Limit setting and statistics

Jory Sonneveld

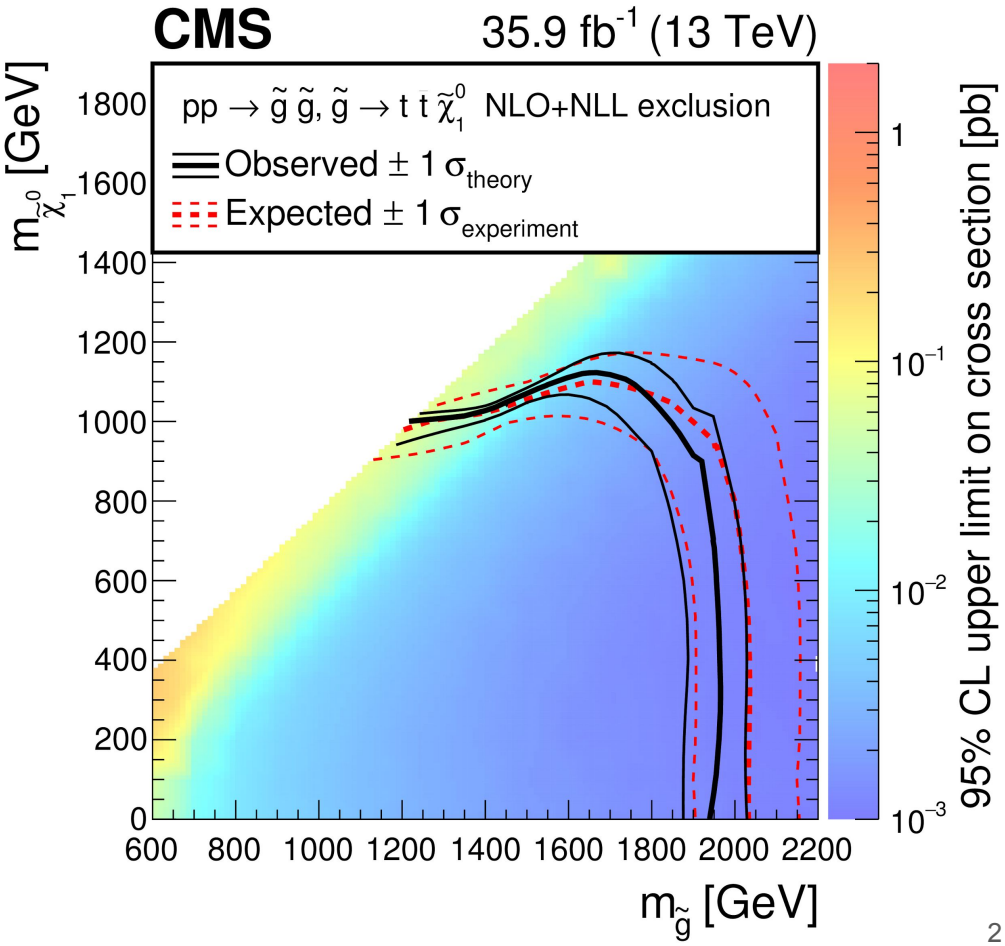
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

jory.sonneveld@cern.ch

An example of a “limit” in high energy physics

From CMS-SUS-16-033

“Limit” on gluino+LSP mass

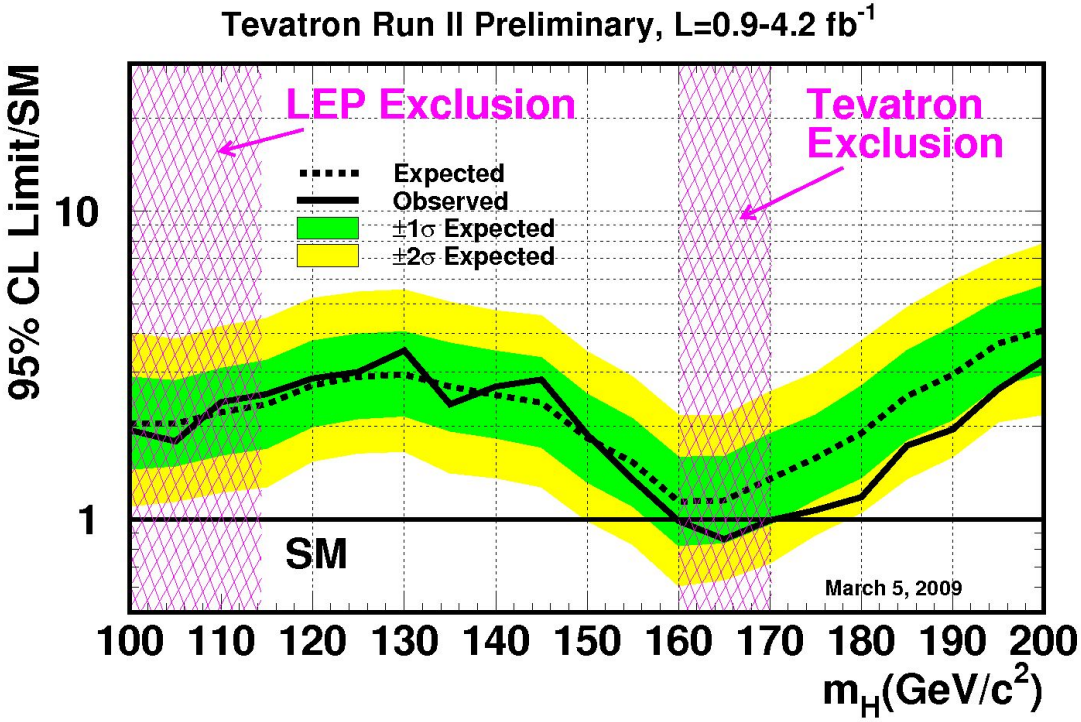


Another example of a “limit” in HEP

Limit on the Higgs cross section in terms of the Standard Model expectation

CL?

→ **confidence level**



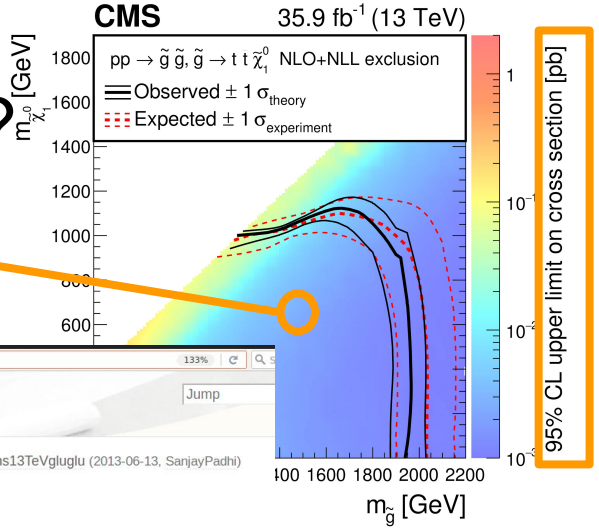
https://www-cdf.fnal.gov/physics/new/hdg/Plain_English_files/higgs-limits-march2009.gif

Where does this "limit" come from?

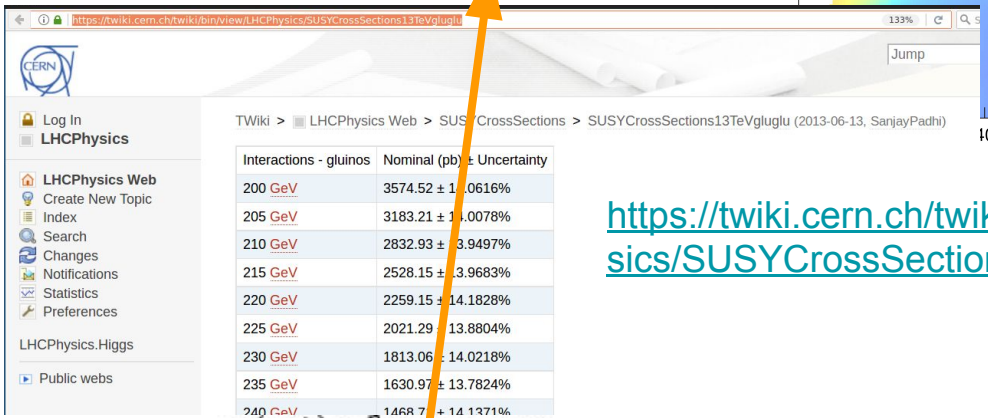
A contour line:

$$\frac{\sigma^{\text{theo}}_{\text{pred}}}{\sigma^{\text{UL}}_{\text{obs}}} = 1$$

$\sigma^{\text{UL}}_{\text{obs}}$



A theory prediction:



https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections13TeVgluglu

$$\mathrm{CL}_{s+b} = P_{s+b}(X \le X_{\text{obs}}) = \frac{e^{-(s+b)} \sum_{n=0}^{n_{\text{obs}}} (s+b)^n}{n!}.$$

An upper limit: $\mathrm{CL}_s = \mathrm{CL}_{s+b} / \mathrm{CL}_b$

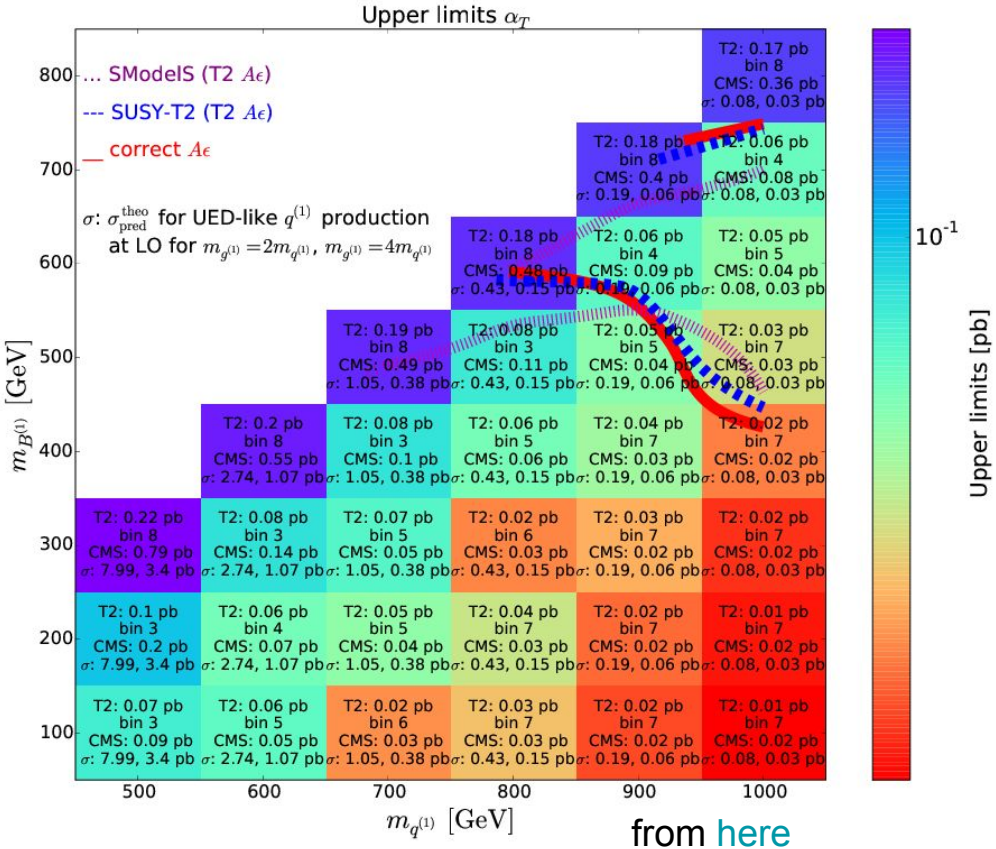
See T. Junk
http://arxiv.org/abs/hep-ex/9902006

$$\mathrm{CL}_{s+b} = P_{s+b}(X \le X_{\text{obs}}) = \frac{e^{-(s+b)} \sum_{n=0}^{n_{\text{obs}}} (s+b)^n}{n!}.$$

The contour line

For each mass point:

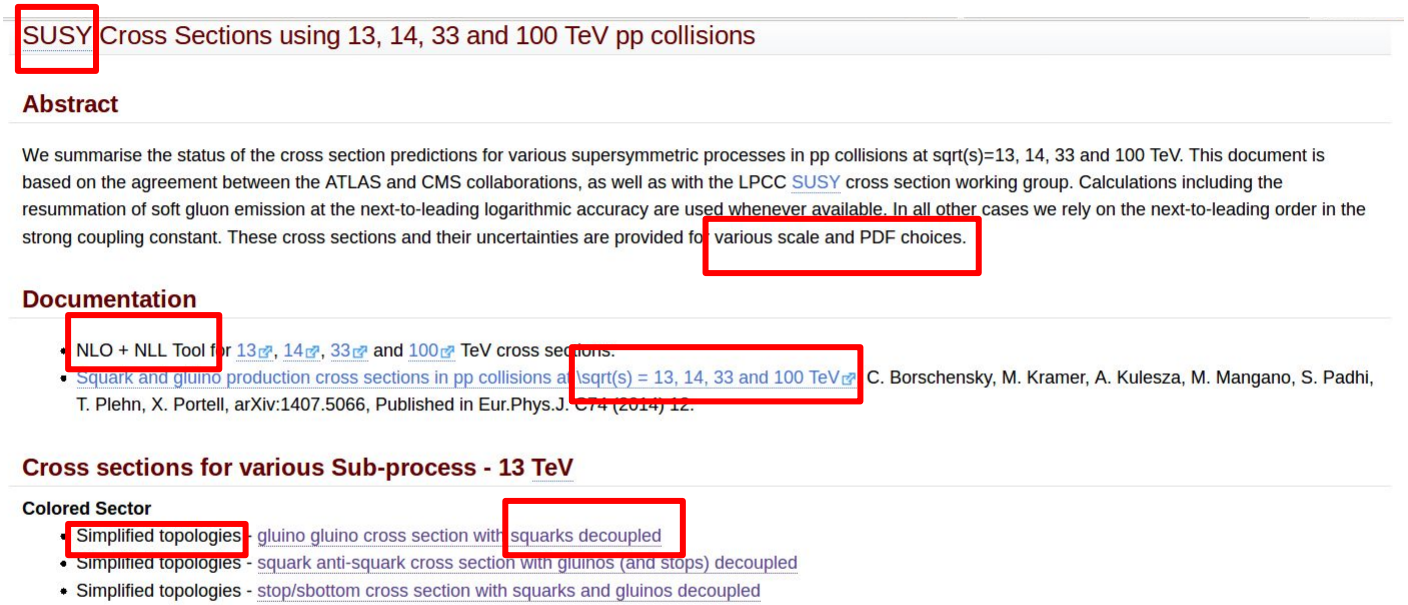
- An **upper limit** from CMS
- An **upper limit** from MadAnalysis (→ color)
- A **theory prediction** for a cross section a KK gluon mass of twice the KK quark mass
- A **theory prediction** for a cross section for a KK gluon mass of four times the KK quark mass
- The **bin** for which the upper limit was computed



from here

The theory prediction

https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections#SUSY_Cross_Sections_using_13_14



SUSY Cross Sections using 13, 14, 33 and 100 TeV pp collisions

Abstract

We summarise the status of the cross section predictions for various supersymmetric processes in pp collisions at sqrt(s)=13, 14, 33 and 100 TeV. This document is based on the agreement between the ATLAS and CMS collaborations, as well as with the LPCC SUSY cross section working group. Calculations including the resummation of soft gluon emission at the next-to-leading logarithmic accuracy are used whenever available. In all other cases we rely on the next-to-leading order in the strong coupling constant. These cross sections and their uncertainties are provided for various scale and PDF choices.

Documentation

- NLO + NLL Tool for 13, 14, 33 and 100 TeV cross sections.
- Squark and gluino production cross sections in pp collisions at \sqrt(s) = 13, 14, 33 and 100 TeV, C. Borschensky, M. Kramer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn, X. Portell, arXiv:1407.5066, Published in Eur.Phys.J. C74 (2014) 12.

Cross sections for various Sub-process - 13 TeV

Colored Sector

- Simplified topologies - gluino gluino cross section with squarks decoupled
- Simplified topologies - squark anti-squark cross section with gluinos (and stops) decoupled
- Simplified topologies - stop/sbottom cross section with squarks and gluinos decoupled

Depends on:

- Model
- PDFs
- Tool (Pythia, NLLFast)
- Masses of other particles
- ...

The upper limit

See T. Junk
http://arxiv.org/abs/hep-ex/9902006

$$\mathrm{CL_{s+b}} = P_{s+b}(X \leq X_{\mathrm{obs}}) = \frac{\mathrm{e}^{-(s+b)} \sum_{n=0}^{n_{\mathrm{obs}}} (s+b)^n}{n!}.$$

$$\mathrm{CL_s} = \mathrm{CL_{s+b}} / \mathrm{CL_b}$$

CL: **confidence level**

X: a **test statistic** or **discriminant**

b: number of expected background events (estimated)

n_{obs}: number of observed events in data (counted)

s: varied until a certain **confidence level**, e.g. 95%, is reached → $s_{\mathrm{obs}}^{\mathrm{UL},95}$

The cross section is related to the number of events: $n_i = \sigma_i \times A\epsilon \times \mathcal{L}$, or xsec * **efficiency** * luminosity

Confidence level and upper limit

$$\mathrm{CL_{s+b}} = P_{s+b}(X \leq X_{\mathrm{obs}}) = \frac{\mathrm{e}^{-(s+b)} \sum_{n=0}^{n_{\mathrm{obs}}} (s+b)^n}{n!}.$$

$$\mathrm{CL_s} = \mathrm{CL_{s+b}} / \mathrm{CL_b}$$

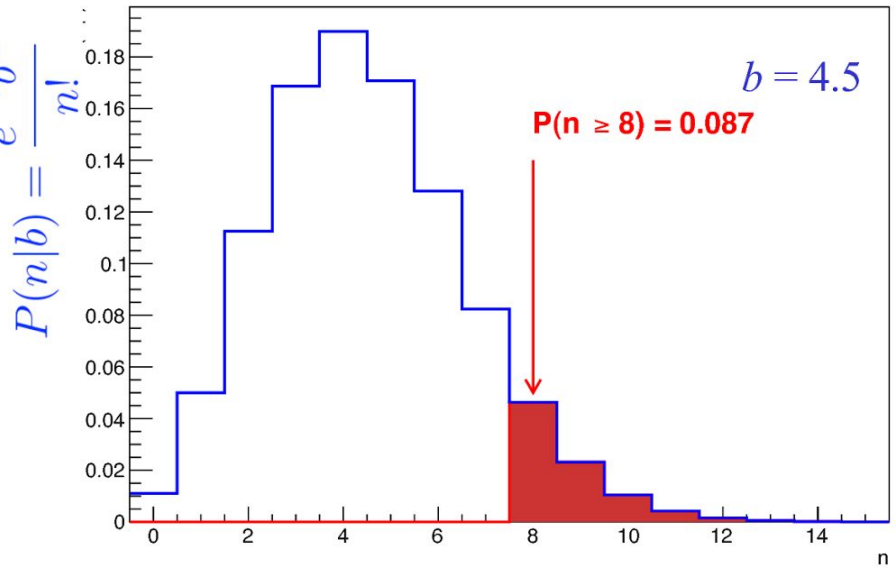
$$\mathrm{CL_b} = \mathrm{CL_{s+b}}_{(s=0)}$$

s: varied until a certain **confidence level**, e.g. 95%, is reached → $s_{\mathrm{obs}}^{\mathrm{UL},95}$

So set for example CL := 0.95 → **compute upper limit by solving for s**

A p-value

Probability that (in this case) background gives at least 8 observed events if background is correct, or an overfluctuation would give >= 8 events:



Not the probability that the background is true!

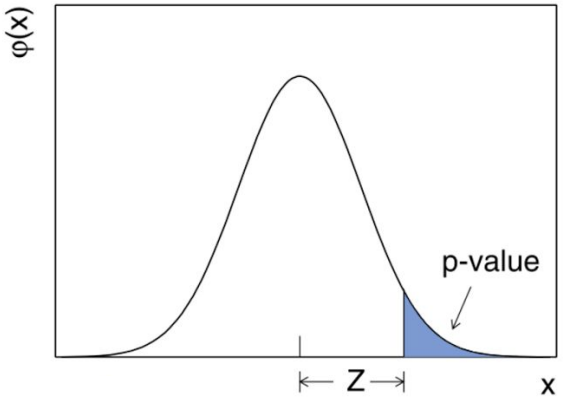
P-values and significance

$$p = \int_{n\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x = 1 - \frac{1}{2}\mathrm{erf}\left(\frac{n}{\sqrt{2}}\right)$$

or:

$$Z = n = \Phi^{-1}(1-p)$$

Φ = cumulative of a normal distribution



One can convert the pvalue to a number of standard deviations or ‘Z-significance’:

p < 0.05 (i.e.: 95% confidence level): Z = 1.64
p < 0.10 (i.e.: 90% confidence level): Z = 1.28

Do it yourself: a single-sided confidence interval

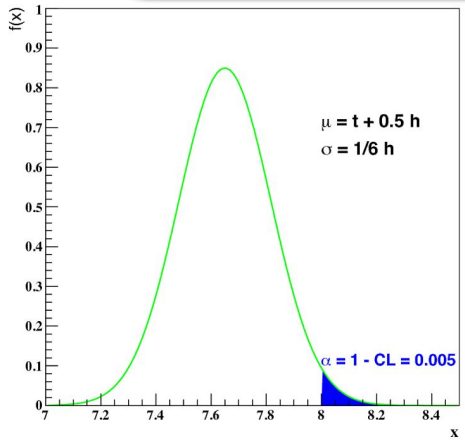
single-sided: $x < x_{max}$ at 95% CL

when do we need it? when we are lazy and want to sleep more, still not being fired for coming late at work

An employee needs to be at work at 8:00 o'clock sharp. The journey takes 30 minutes on average, with a Gaussian uncertainty of σ = 10 minutes due to traffic.

When must he leave home to be late only once a year (~0.5%)? Or to be in time in 99.5% of the cases?

in 2016 in Hamburg there are 254 working days, 30 days vacation: 1/224 ~0.0045



$$\int_{x_{arrival}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - CL$$

mean $\mu = x_{start} + 30'$

width $\sigma = 10'$

limit $x_{arrival} = 08{:}00$

x_{start} (0.5%) = ?

Try this yourself:

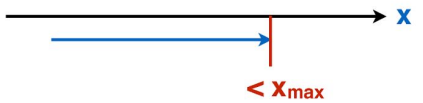
```
>>> from scipy import stats
>>> pvalue = 0.005
>>> stats.norm.interval(1 - 2*pvalue)
```

What do you get? What does it mean?

single-sided!

A confidence interval

single-sided: $x < x_{max}$ at 95% CL

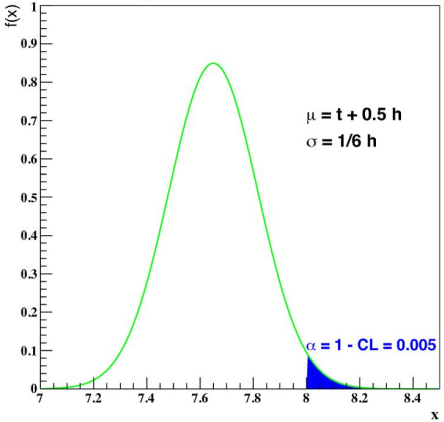


when do we need it? when we a
want to sleep more, still not being fired fo

An employee needs to be at work at 8:00 o'clock sharp. The journey takes 30 minutes on average, with a Gaussian uncertainty of σ = 10 minutes due to traffic.

When must he leave home to be late only once a year (~0.5%)? Or to be in time in 99.5% of the cases?

in 2016 in Hamburg there are 254 working days, 30 days vacation: 1/224 ~0.0045



1 σ	84.13 %
2 σ	97.72 %
2.3 σ	98.92 %
2.6 σ correspond to	**99.53 %**
2.7 σ	99.65 %
3 σ	99.86 %

x_{start} (0.5%) = 08:00 - 00:30 - 00:26 = 07:04

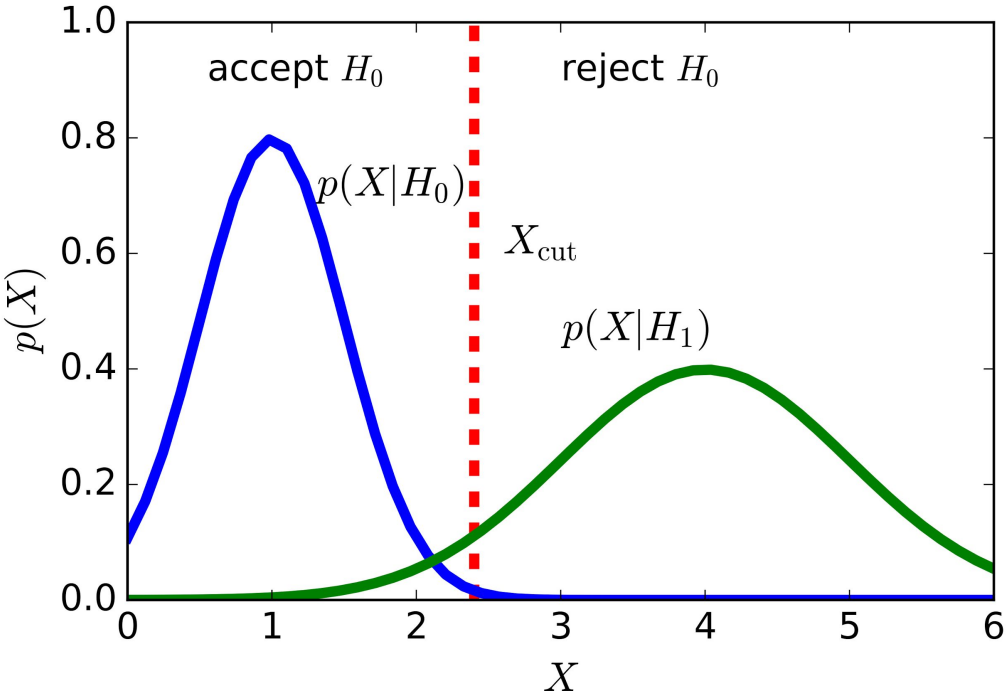
```
>>> number_of_sigma =
stats.norm.interval(1
- 2*tail)[1]
>>> number_of_sigma
2.5758293035489004
>>> minutes =
number_of_sigma*10
>>> total = minutes +
30
>>> total
55.758293035489004
```

http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

Testing hypotheses and discrimination

See also the book: Statistical Data Analysis by Glenn Cowan

Accept null hypothesis H0



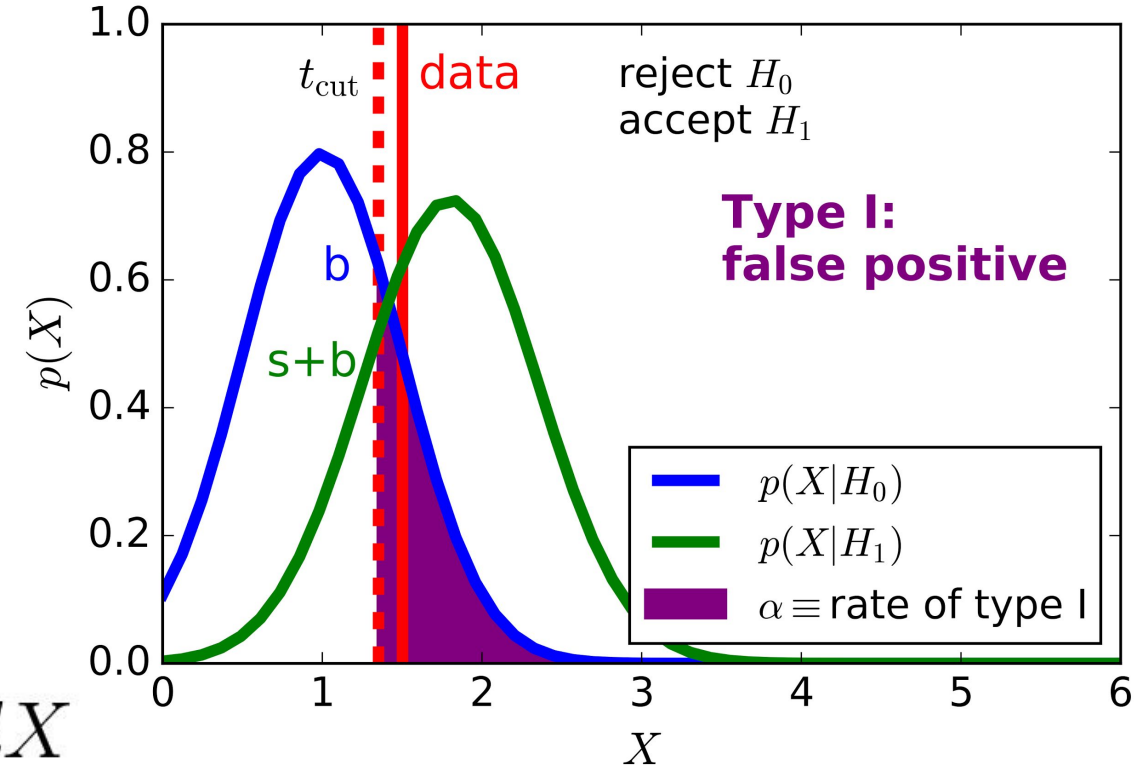
Reject null hypothesis

What can go wrong here?

Type I error: incorrect rejection of H_0

Type I: incorrect rejection of a true null hypothesis (a "false positive"): α

When 'testing for H1': True negative rate: *specificity* = 1 - α

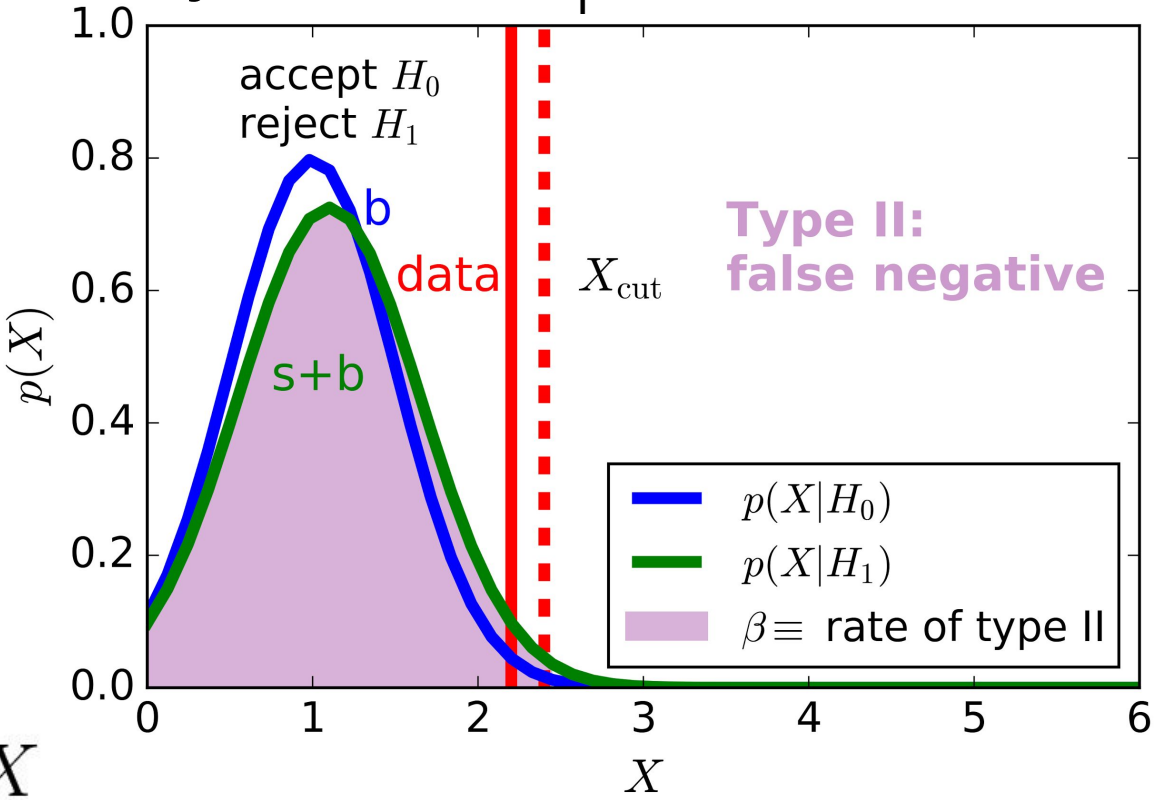


$$\alpha = \int_{X_{\mathrm{crit}}}^{\infty} p(X|H_0)dX$$

Type II error: incorrect rejection of H_1

Type II: incorrectly retaining a false null hypothesis (a "false negative"): ß

When 'testing for H1': True positive rate: *sensitivity* = 1 - ß



$$\beta = \int_{-\infty}^{X_{\text{crit}}} p(X|H_1)dX$$

Type I and Type II errors

Testing a hypothesis:

H_0 null hypothesis: *user is logging in*

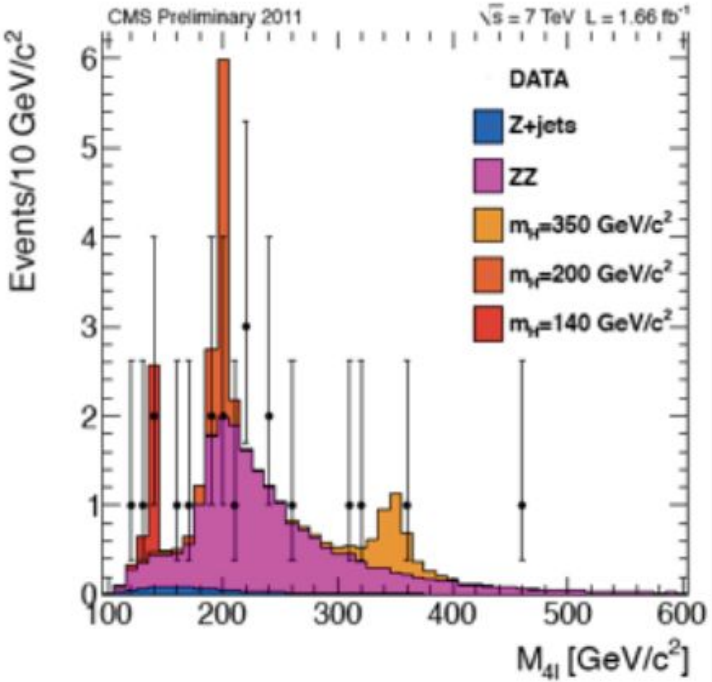
H_1 alternative hypothesis: *imposter is logging in*

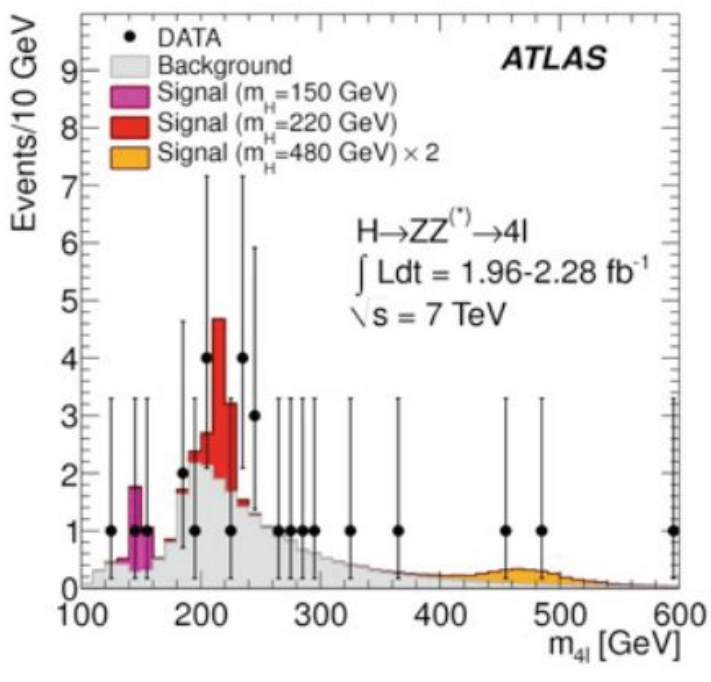
Imposter src

I. *classify authorized users as imposters*
   incorrect rejection of a true null hypothesis (a "false positive"): α
II. *classify imposters as authorized users*
   incorrectly retaining a false null hypothesis (a "false negative"): ß

True positive rate: *sensitivity* = 1 - ß
True negative rate: *specificity* = 1 - α

m(Higgs) = 220 GeV?





How can we easily recognize a ‘fake’ signal? And why?

https://profmattstrassler.files.wordpress.com/2011/12/hcp_h2zz1.png
https://profmattstrassler.com/articles-and-posts/the-higgs-particle/the-standard-model-higgs/seeking-and-studying-the-standard-model-higgs-particle/

Reliability of a test

You are given a test Findzvirez that tests for virez. It is written on the package that it is ‘99% reliable’.

What is the probability that you have the disease virez if your test comes out negative?

What when it comes out positive?

‘Reliable’: p(+|has virez) = 0.99

p(- | has virez) = 0.01 = false negative rate

p(has virez | +) = ?

p(has virez | -) = ?

Sensitivity and specificity of a test

You are given a test Findzvirez that tests for virez. It is written on the package that it is ‘99% reliable’.

You do some research and find that the test is 99% sensitive and 92% specific.

What is the probability that you have virez when the test comes out positive?
p(has virez | +) = ?

True positive rate: *sensitivity* = 1 - ß
True negative rate: *specificity* = 1 - α

p(- | no virez) = 0.92
p(+| no virez) = 0.08 = false positive rate
p(+) = p(+ | no virez)p(no virez) + p(+ | has virez)p(has virez)

Probabilities and Bayes' theorem

Solve this using **Bayes' theorem**: **the probability of an event based on prior knowledge of conditions that may be related to the event**

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

Many false positives: airport security screening

US breast cancer mammography: false positive rate 15%
→ After 10 years of screening, half of the American population has one false positive

We now have all information

You are given a test Findzvirez that tests for virez that is ‘99% reliable’.

You did some research and find that the test is 99% sensitive and 92% specific.

You did more research and find that in general 0.5% of the entire world population actually carries virez.

What is the probability that you have virez when the test comes out positive?

p(- | no virez) = 0.92

p(+| no virez) = 0.08 = false positive rate

p(+| has virez) = 0.99

p(- | has virez) = 0.01 = false negative rate

p(has virez) = 0.005

p(+) = p(+ | no virez)p(no virez) + p(+ | has virez)p(has virez)

p(has virez | +) = ?

No need to worry on testing positive

What is the probability that you have virez when the test comes out positive?

p(+| no virez) = 0.08 = false positive rate

p(+| has virez) = 0.99

p(has virez) = 0.005

p(+) = p(+ | no virez)p(no virez) + p(+ | has virez)p(has virez)

p(has virez | +) = p(+ | has virez)p(has virez) / p(+)

= 0.99*0.005 / (0.08*0.995 + 0.99*0.005)

= 0.0585452395

Don’t worry, but do go for your next (much better) test =)

The upper limit

See T. Junk
http://arxiv.org/abs/hep-ex/9902006

$$\mathrm{CL_{s+b}} = P_{s+b}(X \leq X_{\mathrm{obs}}) = \frac{\mathrm{e}^{-(s+b)} \sum_{n=0}^{n_{\mathrm{obs}}} (s+b)^n}{n!}.$$

$$\mathrm{CL_s} = \mathrm{CL_{s+b}} / \mathrm{CL_b}$$

CL: **confidence level**

X: a **test statistic** or **discriminant**

b: number of expected background events (estimated)

n_{obs}: number of observed events in data (counted)

s: varied until a certain **confidence level**, e.g. 95%, is reached → $s_{\mathrm{obs}}^{\mathrm{UL},95}$

The cross section is related to the number of events: $n_i = \sigma_i \times A\epsilon \times \mathcal{L}$, or xsec * **efficiency** * luminosity

Likelihood

Likelihood ratio:

$$\frac{p(X|H_0)}{p(X|H_1)}$$

Likelihood of a theory given the data:

L(theory, data) = p(data | theory)

Choose **test statistic** X to be able to discriminate well among hypotheses/theories:

$$X = \frac{L(\theta)}{L_{\max}(\hat{\theta})}$$

- the denominator the maximum likelihood.
- θ the theory parameters, θ

Example of a test statistic: likelihood ratio

$$Q(n_{\mathrm{obs}}) = -2\log\frac{L(\lambda_s, n_{\mathrm{obs}}, n_{\mathrm{b}}, \delta_{\mathrm{b}})}{L_{\mathrm{max}}(\hat{\lambda}_s, n_{\mathrm{obs}}, n_{\mathrm{b}}, \delta_{\mathrm{b}})}$$

~χ^2

Likelihood

General likelihood:

$L(\mathrm{data} \mid \theta) = \mathrm{Poisson}(\mathrm{data} \mid s(\theta) + b(\theta)) \cdot p(\tilde{\theta} \mid \theta)$

Poisson distribution (counting experiment)
Can be used to generate pseudodata

Smearing from nuisance parameters: Gaussian, lognormal pdf

Likelihood methods

nuisance parameters: unknown parameters not interesting for the measurement, like detector resolution, uncertainty in backgrounds, background shape modeling, other systematic uncertainties

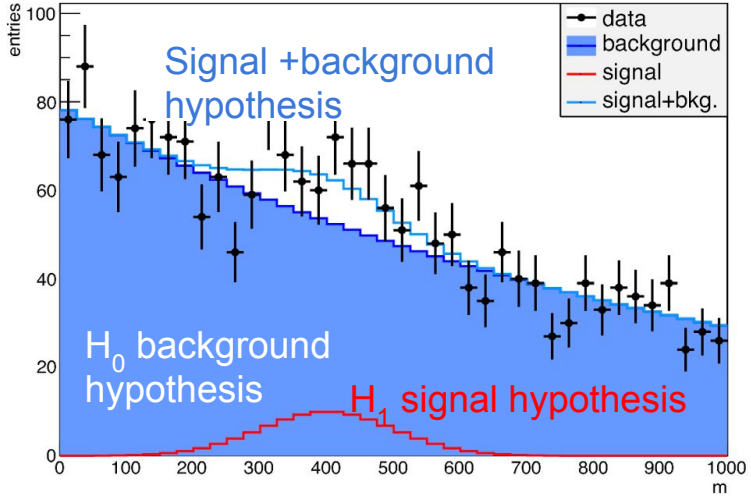
Frequentist: constrain likelihood nuisance parameters with e.g. control sample:

$$L(\text{data, control sample} \mid \theta) = L(\text{ data} \mid \theta, s, b)\, L(\text{control sample} \mid \theta)$$

Bayesian: integrate analytically, numerically, with Markov Chain Monte Carlo

Hybrid: integrate or ***marginalize*** the likelihood function over the nuisance parameters, then frequentist method

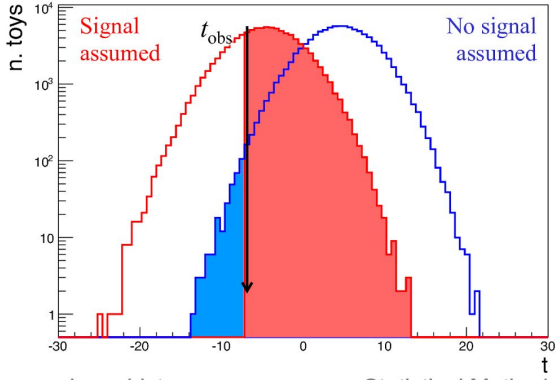
Frequentist method: toy experiments



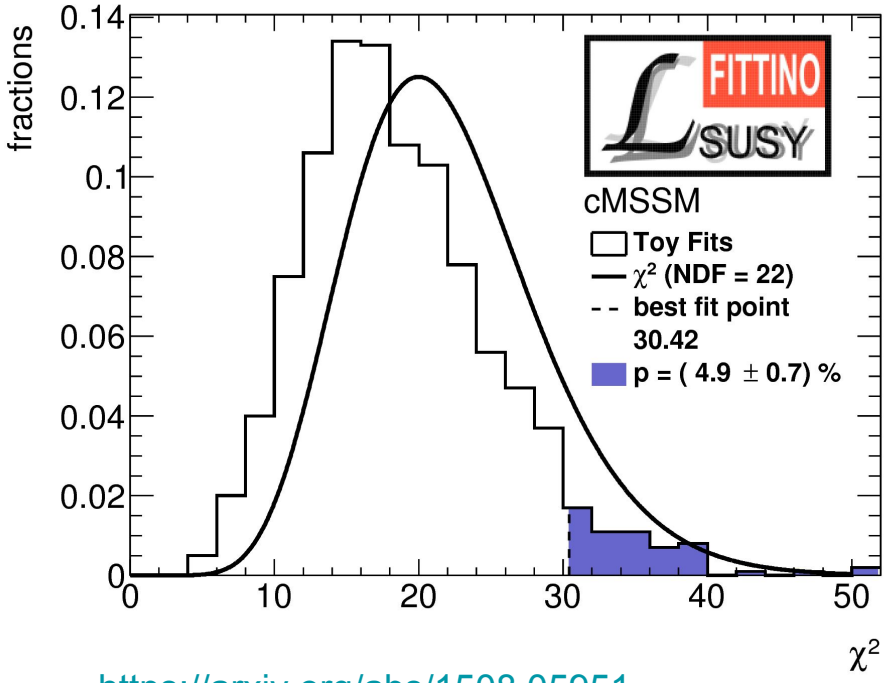
P-value: fraction of toys > t_{obs}

$p = 374/10^5 = 3.7 \pm 0.2\%$

$Z = 2.7$



Example: Fittino toy 'fits' to obtain χ^2 distribution



https://arxiv.org/abs/1508.05951
Eur.Phys.J. C76 (2016) no.2, 96

Bayesian limits

Signal s(θ, x) and background b(x)

Data or *pseudodata*

Nuisance parameters x

Posterior density

$$L(\theta) = \frac{1}{C} \int_x p(\mathrm{data}|\mathrm{s}+\mathrm{b}) \rho_x(x) \pi_s(s) dx$$

Normalization constant

Sometimes: μs with *signal strength modifier* μ

Systematic uncertainties

Prior: usually flat, 0 for s <= 0

$$\int_0^{s^{ul,95\%\mathrm{C.L.}}} L(\theta) d(\theta) := 0.95$$

95% CL upper limit on amount of signal events

Bayesian priors in particle physics

Theory: Allzymmetric(a, other parameters).

Experiment: observable A can only take the values 0 <= A <= 20.
You compute that for Allzymmetric:

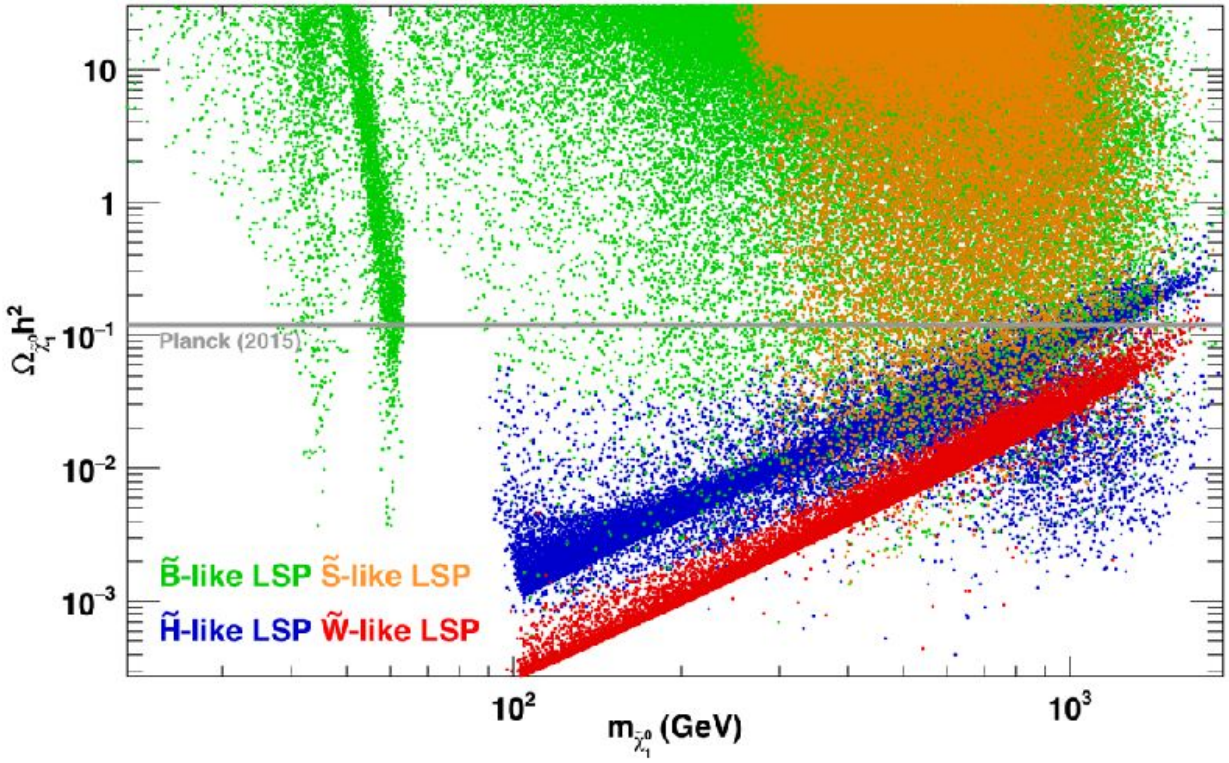
$$0 \leq A \leq 20 \text{ would imply } 0 \leq a \leq 2000.$$

You choose a **prior $\pi(\theta)$** with **θ** your Allzymmetric theory parameters such that

$$\pi(\theta) = 1, \quad 0 \leq a \leq 2000;$$
$$\pi(\theta) = 0 \text{ otherwise.}$$

For example, you do not look at any SUSY parameters that predict $\Omega_{DM} > 0.12$
(see Table 3 of https://arxiv.org/pdf/1502.01589v3.pdf from the Planck results)

Example of a possible prior



From this information one could choose a prior (note: extreme example):
0 if bino-like LSP
1 if wino-, higgsino-like LSP

From Table 3 of https://arxiv.org/pdf/1502.01589v3.pdf

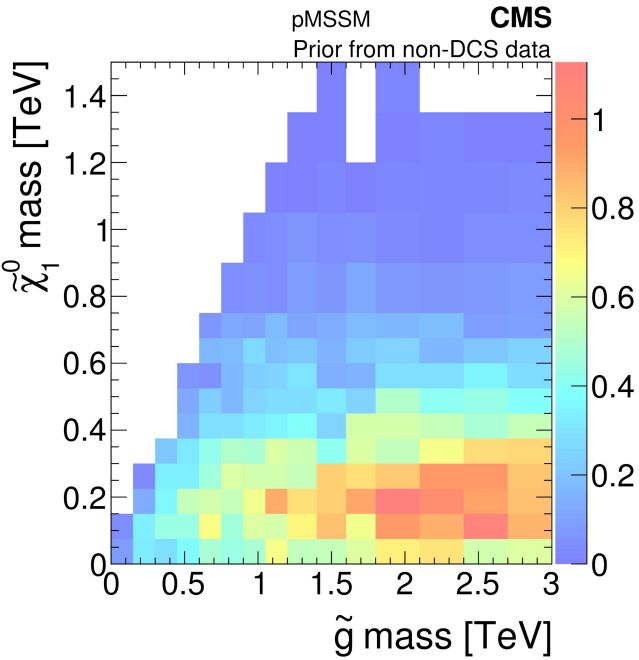
Planck Collaboration: Cosmological paramete

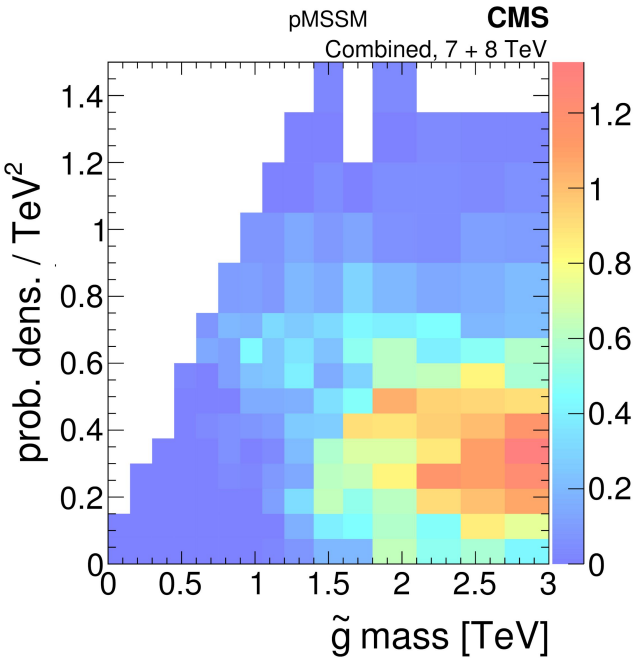
Table 3. Parameters of the base ΛCDM cosmology computed from the 2015 baseline
of parameters determined from the temperature and polarization spectra at high multi
and high multipoles and is the same as column [6] of Table 1. Columns [2] and
multipoles, and only polarization at low multipoles. Column [4] uses the full likeliho
cosmological parameters determined from the *Planck* TT+lowP and *Planck* TT,TE,E

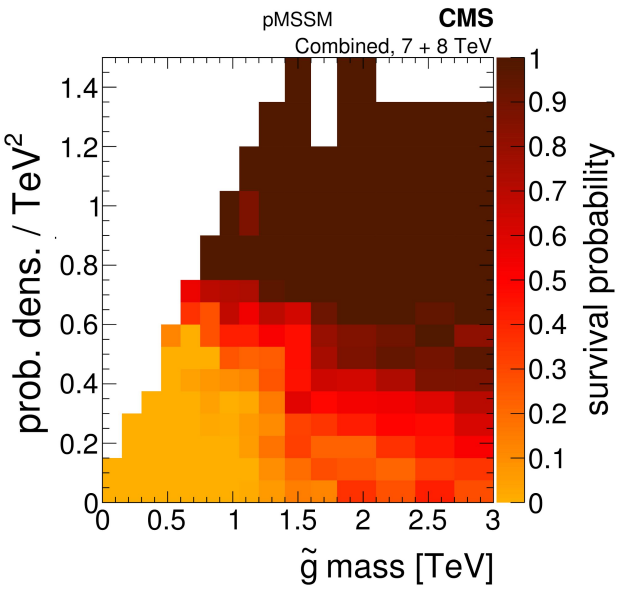
Parameter	[1] *Planck* TT+lowP	[2] *Planck* TE+lowP	[3] *Planck* EE+lowP
$\Omega_b h^2$	0.02222 ± 0.00023	0.02228 ± 0.00025	0.0240 ± 0.0013
$\Omega_c h^2$	0.1197 ± 0.0022	0.1187 ± 0.0021	$0.1150^{+0.0048}_{-0.0055}$
$100\theta_{MC}$	1.04085 ± 0.00047	1.04094 ± 0.00051	1.03988 ± 0.00094
τ	0.078 ± 0.019	0.053 ± 0.019	$0.059^{+0.022}_{-0.019}$
$\ln(10^{10}A_s)$	3.089 ± 0.036	3.031 ± 0.041	$3.066^{+0.046}_{-0.041}$
n_s	0.9655 ± 0.0062	0.965 ± 0.012	0.973 ± 0.016
H_0	67.31 ± 0.96	67.73 ± 0.92	70.2 ± 3.0
Ω_m	0.315 ± 0.013	0.300 ± 0.012	$0.286^{+0.027}_{-0.038}$
σ_8	0.829 ± 0.014	0.802 ± 0.018	0.796 ± 0.024
$10^9 A_s e^{-2\tau}$	1.880 ± 0.014	1.865 ± 0.019	1.907 ± 0.027

From Fig. 7 of https://arxiv.org/pdf/1505.06243.pdf where also the overabundant points are shown.
G. Belanger, J. Da Silva, U. Laa, A. Pukhov

Bayesian posterior density: a pMSSM example







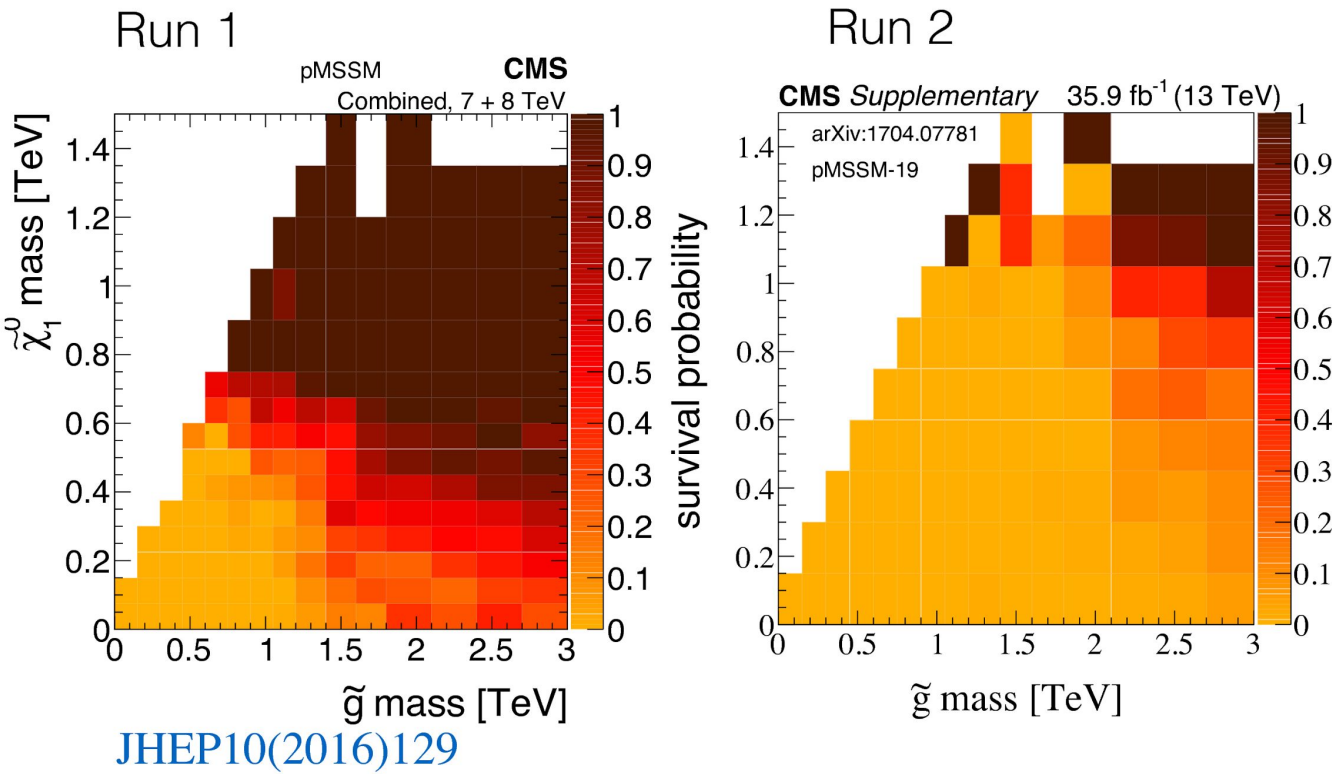
See also
JHEP 10 (2016) 129
arXiv:1606.03577

$$p(\theta|d) \propto L(d^{\text{CMS-SUSY}}|\theta)L(d^{\text{pre-CMS-SUSY}}|\theta)p(\theta)$$

Note: NO dark matter constraints in here!

http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS-15-010/index.html

Aside: these densities/probabilities went down



JHEP10(2016)129

Prior used in CMS pMSSM interpretation

Table 1: The measurements that form the basis of the non-DCS prior $p^{\text{non-DCS}}(\theta)$ for the pMSSM parameters, their observed values and likelihoods. The observables are the decay branching fractions $\mathcal{B}(\mathrm{b} \to \mathrm{s}\gamma)$ and $\mathcal{B}(\mathrm{B_s} \to \mu\mu)$, the SUSY to SM ratio for the branching fraction of the decay $\mathrm{B} \to \tau\nu$, $R(\mathrm{B} \to \tau\nu)$, the difference in the muon anomolous magnetic moment from its SM prediction Δa_μ, the strong coupling constant at the Z boson mass $\alpha_\mathrm{s}(m_Z)$, the top and bottom quark masses m_t and $m_\mathrm{b}(m_\mathrm{b})$, the Higgs boson mass m_h and signal strength μ_h, and sparticle mass limits from LEP. All data except μ_h were used in the initial MCMC scan. Details are given in the text.

i	Observable $\mu_i(\theta)$	Constraint $D_i^{\text{non-DCS}}$	Likelihood function $L[D_i^{\text{non-DCS}}\|\mu_i(\theta)]$	Comment
1	$\mathcal{B}(\mathrm{b} \to \mathrm{s}\gamma)$ [45]	$(3.43 \pm 0.21^{\text{stat}} \pm 0.24^{\text{th}} \pm 0.07^{\text{sys}}) \times 10^{-4}$	Gaussian	reweight
2	$\mathcal{B}(\mathrm{B_s} \to \mu\mu)$ [46]	$(2.9 \pm 0.7 \pm 0.29^{\text{th}}) \times 10^{-9}$	Gaussian	reweight
3	$R(\mathrm{B} \to \tau\nu)$[45]	1.04 ± 0.34	Gaussian	reweight
4	Δa_μ [47]	$(26.1 \pm 6.3^{\text{exp}} \pm 4.9^{\text{SM}} \pm 10.0^{\text{SUSY}}) \times 10^{-10}$	Gaussian	
5	$\alpha_\mathrm{s}(m_Z)$ [48]	0.1184 ± 0.0007	Gaussian	
6	m_t [49]	$173.20 \pm 0.87^{\text{stat}} \pm 1.3^{\text{sys}}$ GeV	Gaussian	reweight
7	$m_\mathrm{b}(m_\mathrm{b})$ [48]	$4.19^{+0.18}_{-0.06}$ GeV	Two-sided Gaussian	
8	m_h	LHC: $m_\mathrm{h}^{\text{low}} = 120$ GeV, $m_\mathrm{h}^{\text{high}} = 130$ GeV	1 if $m_\mathrm{h}^{\text{low}} \le m_\mathrm{h} \le m_\mathrm{h}^{\text{high}}$ 0 if $m_\mathrm{h} < m_\mathrm{h}^{\text{low}}$ or $m_\mathrm{h} > m_\mathrm{h}^{\text{high}}$	reweight
9	μ_h	CMS and ATLAS in LHC Run 1, Tevatron	LILITH 1.01 [50, 51]	post-MCMC
10	sparticle masses	LEP [52] (via MICROMEGAS [53–55])	1 if allowed 0 if excluded	

The upper limit

See T. Junk
http://arxiv.org/abs/hep-ex/9902006

$$\mathrm{CL_{s+b}} = P_{s+b}(X \leq X_{\mathrm{obs}}) = \frac{\mathrm{e}^{-(s+b)} \sum_{n=0}^{n_{\mathrm{obs}}} (s+b)^n}{n!}.$$

$$\mathrm{CL_s} = \mathrm{CL_{s+b}} / \mathrm{CL_b}$$

CL: **confidence level**

What is CLs?
Why not CLs+b?

X: a **test statistic** or **discriminant**

b: number of expected background events (estimated)

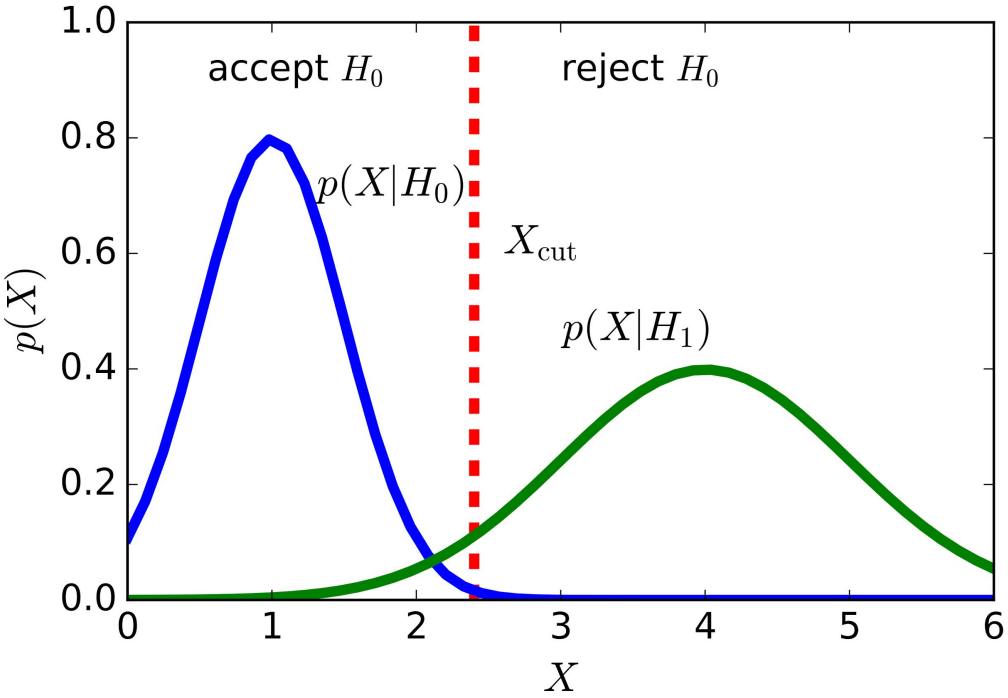
n_{obs}: number of observed events in data (counted)

s: varied until a certain **confidence level**, e.g. 95%, is reached → $s_{\mathrm{obs}}^{\mathrm{UL},95}$

The cross section is related to the number of events: $n_i = \sigma_i \times A\epsilon \times \mathcal{L}$, or xsec * **efficiency** * luminosity

Testing hypotheses: probability distributions

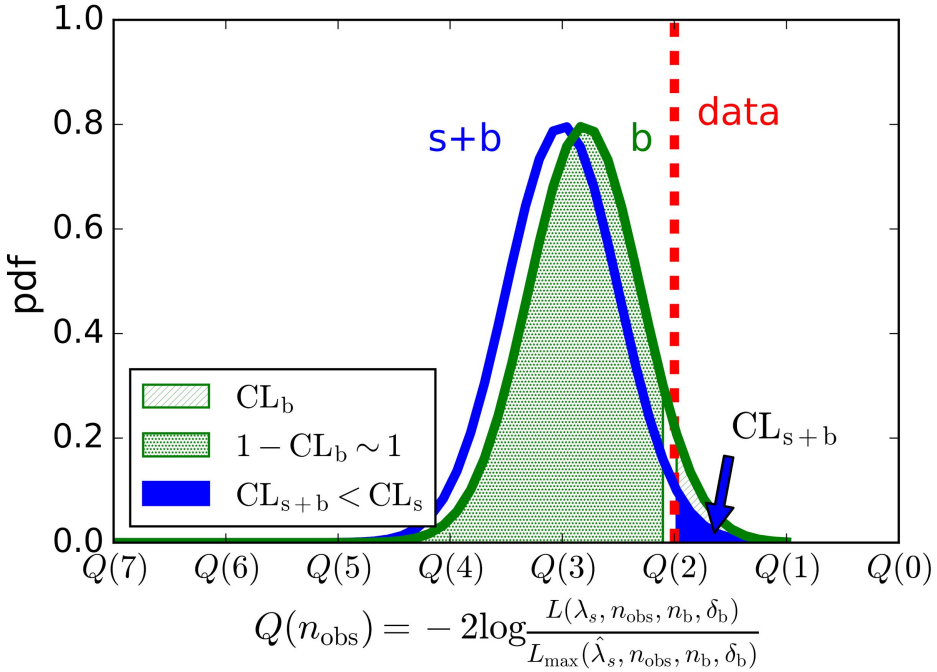
Accept null hypothesis H0

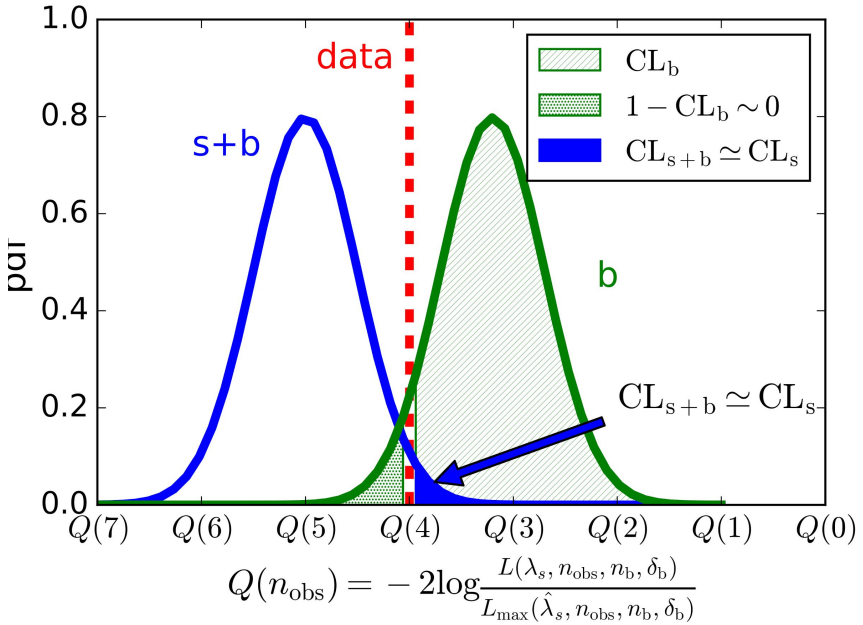


Reject null hypothesis

What are CLs, CLb, and CLs+b?

Confidence level: CLs, CLb, and CLs+b



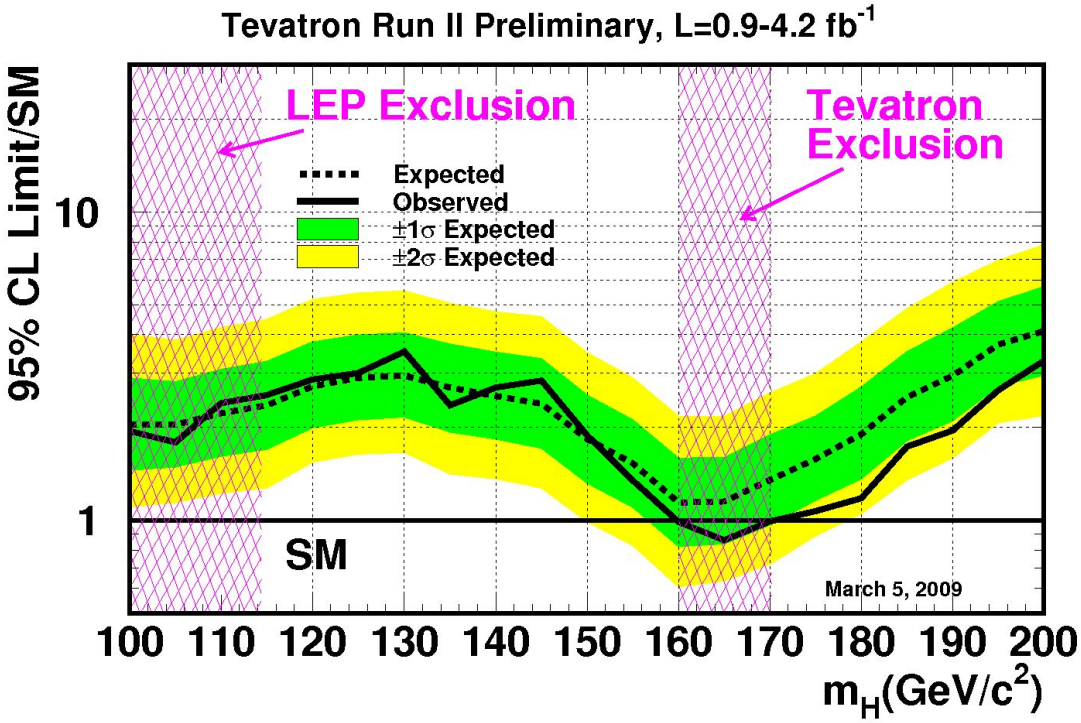


CLs = CLs+b/CLb

Another example of a “limit”

Limit on the Higgs cross section in terms of the Standard Model expectation

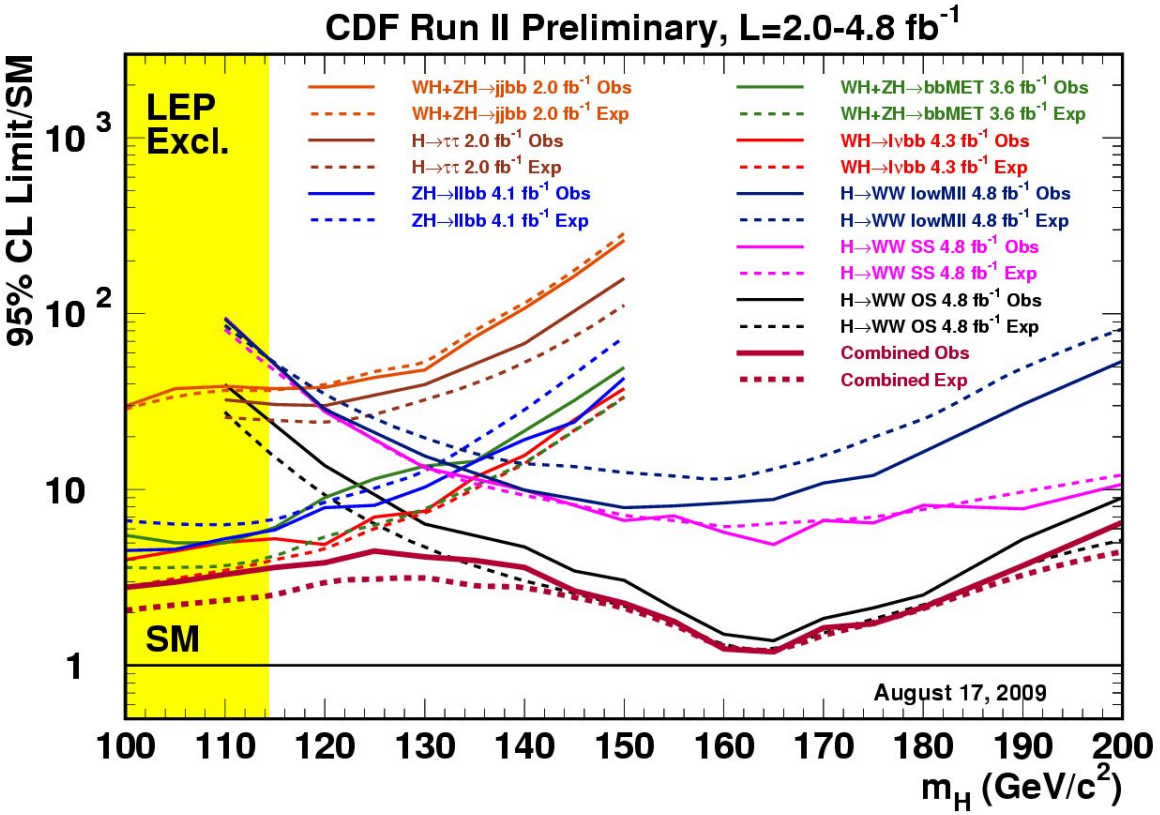
- Dashed line: expected limit
- Solid line: observed limit



https://www-cdf.fnal.gov/physics/new/hdg/Plain_English_files/higgs-limits-march2009.gif

Many results combined

We (outside the experiment) usually are not able to combine results without knowing all correlations

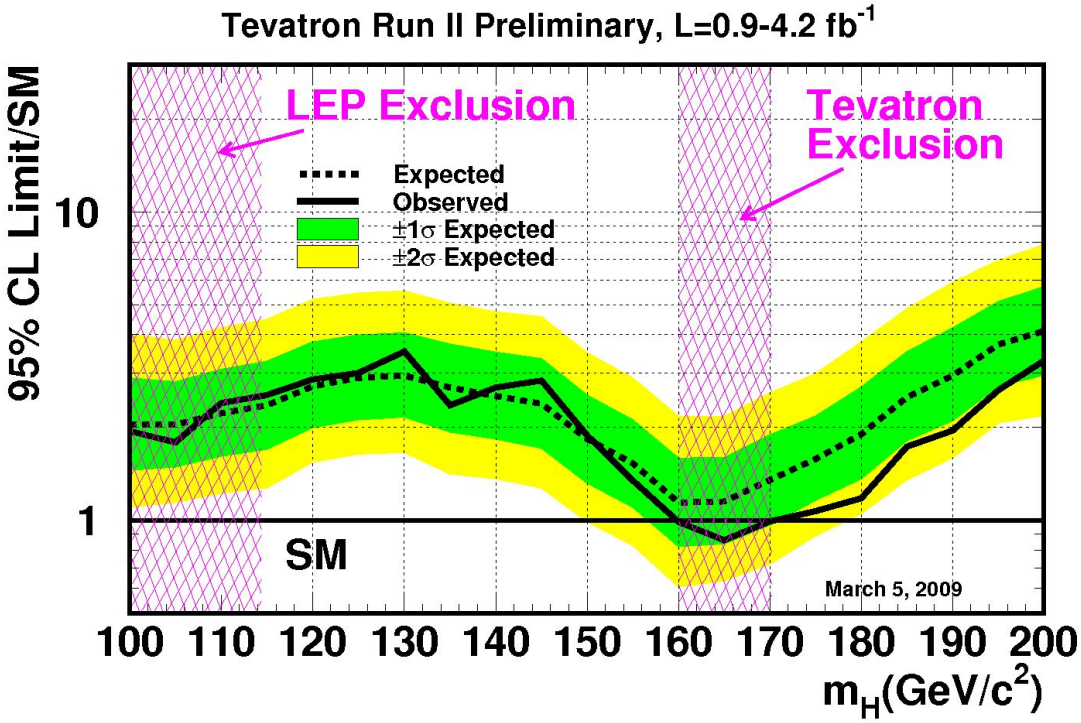


Another example of a “limit”

Limit on the Higgs cross section in terms of the Standard Model expectation

“Expected”: limit on Higgs mass given only pseudoexperiments (not data) → gives sensitivity

Y-axis: limit in terms of SM predicted cross section of Higgs



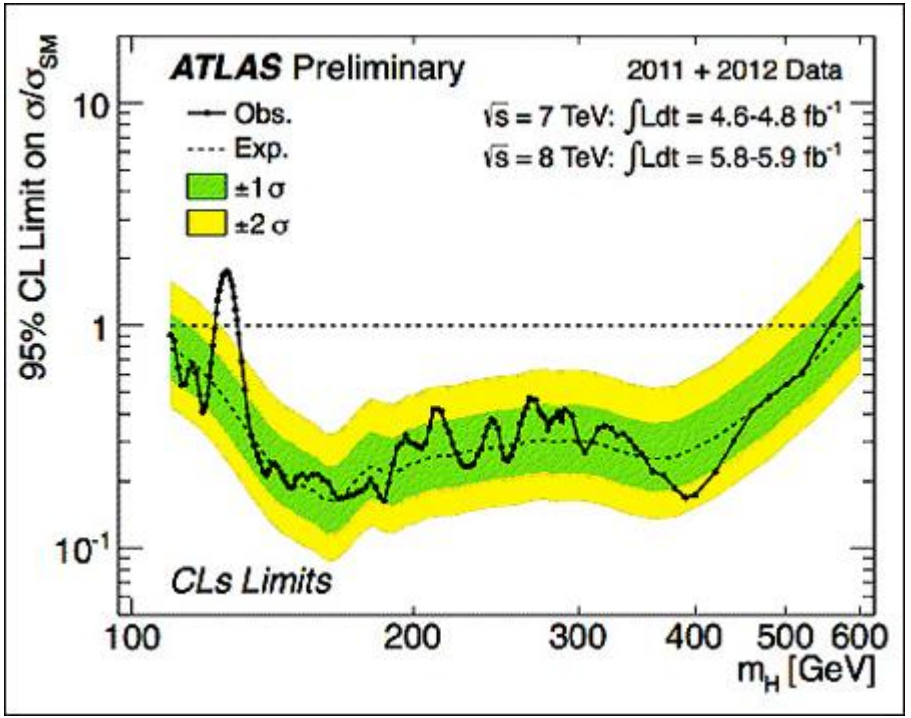
Can we find a Higgs boson here?

https://www-cdf.fnal.gov/physics/new/hdg/Plain_English_files/higgs-limits-march2009.gif

Sensitivity

Limit on cross section of a Higgs boson

- 2*SM expected → need twice the SM expected Higgs cross section to see a signal
- 1*SM expected → can identify Higgs boson if it were there

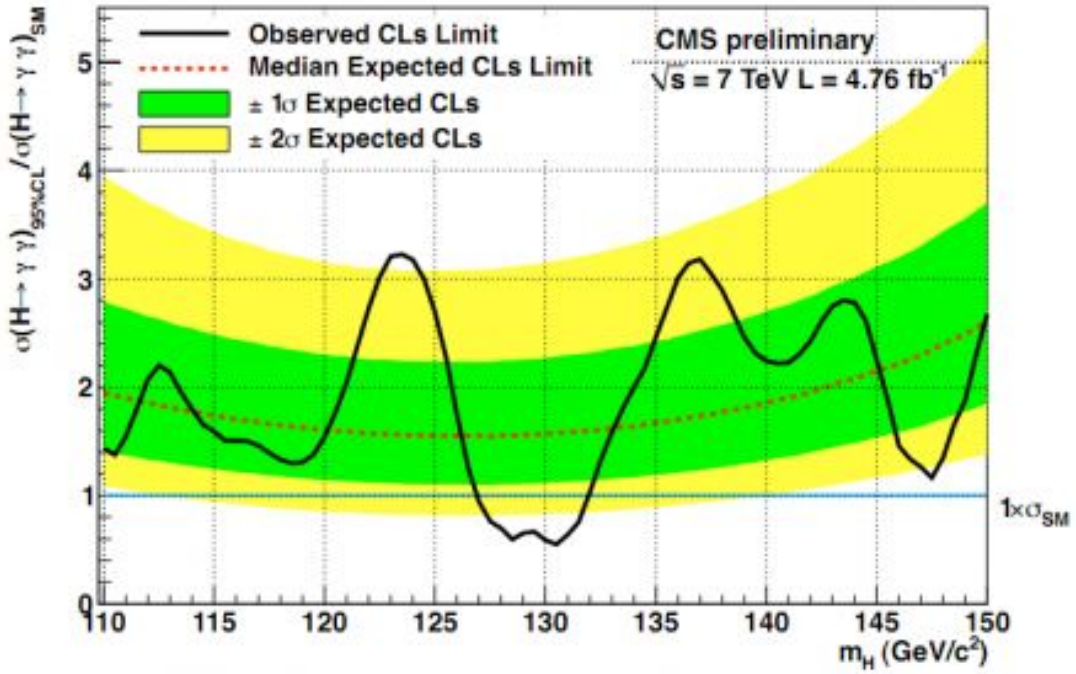


Can we find a Higgs boson here?

A signal

- Observed > expected: upward fluctuation, signal
- Observed < expected: downward fluctuation, exclusion

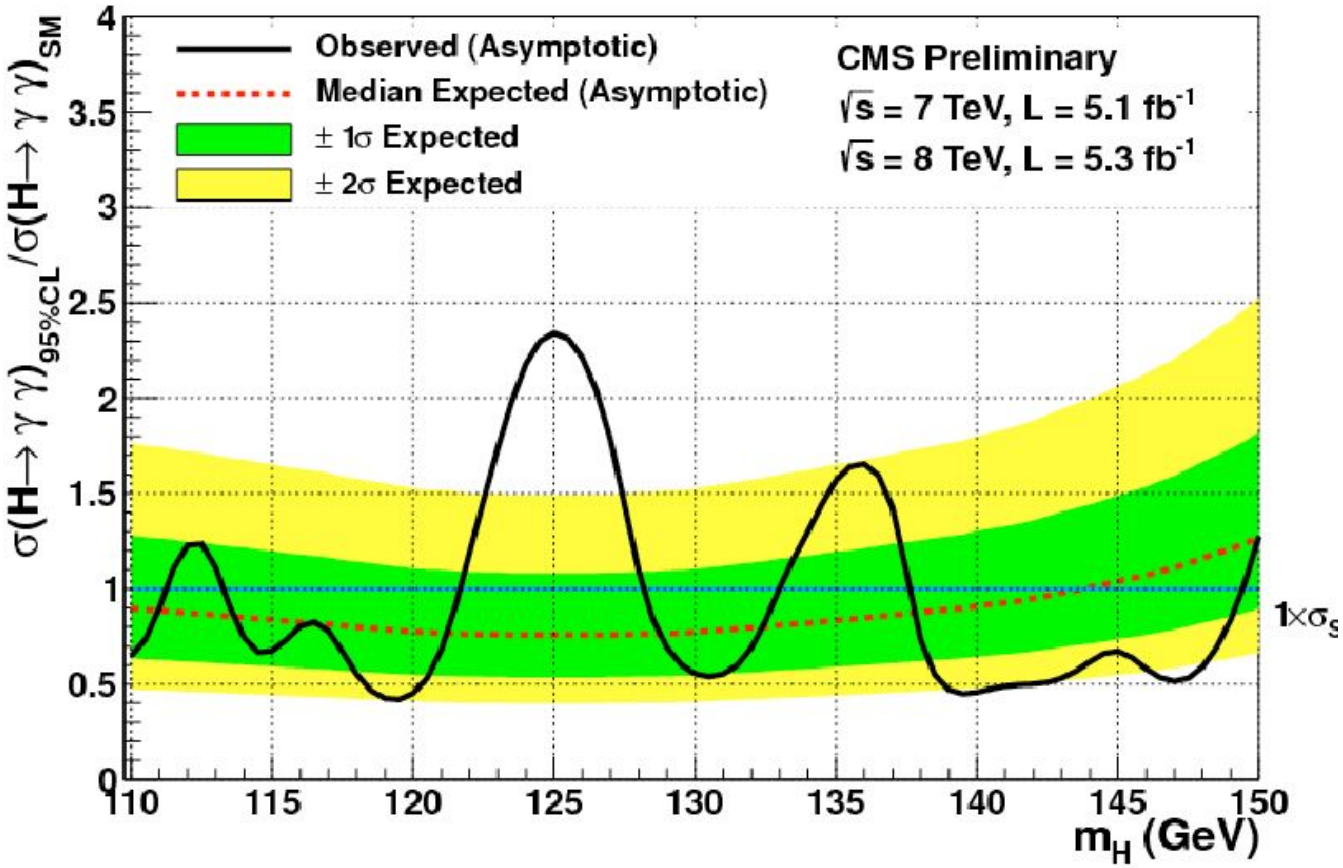
One expects ⅔ of observed limit to fall within bands of pseudoexperiments.



What do we see here? Did we exclude? Discover?

More data

What do we see now?



Methods at LEP, Tevatron, LHC

Table 11: Comparison of CL_s definitions as used at LEP, Tevatron, and adopted for the summer 2011 Higgs combination at LHC.

	Test statistic	Profiled?	Test statistic sampling
LEP	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\tilde{\theta})}{\mathcal{L}(data\|0,\tilde{\theta})}$	no	Bayesian-frequentist hybrid
Tevatron	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|0,\hat{\theta}_0)}$	yes	Bayesian-frequentist hybrid
LHC	$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|\hat{\mu},\hat{\theta})}$	yes $(0 \leq \hat{\mu} \leq \mu)$	frequentist

https://cdsweb.cern.ch/record/1375842/files/ATL-PHYS-PUB-2011-011.pdf

From madanalysis/misc/run_recast.py

MadAnalysis: computing limits and confidence level

```
def CLs(SigHypothesis, NumObserved, ExpectedBG, BGError, NumToyExperiments, cl=0.95):

    ExpectedBGs = stats.norm.rvs(loc=ExpectedBG, scale=BGError, \
    size=NumToyExperiments)

    ToyBGs = stats.poisson.rvs(ExpectedBGs)

    p_b = stats.percentileofscore(ToyBGs, NumObserved, kind='weak')*.01

    ExpectedBGandS = [expectedbg + SigHypothesis for expectedbg in ExpectedBGs]

    ToyBplusS = stats.poisson.rvs(ExpectedBGandS)

    p_SplusB = stats.percentileofscore(ToyBplusS, NumObserved, kind='weak')*.01

    return 1. - (p_SplusB / p_b) - cl # 1 - CLs
```

number of predicted signal events

number of observed events: data

number of background events

error on estimated background

number of toy experiments for estimating expected distributions

desired confidence level

1. `ExpectedBGs`: randomly generated normal distribution around **given** number of background events `ExpectedBG`
2. `ToyBGs`: 'toy backgrounds', or random poisson generated values around each 'expected background'
3. `p_b`: probability of background to fluctuate as low as toy backgrounds **given** data `NumObserved`
4. `p_SplusB`: probability of toy background + **given** signal `SigHypothesis` to be larger than the **given** data (computed from fraction using toy distribution)
5. `CLs = p_SplusB/p_b` confidence level

```
# !! Create a hepfile at path-to-your-hepfile.hep (or directory with
files)
./bin/ma5 -R
set main.recast = on
set main.recast.store_root = True
import path-to-your-hepfile.hep as test
set test.xsection = 10
Submit
N
```

Try it yourself in MadAnalysis

See output in (the '0' increments!):
ANALYSIS_0/Output/CLs_output_summary.dat

Not shown: uncertainties, dataset name (test)

What do these outputs mean?

analysis name	signal region	best?	sig95(exp)	sig95(obs)	CLs		efficiency
ATLAS_1604_07773	EM1	0	22182.5588293	23543.2331078	0.0000000	||	0.0000116
ATLAS_1604_07773	EM2	0	10547.9918475	8056.7515372	0.0000000	||	0.0000158
ATLAS_1604_07773	EM3	0	6049.2112595	3655.2255716	0.0029150	||	0.0000183
ATLAS_1604_07773	EM4	0	2505.5169811	2875.6391430	0.0095945	||	0.0000365
ATLAS_1604_07773	EM5	0	978.0965381	1400.0154618	0.0056240	||	0.0000357
ATLAS_1604_07773	EM6	0	783.8631224	551.9835762	0.0493750	||	0.0000257
ATLAS_1604_07773	EM7	1	167.0462343	215.5491163	0.0477806	||	0.0000897
ATLAS_1604_07773	IM1	0	2494.4494597	2267.4810428	0.0094842	||	0.0002333
ATLAS_1604_07773	IM2	0	1608.4411402	1298.2146926	0.0100975	||	0.0002217
ATLAS_1604_07773	IM3	0	1043.7782258	912.3879836	0.0133353	||	0.0002059
ATLAS_1604_07773	IM4	0	679.5583044	804.8926534	0.0151351	||	0.0001876
ATLAS_1604_07773	IM5	0	342.8603425	410.3009915	0.0260186	||	0.0001511
ATLAS_1604_07773	IM6	0	236.2176592	206.2989571	0.0878135	||	0.0001154
cms_b2g_14_004	control region 1	0	16511304.577969116524165.14609540.0002901			||	0.0000
cms_b2g_14_004	control region 2	0	-1	-1	0.0000000	||	0.0000000
cms_b2g_14_004	signal region	1	762.7432505	878.2088319	0.0125393	||	0.0000008
cms_b2g_12_022	S0b	0	-1	-1	0.0000000	||	0.0000000
cms_b2g_12_022	S1b	0	-1	-1	0.0000000	||	0.0000000

Backup

References and links

- T. Junk http://arxiv.org/abs/hep-ex/9902006
- A.L Read on CLs Journal of Physics G: Nuclear and Particle Physics 28 no. 10, (2002) 2693
- Statistical Data Analysis by Glenn Cowan
- Likelihoods https://cdsweb.cern.ch/record/1375842/files/ATL-PHYS-PUB-2011-011.pdf
- Simplified likelihoods in CMS https://cds.cern.ch/record/2242860/
- Data Analysis, a Baysian Tutorial by Sivia and Skilling
- Slides by Luca Lista
- Azatov, Contino, Galloway https://arxiv.org/pdf/1202.3415v3.pdf
- Wikipedia on sensitivity, specificity, Bayes' theorem, type I and II errors, …
- Ipython notebook examples: https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks#machine-learning-statistics-and-probability

And also:

- How Randomness Rules Our Lives by Leonard Mlodinow
- Critical Mass by Philip Ball
- Thinking, Fast and Slow by Daniel Kahneman

Try it yourself: three doors problem

Note: I would simulate this =)

One student plays the contestant, and another, the host. Label three paper cups #1, #2, and #3. While the contestant looks away, the host randomly hides a penny under a cup by throwing a die until a 1, 2, or 3 comes up. Next, the contestant randomly points to a cup by throwing a die the same way. Then the host purposely lifts up a losing cup from the two unchosen. Lastly, the contestant "stays" and lifts up his original cup to see if it covers the penny. Play "not switching" two hundred times and keep track of how often the contestant wins.

Then test the other strategy. Play the game the same way until the last instruction, at which point the contestant instead "switches" and lifts up the cup not chosen by anyone to see if it covers the penny. Play "switching" two hundred times, also.

https://en.wikipedia.org/wiki/Monty_Hall_problem
http://marilynvossavant.com/game-show-problem/

Change door!

	DOOR 1	DOOR 2	DOOR 3	RESULT
GAME 1	AUTO	GOAT	GOAT	Switch and you lose.
GAME 2	GOAT	AUTO	GOAT	Switch and you win.
GAME 3	GOAT	GOAT	AUTO	Switch and you win.
GAME 4	AUTO	GOAT	GOAT	Stay and you win.
GAME 5	GOAT	AUTO	GOAT	Stay and you lose.
GAME 6	GOAT	GOAT	AUTO	Stay and you lose.

Confidence level: confidence intervals

See also the book: Statistical Data Analysis by Glenn Cowan

Want to measure an unknown parameter θ:

We do not know we know θ, but if we know the pdf g for an **estimator** $\hat{\theta}$ for a given value of θ, we can define:

- The probability to find θ in an interval higher than a lower value α:

$$\alpha = p(\hat{\theta} \geq \hat{\theta}^{\alpha}_{\text{lower}}(\theta)) = \int_{\hat{\theta}^{\alpha}_{\text{lower}}(\theta)}^{\infty} g(\hat{\theta};\theta)d\hat{\theta} = 1 - G(\hat{\theta}^{\alpha}_{\text{lower}}(\theta);\theta)$$

- The probability to find θ in an interval lower than a higher value β:

$$\beta = p(\hat{\theta} \geq \hat{\theta}^{\beta}_{\text{upper}}(\theta)) = \int_{-\infty}^{\hat{\theta}^{\beta}_{\text{upper}}(\theta)} g(\hat{\theta};\theta)d\hat{\theta} = G(\hat{\theta}^{\beta}_{\text{upper}}(\theta);\theta).$$

- From this we find **the probability to have a true value θ in the interval [a,b]**

$$p(\hat{\theta}^{\beta}_{\text{upper}}(\theta) \leq \hat{\theta} \leq \hat{\theta}^{\alpha}_{\text{lower}}(\theta)) = 1 - \alpha - \beta. \rightarrow \quad p(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta$$

Confidence interval: single-sided

$$p(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta$$

A **one-sided confidence interval** or **limit** is the **lower limit a on θ** so that

$a < \theta$

with the **probability 1 - α**

False positive rate

or the **upper limit b on θ** so that

$\theta < b$

with the **probability 1 - β**

False negative rate

A central interval can be defined with probability 1 - α - β

Nuisance parameters

nuisance parameters: unknown parameters not interesting for the measurement:

- detector resolution
- uncertainty in backgrounds
- background shape modeling
- other systematic uncertainties
- ...

Then one can:

- Add the nuisance parameters to your likelihood model (easier to incorporate in a fit than in upper limits)
- “Integrate it away” (Bayesian way)

Likelihood methods

Frequentist: constrain likelihood nuisance parameters with e.g. control sample:

$$L(\text{data, control sample} \mid \theta) = L(\text{ data} \mid \theta, s, b)\, L(\text{control sample} \mid \theta)$$

a large fraction (68% or 95%, usually) of the experiments contain the fixed but unknown true value within the confidence interval $[\theta^{est} - \delta_1 , \theta^{est} + \delta_2]$

Bayesian: integrate analytically, numerically, with Markov Chain Monte Carlo

The posterior PDF for θ is maximum at θ est and its integral is 68% within the range $[\theta^{est} - \delta_1 , \theta^{est} + \delta_2]$

Hybrid: integrate or ***marginalize*** the likelihood function over the nuisance parameters