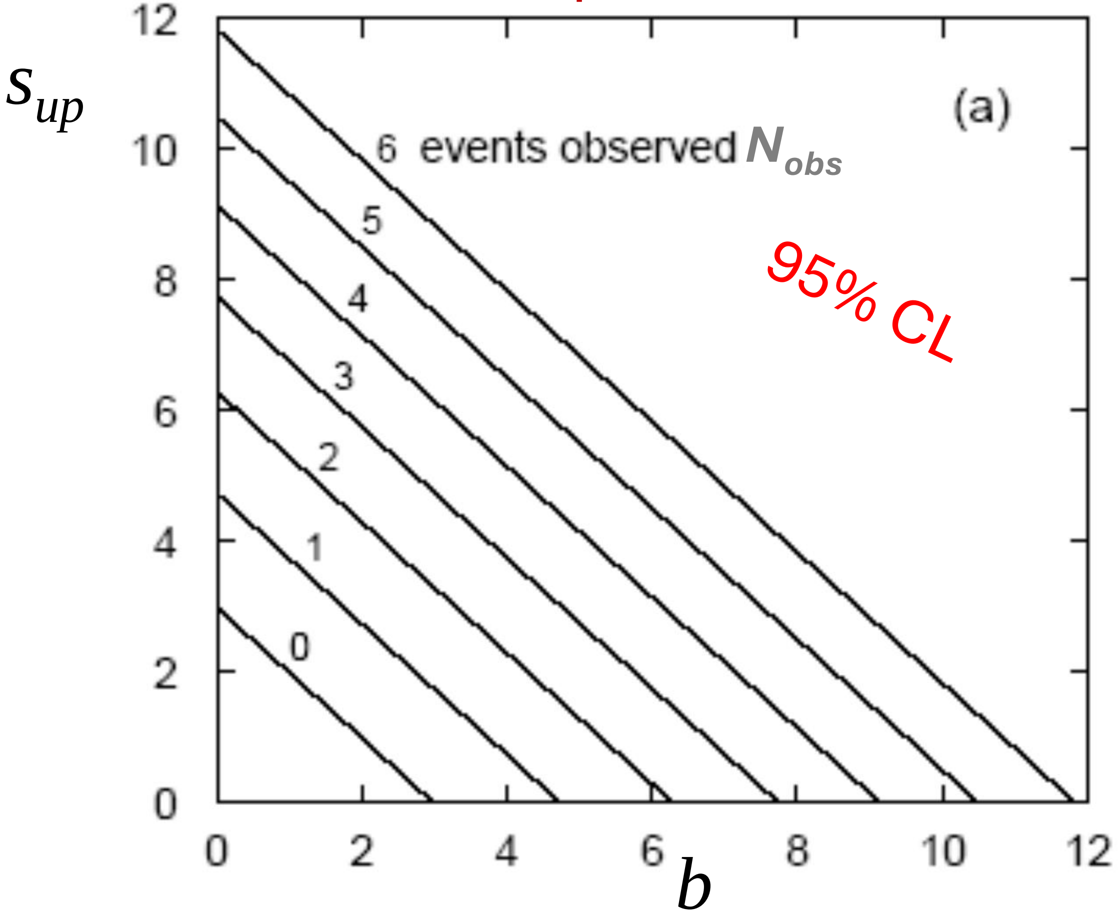
$b \propto L_{int}$

$N_{obs}=0$ → limits on cross section scale linearly with the luminosity.

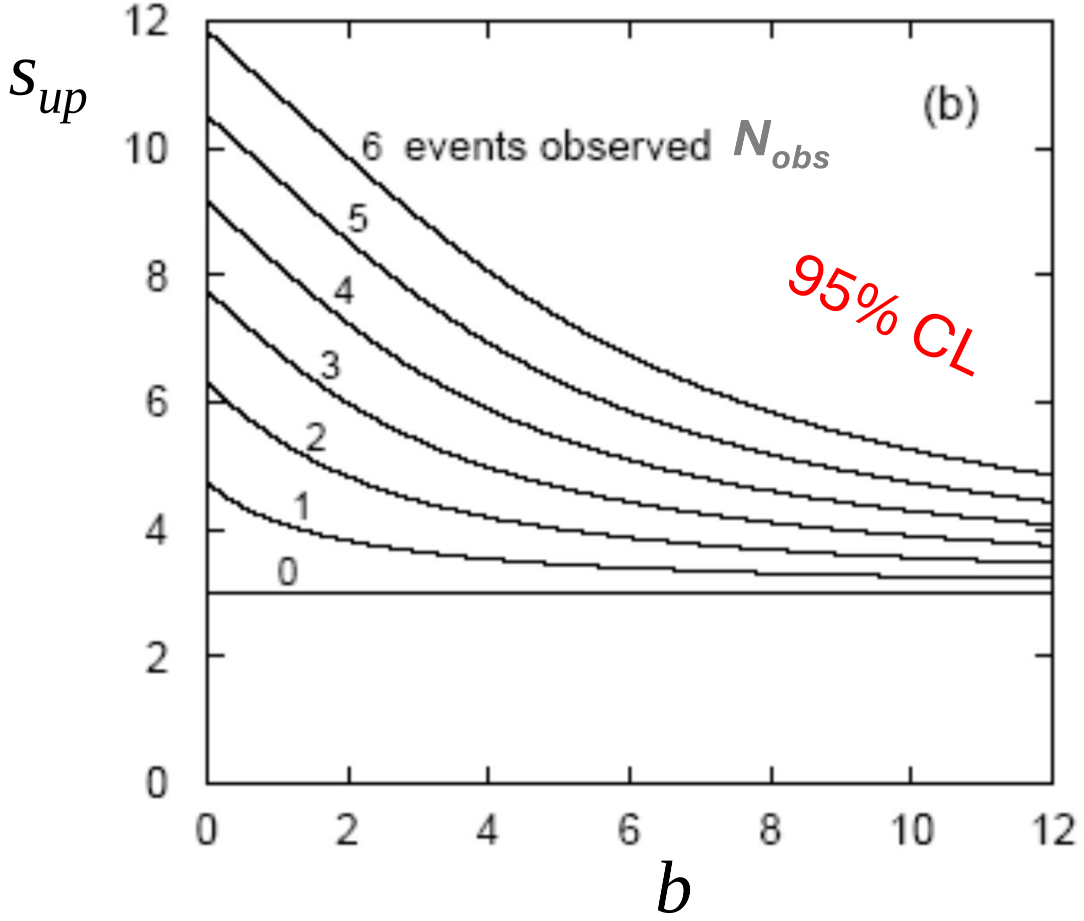
Limit setting in practice: a simple example

Frequentist vs Bayesian

frequentist



Bayesian



Bayesian limits in general greater (conservative?) than frequentist and never go negative

Bayesian priors

Dependence on prior is undoubtly the main weakness of Bayesian methods

Flat prior $\pi(s)$ often used. But flat on what?

- Flat in s $\rightarrow$ flat in cross section $\rightarrow$ not flat in coupling.
- In general, not flat for functions of the parameter.

Not the only possible choice. A **log-flat** prior would express our belief (wherever it comes from) that the probability of a non-zero signal extends uniformly over orders of magnitude: $\pi(s) \propto 1/s$.

On the other hand, dependence of final results upon the assumed prior is often negligible or small (remember: limits are not precision measurements!)

- This dependence have to be always checked: **sensitivity analysis.**

More on Bayesian priors

Attempts to subtract some degree of subjectivity by deriving the prior probabilities from formal rules: to satisfy certain invariance principles or to provide maximum information gain.
Often called 'objective priors' (not to be taken too literally), as opposed to 'subjective priors'.

They don't fully express a degree of belief: useful in comparing results obtained with subjective priors, producing intervals whose (even frequentist) properties can be studied.

Jeffreys' priors

$$\pi(\theta) \propto \sqrt{\det(\mathrm{I}(\theta))}$$

$$I_{ij}(\theta) = -E\left[\frac{\partial^2 \ln L(x|\theta)}{\partial\theta_i \partial\theta_j}\right] = -\int \frac{\partial^2 \ln L(x|\theta)}{\partial\theta_i \partial\theta_j} L(x|\theta)\, dx$$ Fisher information matrix

This is shown to lead to invariance under transformation of parameters. For Poisson(θ) $\propto \frac{1}{\sqrt{\theta}}$

In our case → $\pi(s) \propto \frac{1}{\sqrt{s+b}}$

Surprised?

Bayesian limits with non-fixed background

Conceptually (but not necessarily computationally) trivial extension to the case of non-fixed b: introducing uncertainties, both statistical and systematic, on the background

We introduce a prior $\pi_b(b)$ for the background: eg Gaussian, to parametrize systematic uncertainties of size σ_b on the background b taken from Monte Carlo

$$\pi_b(b) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{\frac{-(b'-b)^2}{2\sigma_b^2}}$$

... and then we marginalize

$$h(s|N_{obs}) = \frac{\int_0^\infty \frac{(s+b)^{N_{obs}} e^{-(s+b)}}{N_{obs}!} \pi(s) \pi_b(b') db'}{\int_0^\infty \int_0^\infty \frac{(s'+b)^{N_{obs}} e^{-(s'+b)}}{N_{obs}!} \pi(s') ds' \pi_b(b') db'}$$

As for the signal, $b \geq 0$. Not guaranteed by a Gaussian, unless σ_b is small compared to b. Otherwise, one is forced to truncate and renormalize. Does this still represent true uncertainty?

More on Bayesian priors for the background

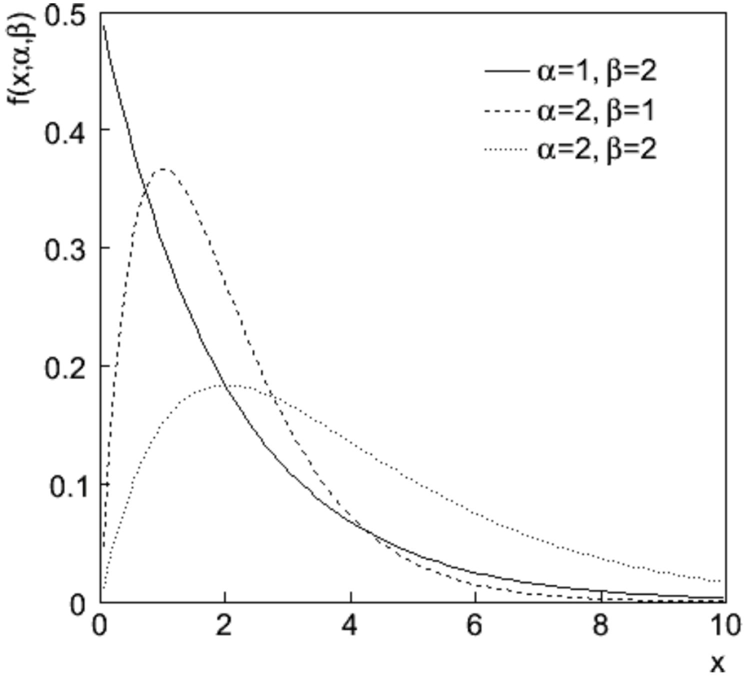
Here is an alternative.

The estimate of the background comes from a control sample.
We search for signal in a Poisson($s+b$) distribution and we evaluate the background in a Poisson(kb) distribution, with a (known) scale factor k.

Example (simplified!): we search for a resonance that decays in $\mu^+\mu^-$, we estimate the background looking at $e^+\mu^- + e^-\mu^+$ with k taking into account the different PID and combinatorial.

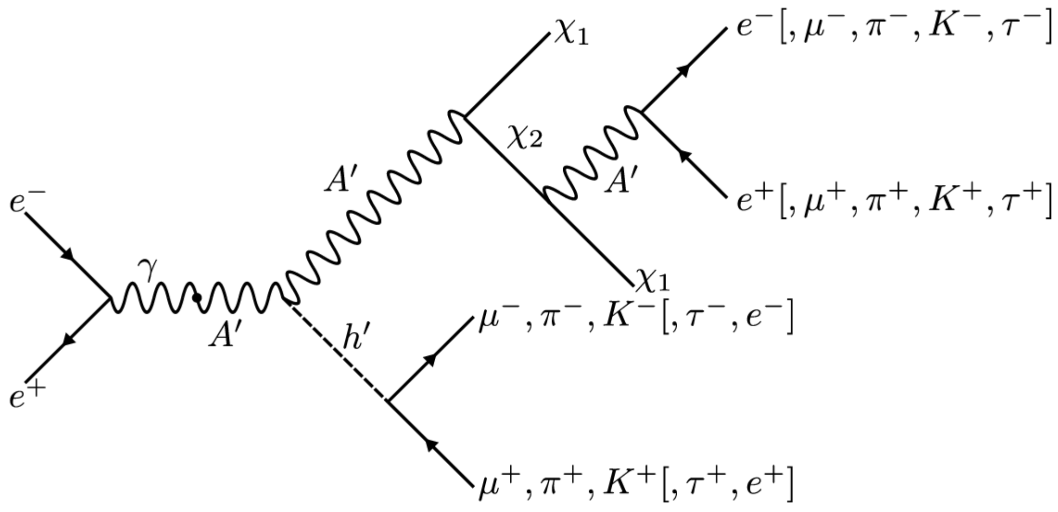
Posterior $\pi(b)$ after looking at $e\mu$ events
$$\pi(b) = \frac{kb^{N_{obs}^{e\mu}} e^{-kb}}{N_{obs}^{e\mu}!} \pi_0(b)$$

If the 'original' $\pi_0(b)$ is assumed flat, then $\pi(b)$, **which is the background prior for the $\mu^+\mu^-$ search**, is a Gamma distribution, with better properties than the Gaussian

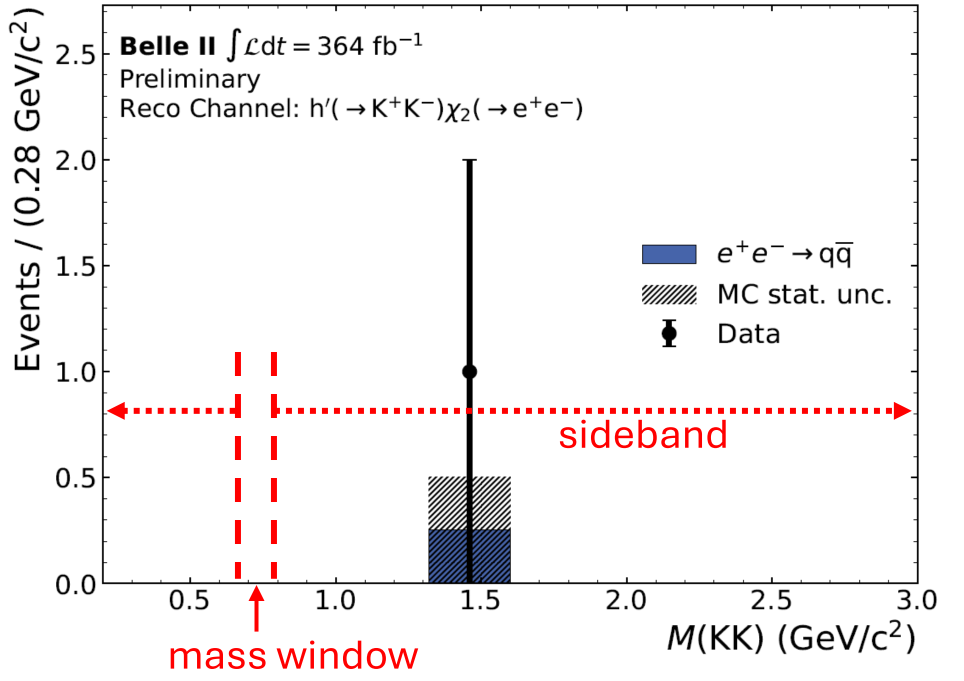


Priors for the background: a Belle II example

Inelastic dark matter + dark Higgs h'



Search for a peak $h' \to \mu^+\mu^-$, $\pi^+\pi^-$, K^+K^-



Expected background ~ 0.

Measured directly in data through sidebands.

Mass windows of width 1-5 MeV. Sideband is the full mass spectrum excluding the mass window.

Ratio of sideband width to mass window width $f \sim 1000$.

Priors for the background: a Belle II example

Inelastic dark matter + dark Higgs h'

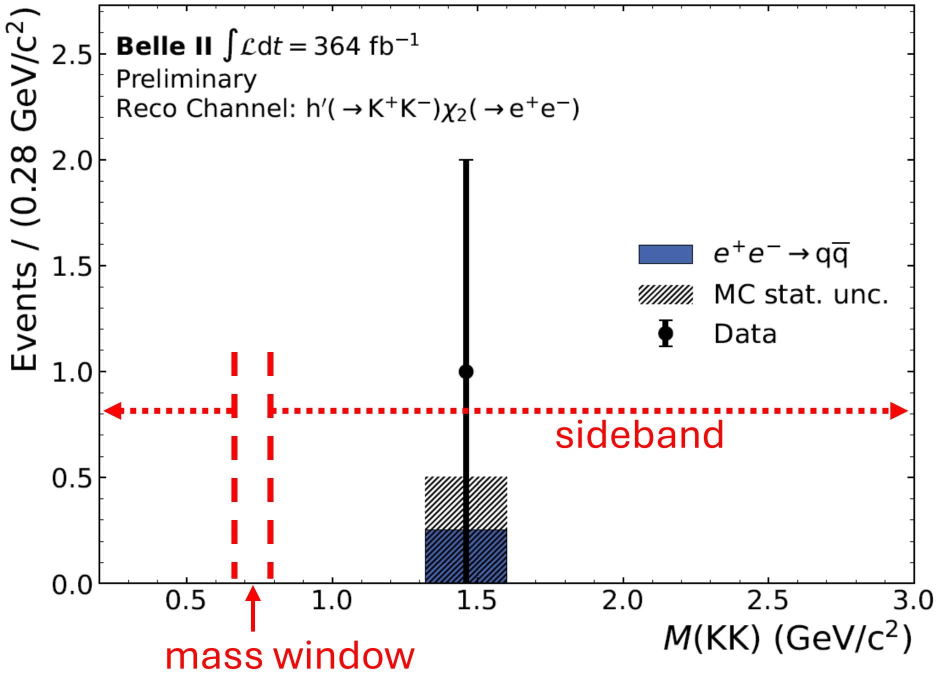
Search for a peak $h' \to \mu^+\mu^-$, $\pi^+\pi^-$, K^+K^-

Assume uniform background.

Expected background in sideband $b_{SB} \to$ nuisance parameter

Expected background in mass window b_{SB}/f

Count events N_{obs} and N_{obs}^{SB} in mass window and sideband



$$L(s, b_{SB}) = \frac{(s + b_{SB}/f)^{N_{obs}} e^{-(s+b_{SB}/f)}}{N_{obs}!} \times \frac{b_{SB}{}^{N_{obs}^{SB}} e^{-b_{SB}}}{N_{obs}^{SB}!}$$

$$L(s) = \int_0^\infty db_{SB}\; L(s, b_{SB})$$

Even with 0 observed events in the sideband (in Monte Carlo too!) and in the signal window, this is perfectly manageable and accounts for all statistical fluctuations through the two Poissonians.

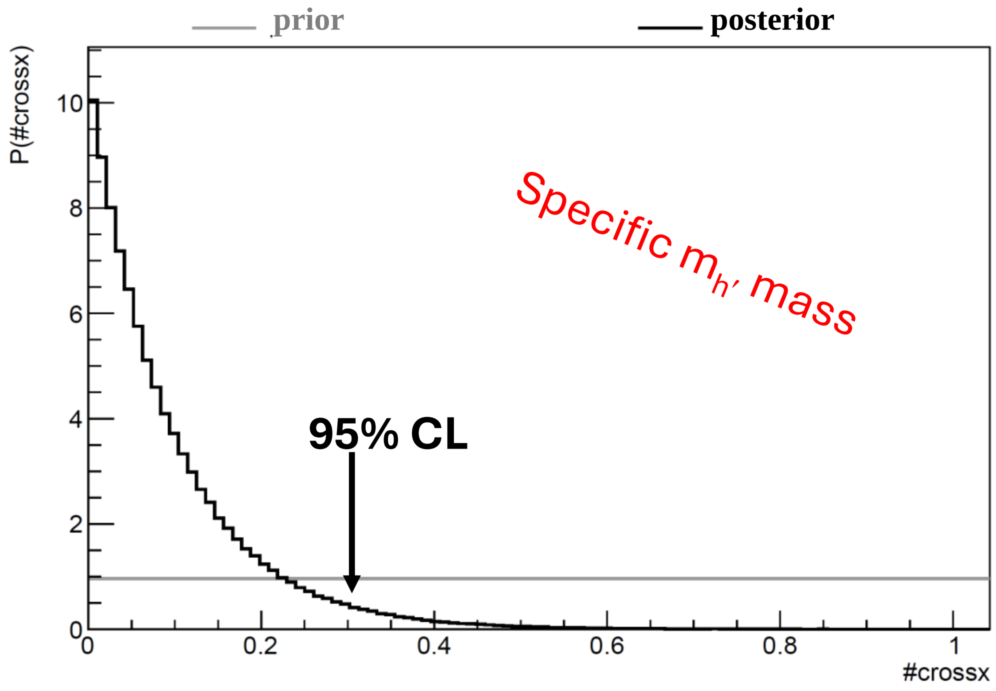
Priors for the background: a Belle II example

Inelastic dark matter + dark Higgs h'

Assume uniform background, but add a systematic uncertainty Δ to keep into account possible departures from uniformity.

$b_{SB}/f \to b_{SB}(1+\Delta)/f$ $\pi(\Delta)$ Gaussian with width σ_Δ

Include in the model and marginalize



$$L(s, b_{SB}, \Delta) = \frac{(s + b_{SB}(1+\Delta)/f)^{N_{obs}} e^{-(s+b_{SB}(1+\Delta)/f)}}{N_{obs}!} \times \frac{{b_{SB}}^{N_{obs}^{SB}} e^{-b_{SB}}}{N_{obs}^{SB}!} \times g(\Delta, \sigma_\Delta)$$

$$L(s) = \int_{-\infty}^{\infty} d\Delta \int_0^{\infty} db_{SB}\, L(s, b_{SB}, \Delta)$$

Simple, isn't it? ☺

Bayesian limits: summary

Very well suited for counting experiments

Clear framework for the treatment of nuisance parameters and thus systematic uncertainties

- Marginalize, marginalize, marginalize, ...

Main issue: subjectivity in the choice of priors

- Use the flat prior
- Compare results with at least another prior: log-flat, Jeffreys, ...

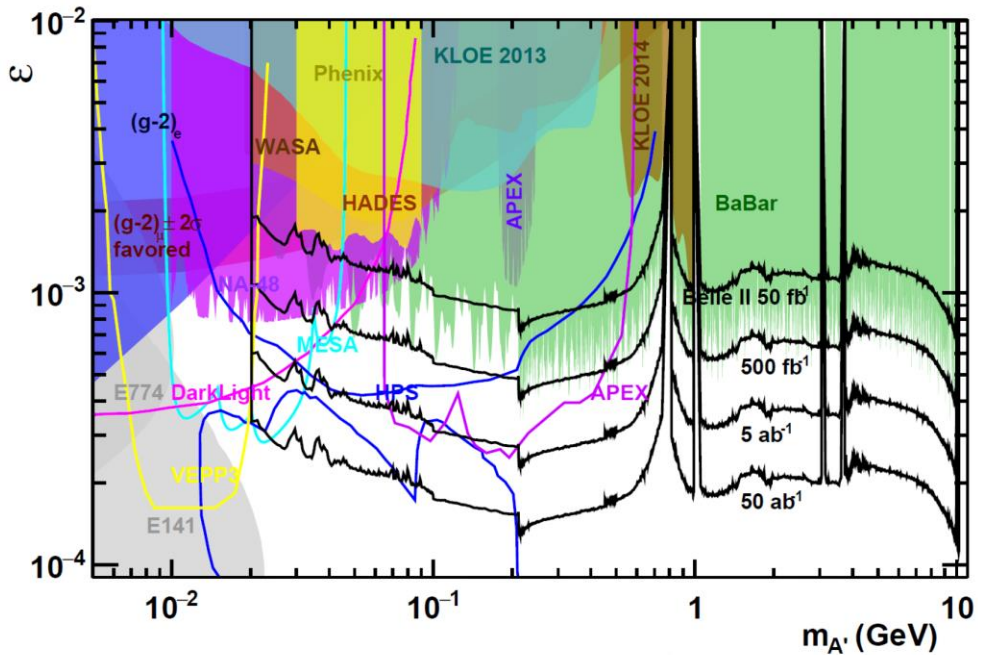
Models (likelihoods) can easily get very complicated

- Very rarely (semi)-analytically solvable
- You need some tool to perform numerical multidimensional integrations: most of them based on Monte Carlo Markov Chains aka MCMC: BAT (Bayesian Analysis Toolkit), BPULE (Bayesian Poissonian Upper Limit Estimator)
- ... and then you have the pyhf world

Coverage to be checked: typically with toys

Back to frequentist: spurious exclusions

Poissonian event counting: **frequentist** (reprise)



Imagine we get a limit on $\varepsilon < 0$ for some $m_{A'}$.
Let's close an eye about the *unphysicality*.

- Should we be happy that the constraint is so tight?
- Would this mean that games are over for that $m_{A'}$?
- Would really SM ($\varepsilon = 0$) be **excluded** in that region??
- Should we believe at all in this result?

The answers are of course **NO**.
This is a real problem that we want to solve (we will ☺)

Limit setting in practice: spurious exclusions

Limit $s_{up}<0$, in presence of a physics boundary $s\geq 0$.

Problem of excluding parameter values with not sufficient information to distinguish between the b and $s+b$ hypotheses (small signal rates, background ~signal, lack of discrimination due to physics or experimental resolution).

Ideally, we would prefer that, in these cases, the signal is not excluded.

To spot the problem, we can always compare the ***'observed'*** limit with the ***'expected'*** one, based on b and the set of all possible outcomes N. If they differ consistently, with $s_{expected} >> s_{observed}$, the problem is likely there.

Fortunately, we have (at least) one solution: the modified frequentist CL_s method

Modified frequentist CL_s method

... or, following authors' words, 'frequentist-motivated' CL_s method

«Excluding zero signal tends to say more about the probability of observing a similar or stronger exclusion in future experiments with the same expected signal and background than about the non-exixtence of the signal itself» (A. Read)

We are way more interested in statements about existence/non-existence of the signal rather than obtaining results sensitive to fluctuations of the background above a (hypothetical) signal.

The idea is to normalize the confidence level CL_{s+b} to the confidence level CL_b observed for the background-only hypothesis.

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1 - p_b}$$

and reject the hypothesis if $CL_s \leq \alpha$.

Modified frequentist CL_s method

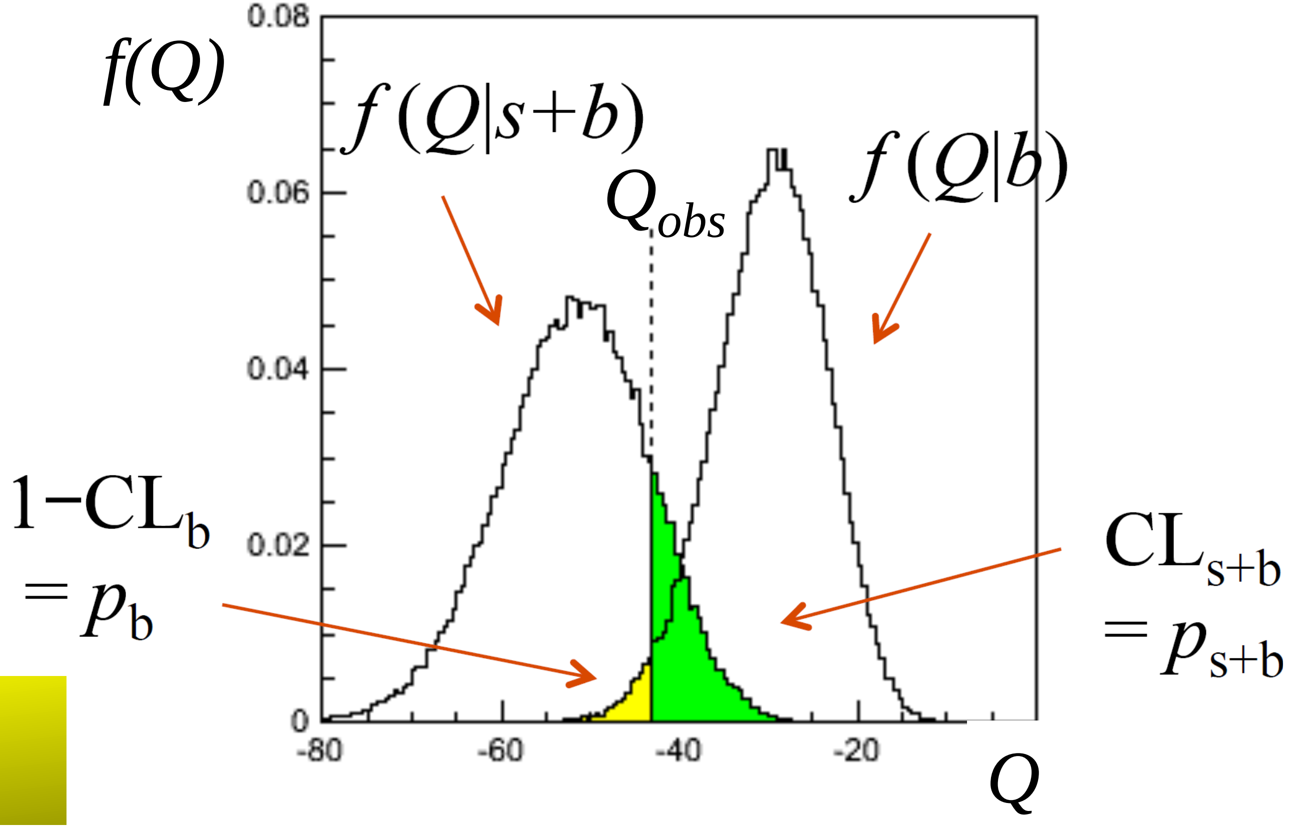
Formulated by A.Read *et al* ≈< 2000, at the time of the Higgs search at LEP

CL_{s+b} and CL_b are pure frequentist probabilities, but not their ratio (hence '**modified**')

Based on the distributions of:

$$Q(s+b) = -2 \log L(s+b)$$

$$Q(b) = -2 \log L(b)$$



Well separated distributions → 1-CL_b small → CL_b≈1
$CL_s \sim CL_{s+b}$ the ordinary p-value of $s+b$ hypothesis

Modified frequentist CL_s method

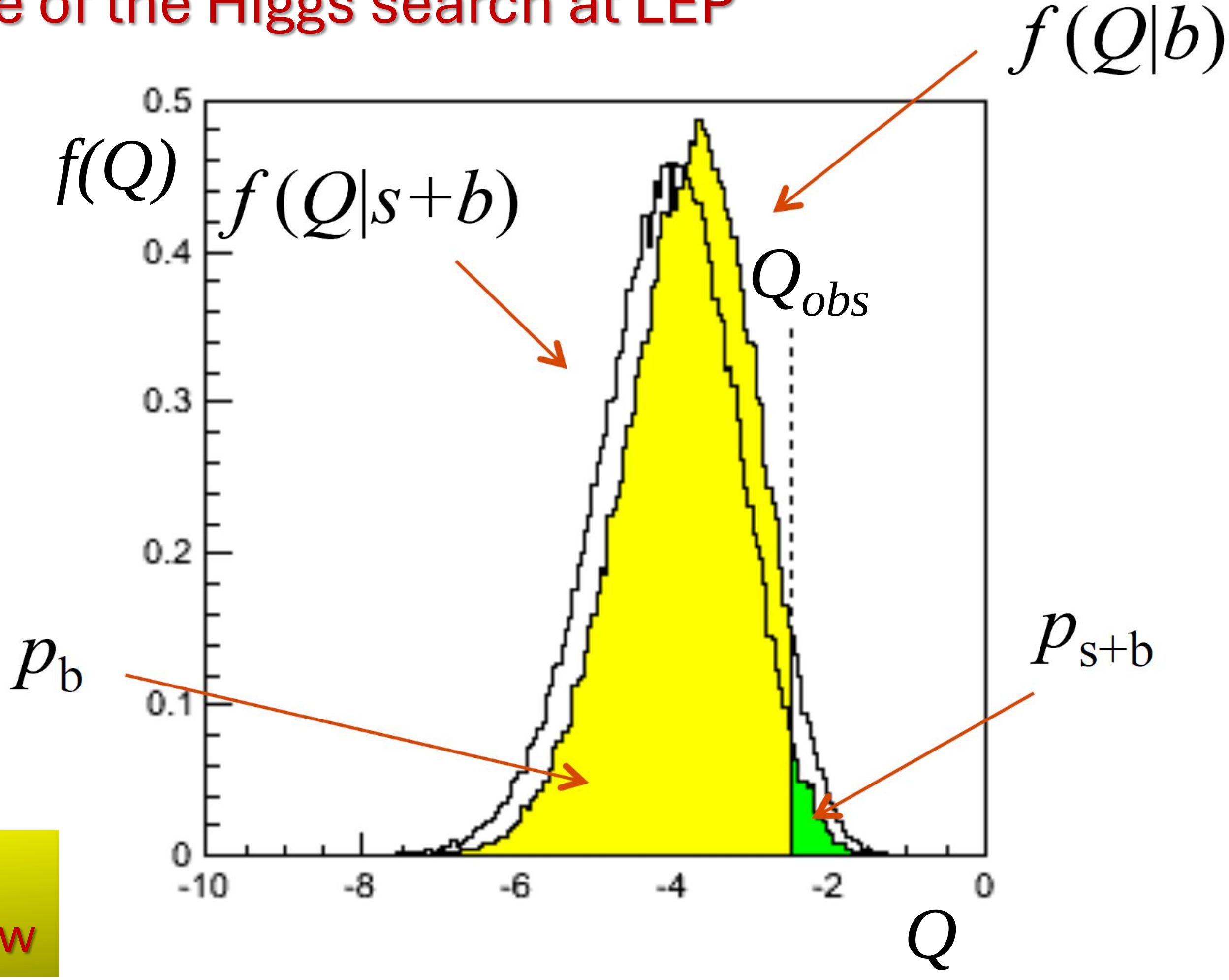
Formulated by A.Read *et al* ≈< 2000, at the time of the Higgs search at LEP

CL_{s+b} and CL_b are pure frequentist probabilities, but not their ratio (hence '**modified**')

Based on the distributions of:

$$Q(s+b) = -2 \log L(s+b)$$

$$Q(b) = -2 \log L(b)$$



Close distributions → 1-CL_b large → CL_b small
Prevents small CL_s, avoiding exclusion if sensitivity is low

The price to pay is that results are more conservative

Modified frequentist CL_s method

Formulated by A.Read *et al* ≈< 2000, at the time of the Higgs search at LEP

One of the original motivations for CL_s was to identify a generalization of frequentist upper limits for counting experiments that corresponded to the Bayesian result with a flat prior.

It shouldn't come as a surprise that for fixed b one gets an identical result to the Bayesian case (Didn't we say that Bayesian is more conservative?)

$$1 - CL = e^{-s_{up}} \frac{\sum_{m=0}^{N_{obs}} \frac{(s^{up}+b)^m}{m!}}{\sum_{m=0}^{N_{obs}} \frac{b^m}{m!}}$$

This is reassuring in both directions

- Frequentist CL_s → Bayesian
- Bayesian (flat prior) → frequentist

notably, in the low-statistics (difficult) case

What about coverage?

- Being a modified frequentist method, CL_s does not fully guarantee the coverage.
- In particular, it is known to lead to over-coverage.
- Obtained limits are anyway conventionally declared at the nominal CL.

Modified frequentist CL_s method

Based on the distributions of:

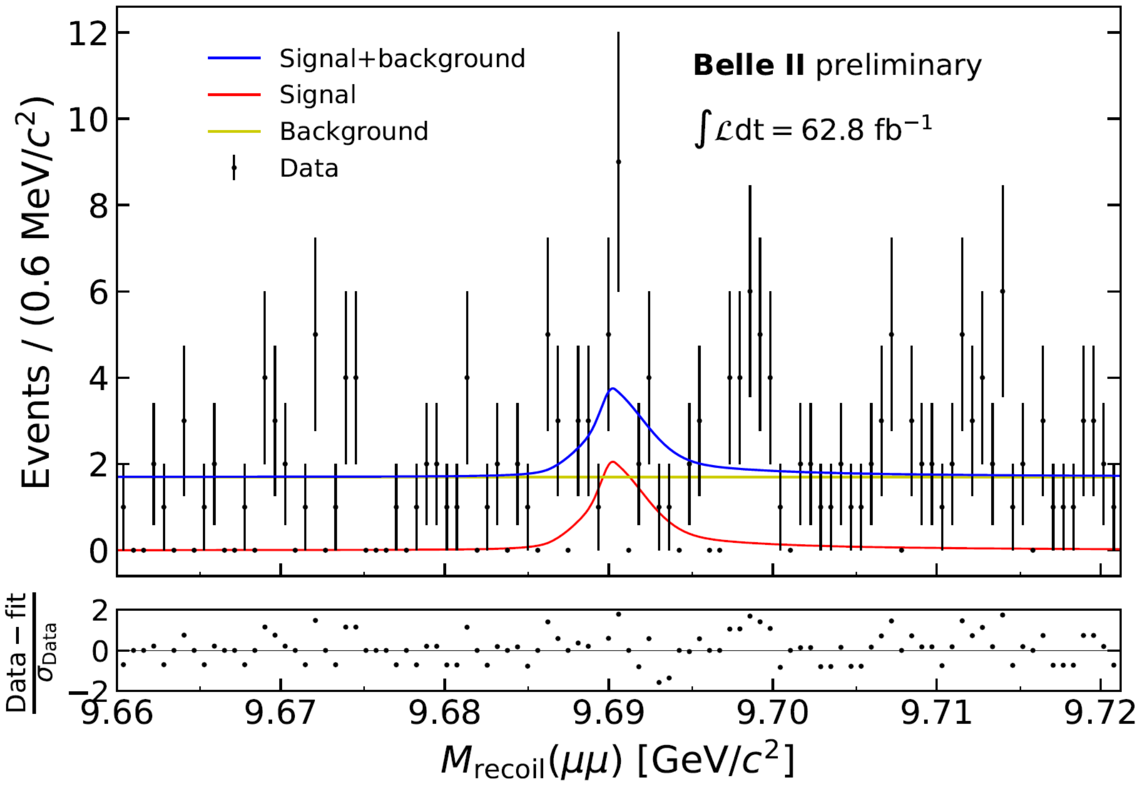
$$Q(s+b) = -2 \log L(s+b)$$

$$Q(b) = -2 \log L(b)$$

allows an immediate generalization beyond the case of the pure counting experiment

- Background fitted directly in data assuming smoothness
- Signal searched as an excess over background through fitting

$Z' \to \tau\tau$ search

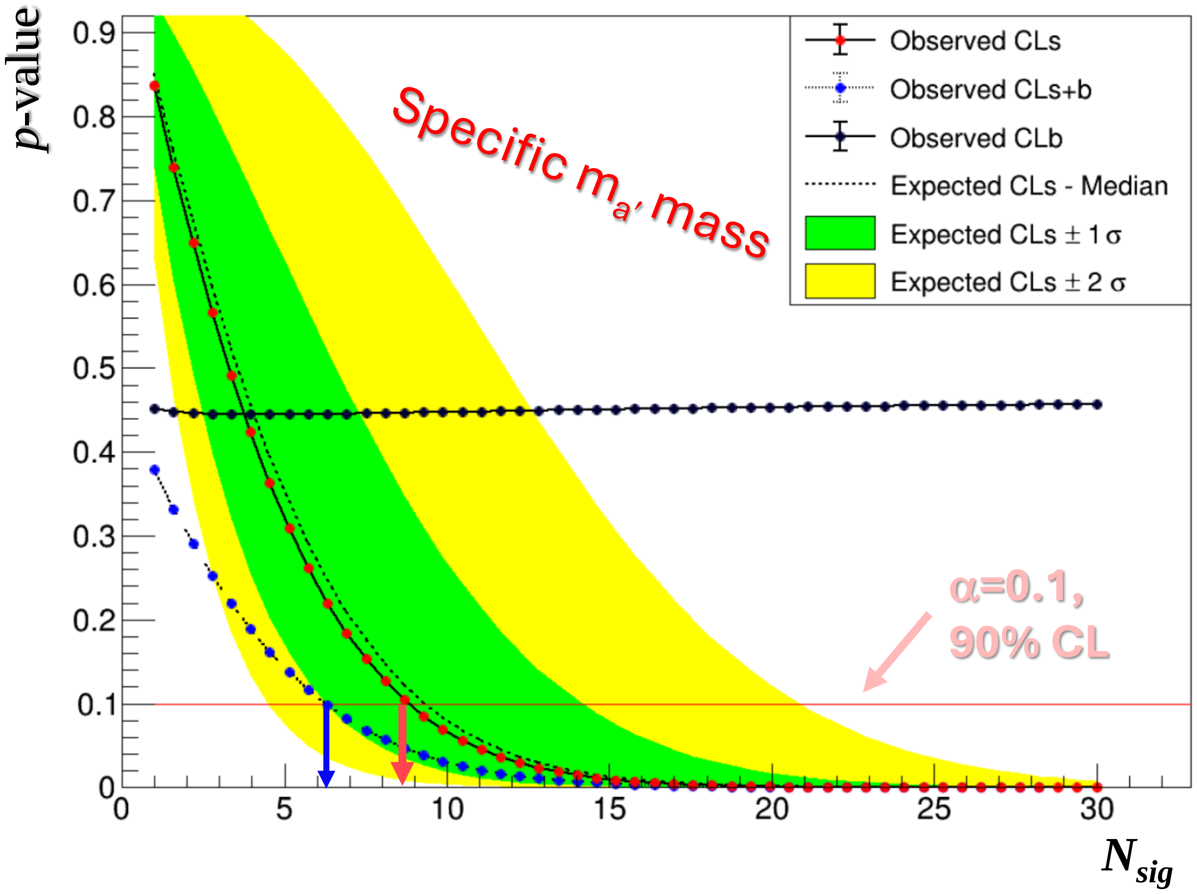


Typically maximum-likelihood fits

Modified frequentist CL_s method

How to read a CL_s limit: a recent Belle example

ALP search in $B \to K^{(*)} a' (a' \to \gamma\gamma)$

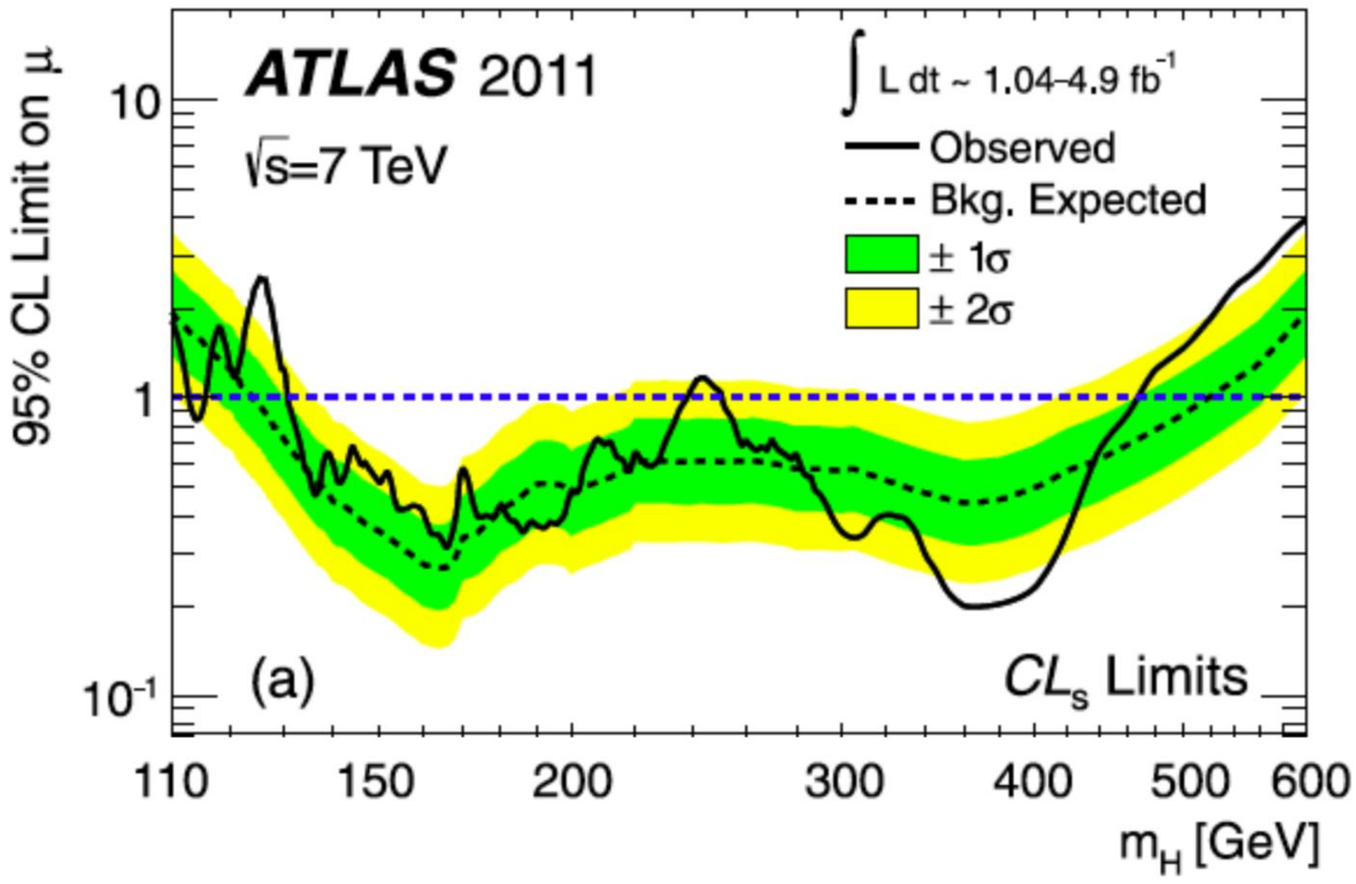


In addition to the observed results, the median and the $\pm 1\sigma$ and $\pm 2\sigma$ values expected for the b only hypothesis are also shown.

$N_{sig} < 9.28$ @ 90%CL

Modified frequentist CL_s method

Possibly the most famous exclusion ever.



For each m_H find the CLs upper limit on μ

Add median and $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands for μ=0 hypothesis

Nuisance parameters: frequentist case

Differently from the Bayesian case, where the treatment of the nuisance parameters is very clearly defined (marginalization, with only computing issues, if any) **there is no general and crystal clear approach to do it in the frequentist case.**

Nuisance parameters ⇔ systematics

Ideally, one would like to add the effect of the nuisance parameters to the likelihood model and proceed. But the model becomes more complex and high dimensional.

Two main approaches:

- Hybrid frequentist – Bayesian
- Likelihood profile

Nuisance parameters: frequentist case

Hybrid frequentist – Bayesian approach (Cousins – Highlands)

Marginalize the likelihood integrating over (all ,or part of) the nuisance parameters θ and then use it in a frequentist way

$$L_{hybrid}(s,b) = \int L(s,b,\theta)\pi(\theta)d\theta$$

The obvious draw back is that a marginalized hybrid likelihood is no longer a ‘real’ likelihood in the frequentist meaning, since θ would not change if we repeated the experiment

Anyway, in many cases numerical studies with toys show that this approach gives very similar results. These checks should always be done to validate the method in specific applications.
For example: for a p-value in the b-only hypothesis from a marginalized likelihood, one should check that is flat-distributed for the background.

Nuisance parameters: frequentist case

Likelihood profile

Replace the likelihood with a 'profiled' likelihood, using the values of the nuisance parameters that maximize $L(s, \theta)$ for each s, and then use the profiled likelihood as much as the original likelihood.

Reduces the dimensionality of the problem. Again, the profiled-likelihood is not a true likelihood, but turns out to be a very good approximation in many cases.

Achieved through the profile likelihood ratio normalized to the value of the likelihood at its maximum, i.e. with the values estimated from a ML fit

$$\lambda(s) = \frac{L(s, \hat{\hat{\theta}}(s))}{L(\hat{s}, \hat{\theta})}$$

← Fix s, fit θ (numerator)

← Fit both s and θ (denominator)

and then study $Q=-2 \log \lambda(s)$ distributions $f(Q)$.

Nuisance parameters: frequentist case

Likelihood profile

p-value calculations for the $s+b$ and b hypotheses require hard integrations. Two approaches available:

- For sufficiently large data sample, the Q distributions $f(Q)$ are asymptotically known through Wilk's theorem and independent on the nuisance parameters: integrals can be performed directly.
  - Actual distributions are distorted wrt χ^2
  - Have a look at Eur.Phys.J.C71:1554,2011 (arXiv) for more details (parametrized as a function of μ strength factor): more complicated Q test statistics are often used.

- Alternatively, they can be evaluated with ensemble of pseudo-experiments, so called toy Montecarlo experiments (or shortly **toys**), randomizing the involved global observables, including those associated to nuisance parameters.
  - Asymptotic functions not assumed. But, knowing that they are approximately independent on θ, allows not to compute p-values for all θ.
  - Have a look also at D.Tonelli's lecture some PWs ago for more details

CL_s with nuisance parameters

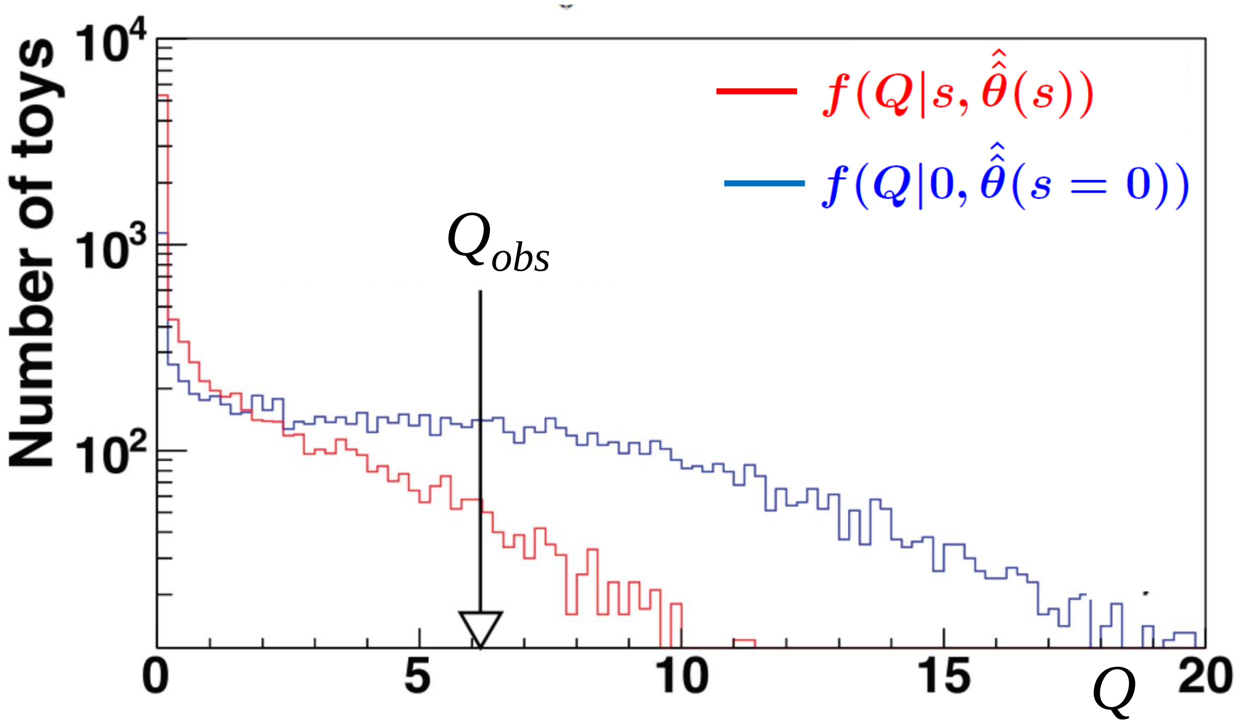
CL_s with asymptotic method →

$$p_{s+b} = \int_{Q_{obs}}^{\infty} f(Q|s, \hat{\hat{\theta}}(s))dQ$$

$$1 - p_b = \int_{Q_{obs}}^{\infty} f(Q|0, \hat{\hat{\theta}}(s=0))dQ$$

$$CL_s = \frac{p_{s+b}}{1 - p_b}$$

CL_s with Montecarlo toys →

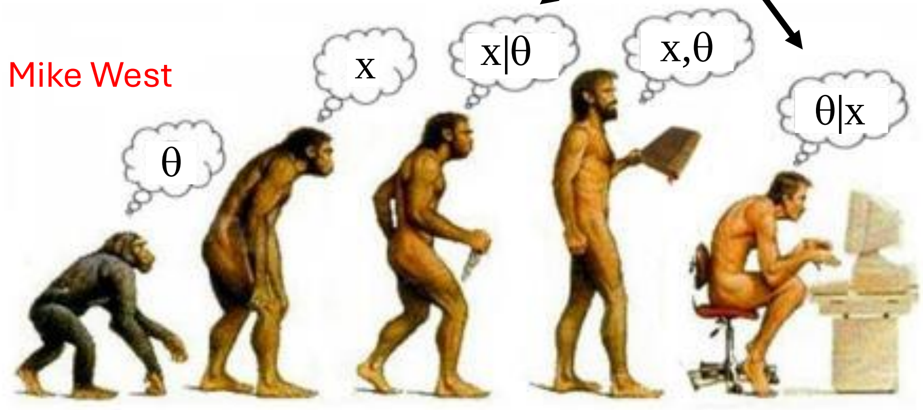


Both approaches implemented in HistFactory.
Available in Roostat/Roofit and pyhf.

Limit setting: summary

Hope that you have now sufficient information to understand where to locate yourself.

@ Mike West



Homo Apriorius | Homo Pragmaticus | Homo Frequentistus | Homo Sapiens | Homo Bayesianus

Location can change, depending on the analysis/problem. Adaptability is what Homo Sapiens used to survive until our days.

More clear ideas from the dedicated pyhf tutorial.

Thursday 17 October

Pyhf. (setup, simple model, fit) *Giordon Stark*
Kobayashi Hall, KEK 15:30 - 16:00

Frequentist inference: CLs limit setting with pyhf
Slavomira Stefkova

Bayesian inference with pyhf *Lorenz Ennio Gaertner*
Kobayashi Hall, KEK 16:30 - 17:00