# Collider Physics - Chapter 4
# Searches and limits

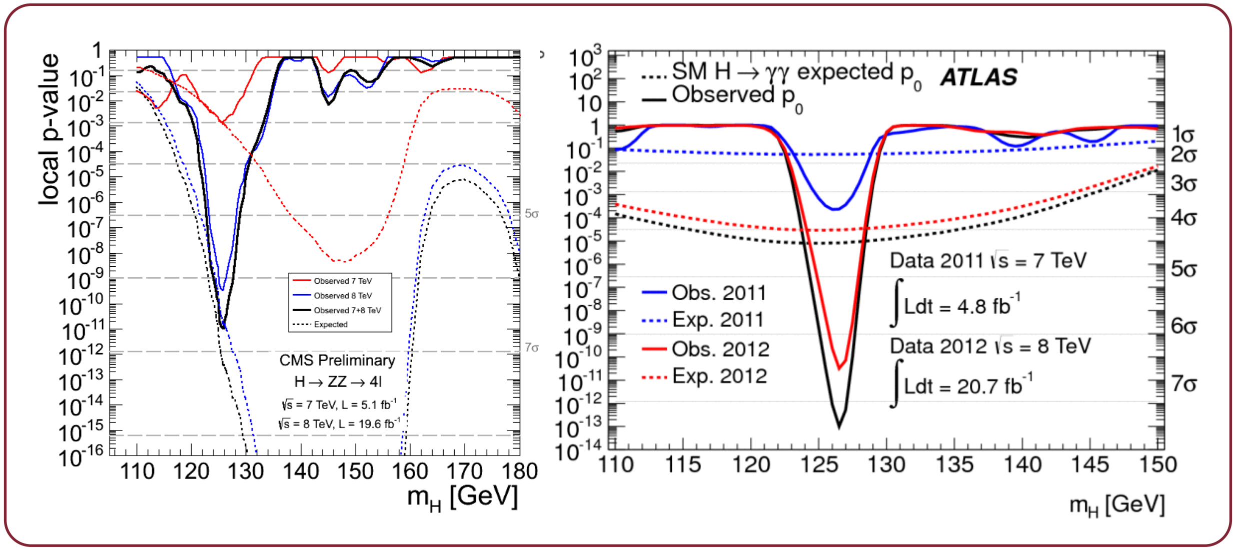
Paolo Bagnaia
SAPIENZA
UNIVERSITÀ DI ROMA

AA 21-22

*last mod.*
*8-Nov-21*

# 4 – Searches and limits

1. Probability
2. Searches and limits
3. Limits
4. Maximum likelihood
5. Interpretation of results




- methods commonly used in all recent searches (e.g. LEP, LHC, gravitational waves);
- also in other lectures (e.g. "Laboratorio di Meccanica", "Physics Laboratory");
- but "repetita juvant" (maybe);
- not a well-organized presentation, beyond the scope of present lectures ($\rightarrow$ references).

*There are three kinds of lies: lies, damned lies, and statistics.*

*B. Disraeli*

*If you torture the data long enough, it will confess to anything.*

*R. Coase*

# probability: a new guest star in the game

- Modern particle physics makes a large use of (relatively) new sciences : **probability** and her sister **statistics**;
- [*we are scientists, not gamblers, and do NOT discuss poker and dice here*];
- in classical physics the resolution function of an observable can be seen as a pdf(*);
- q.m. is probabilistic, at least in its original (Copenhagen) interpretation, since the "theory" (e.g. $d\sigma/d\cos\theta$) is a distribution, while the "measure" produces single values;
- but its use to assess a statement *[e.g. "the probability that we have discovered the Higgs boson in our data"]* is really modern;
- actually the humankind thinks in terms of probability (*risk*, *chance*, *luck* … mean "probability", while *experience*, *past events*, *use*, … mean "statistics").

_______________

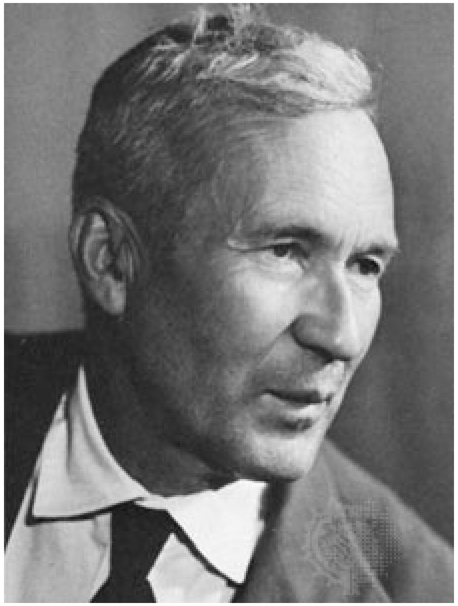
(*) pdf: acronym for *probability distribution function*. (or *probability density function*).

*For [some] readers :*

- *these lectures avoid carefully to enter in the discussion frequentism ↔ bayesianism;*
- *however, a modern particle physicist must understand and use both;*
- *only, (try to) avoid fights.*

# probability: Kolmogorov axioms

Andrei Nikolayevich Kolmogorov [Андрей Николаевич Колмогоров] (1903–1987), a Russian (sovietic) mathematician, in 1933 wrote a fundamental paper on axiomatization of probability; he introduced the space S of the events (A, B, …) and the event probability as a measure $\mathscr{P}$(A) in S.

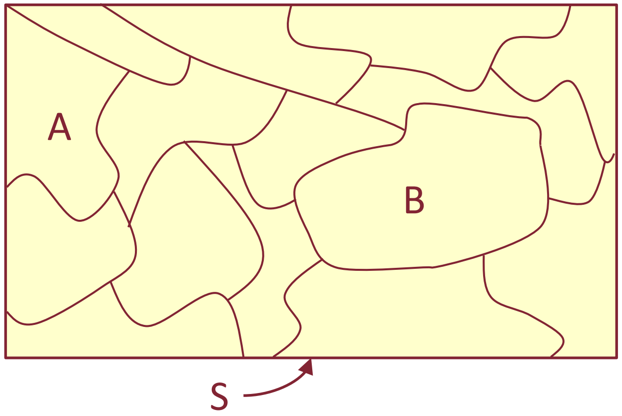


K. axioms are :

1. $0 \leq \mathscr{P}(A) \leq 1 \ \forall \ A \in S$;
2. $\mathscr{P}(S) = 1$;
3. $A \cap B = \emptyset \Rightarrow \mathscr{P}(A \cup B) = \mathscr{P}(A) + \mathscr{P}(B)$.

Some theorems (easily demonstrated):

- $\mathscr{P}(\bar{A}) = 1 - \mathscr{P}(A)$;
- $\mathscr{P}(A \cup \bar{A}) = 1$;
- $\mathscr{P}(\emptyset) = 0$;
- $A \subset B \Rightarrow \mathscr{P}(A) \leq \mathscr{P}(B)$;
- $\mathscr{P}(A \cup B) = \mathscr{P}(A) + \mathscr{P}(B) - \mathscr{P}(A \cap B)$.



# searches and limits

- Sometimes, the result of a measure is NOT the value of an observable x :

  "$x = x_{exp} \pm \Delta x$",

- but, instead, a qualitative "**search**" :

  "the phenomenon $\mathcal{Y}$ does (not) exist",

  **or, alternatively :**

  "the phenomenon $\mathcal{Y}$ does NOT exist in the parameter range $\Phi$".

- [statements with "not" apply if the effect is not found, and an "**exclusion**" (a "**limit**", when $\Phi$ is not full) is established]
- In modern experiments, the searches account for >50% of the published papers, and almost all are negative [*but the Higgs search at LHC, of course*].
- Obviously, a negative result is NOT a failure: if any, it is a failure of the theory under test.
- [but a **discovery** is much more pleasant and rewarding]
- A rigorous procedure, well understood and "easy" to apply, is imperative.
- This method is a major success of the LEP era: it uses math, statistics, physics, common sense and communication skill.
- It MUST be in the panoply of each particle physicist, both theoreticians and experimentalists.

The examples here remain inside the SM:

- Higgs searches at LEP (negative) and LHC (positive) ;
- after the Higgs discovery, the focus has shifted toward "bSM" searches, but the method has not changed (still improving).

# searches and limits: definitions

In the next slides :

- $\mathscr{L}_{int}$ : int. luminosity (sometimes only $\mathscr{L}$);
- $\sigma_s$ : cross section of signal (searched for);
- $\sigma_b$ : cross section of backgrounds (known);
- $\varepsilon_s$ : efficiency for signal (0÷1, larger is better);
- $\varepsilon_b$ : ditto for backgrounds (0÷1, smaller is better);
- s : # expected signal events [$s = \mathscr{L}_{int}\varepsilon_s\sigma_s$];
- b : ditto for backgrounds [$b = \mathscr{L}_{int}\varepsilon_b\sigma_b$];
- n : # expected events [$n = s + b$, or $n = b$];
- N : # found events (N fluctuates around n with Poisson ($\rightarrow$ Gauss) statistics;
- $\mathscr{P}$ : probability, according to a given pdf;
- CL : "confidence level", a limit (< 1) in the cumulative probability;
- $\Lambda$ : likelihood function for signal+bckgd ($\Lambda_s$) or bckgd-only ($\Lambda_b$) hypotheses;
- $\mu$ : parameter defining the signal level [$n = b + \mu s$], used for limit definition;
- p : "p-value", probability to get the same result or another less probable, in the hypothesis of bckgd-only;
- E[x] : expected value of the quantity "x".

a lot of math, but try to understand the underlying physics.

# searches and limits: verify/falsify

- A theory (SM, SUSY, …) predicts a phenomenon (a particle, a dynamic effect), e.g. $e^+$, $\bar{p}$, $\Omega^-$, $W^\pm/Z$, H;
- [in some cases the phenomenon depends on unknown parameter(s), e.g. the Higgs boson mass]
- a new device (e.g. an accelerator) is potentially able to observe the phenomenon [fully or in a range of the parameters space still unexplored];
- therefore, two possibilities:

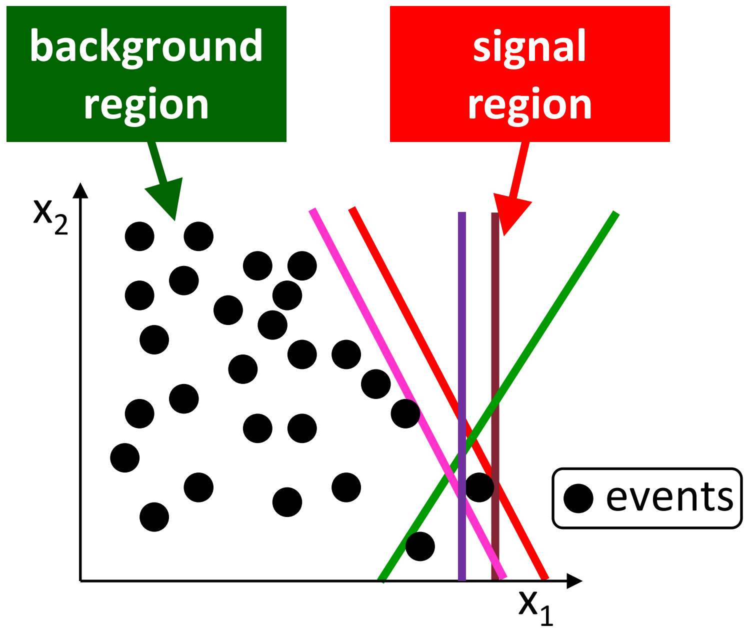
  A. **observation**: the theory is "verified" (à la Popper) [and the free parameter(s) are measured];

  B. **non-observation**: the theory is "falsified" (à la Popper) [or some subspace in the parameter space, e.g. an interval in one dimension, is excluded $\rightarrow$ a "**limit**" is established];

❖ different approach, nowadays less common ("model independent"): look for unknown effects, without theoretical guidance, e.g. $\mathbb{CP}$ violation, $J/\psi$.

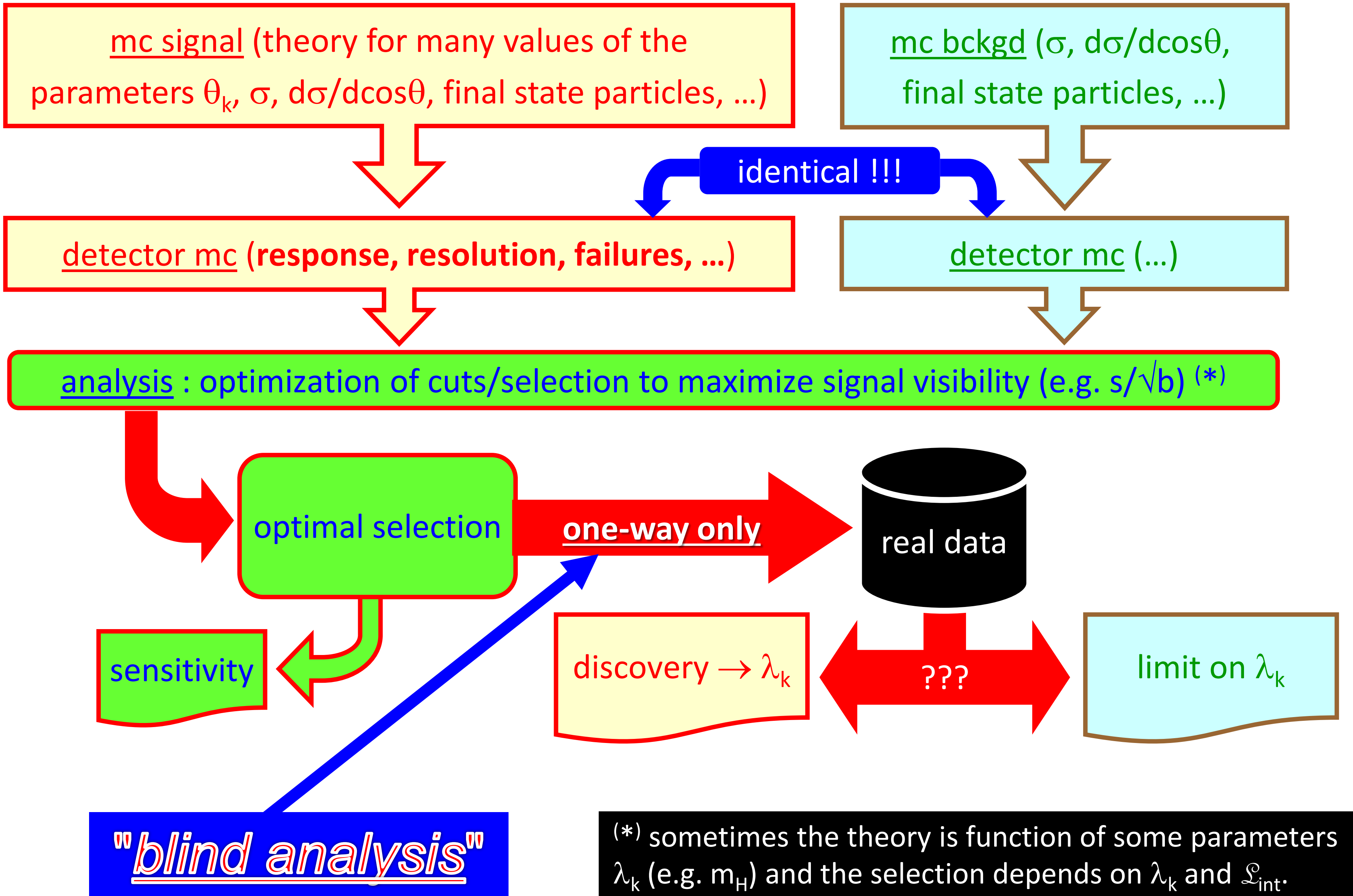
# searches and limits: blind analysis

- the key point: usually $b \gg s$, but $f_b(x)$ and $f_s(x)$ are <u>very</u> different → cuts in the event variables ($x$ : mass, angle, …), such that to enhance s wrt b;
- when n is large ($n \gg \sqrt{n}$), statistical fluctuations (s.f.) do NOT modify the result;
- usually (not only for impatience) n is small and its s.f. are important;
- small variations in the selection (→ small change in b and s) may produce large differences in the result N [*look at the example in two variables: e.g., if b=0.001 after the cuts, when N changes 0 → 1, N=0 or N=1 is totally different*];
- a "neutral" analysis is impossible; <u>a posteriori</u>, it is easy to justify a little change in the selection (e.g. cuts), which strongly affects the results;
- therefore, the only correct procedure consists in defining the selection <u>a priori</u> (e.g. by optimizing the <u>expected</u> visibility on a mc event sample); then, the selection is "blindly" applied to the actual event sample (→ "**<u>blind analysis</u>**").




which is the "right" cut ?

# searches and limits: flowchart



# limits

[*in the "good ole times", life was simpler: if the background is negligible, the first observations led to the discovery, as for* $e^+$, $\bar{p}$, $\Omega^-$, $W^\pm$ *and* Z]

- in most cases, the background (reducible or irreducible) is calculable;
- a discovery is defined as an observation that is incompatible with a +ve statistical fluctuation respect to the expected background alone;
- a limit is established if the observation is incompatible with a –ve fluctuation respect to the expected (signal + background);
- both statements are based on a "reductio ad absurdum"; since all values of N in $[0,\infty]$ are possible, it is compulsory to predefine a CL to "cut" the pdf;
- the CL for discovery and exclusion can be different : usually for the discovery stricter criteria are required;
- a priori the expected signal s can be compared with the fluctuation of the background (in approximation of large number of events, $s \leftrightarrow \sqrt{b}$) : $n_\sigma = s / \sqrt{b}$ is a figure of merit of the experiment;
- a posteriori the observed number (N) is compared with the expected background (b) or with the sum (s + b).

________________________________

*Example. In an exp., expect 100 background events and 44 signal after some cuts; use the "large number" approximation ($\Delta n = \sqrt{n}$) :*

$b = 100$, $\Delta b = \sqrt{b} = 10$;

$n = s + b = 144$ , $\Delta n = 12$.

*The pre-chosen confidence level is "3 σ".*

*The discovery corresponds to an observation of*

$N > (100+3\times10) = 130$ *events.*

*A limit is established if*

$N < (144 - 3\times12) = 108$ *events.*

*There is no decision if $108 < N < 130$.*

*The values $N < 70$ and $N > 180$ are "impossible".*

# limits: problem

Problem (based on previous example) :

You want an *experimentum crucis*. Compute the factor, wrt to previous luminosity, which allows to avoid the "no-decision" region. Assume the selection be the same.

# limits: Poisson statistics

- In general, the searches look for processes with VERY limited statistics (*want to discover asap*);
- therefore the limit ("n large", more precisely n >> √n) cannot be used (neither its consequences, like the Gauss pdf);
- searches are clearly in the "Poisson regime": large sample and small probability, such that the expected number of events (n,s) be not large;
- use the Poisson distribution :

$$\mathcal{P}(N|m) = \frac{e^{-m}m^N}{N!}; \quad \langle N \rangle = m; \quad \sigma_N = \sqrt{m};$$

- therefore, in a search, two cases :

1. the signal does exist :

$$\mathcal{P}(N|b+s) = \frac{e^{-(b+s)}(b+s)^N}{N!}; \quad \langle N \rangle = b+s; \quad \sigma_N = \sqrt{b+s};$$

[s may be known or unknown]

2. the signal does NOT exist :

$$\mathcal{P}(N|b) = \frac{e^{-b}b^N}{N!}; \quad \langle N \rangle = b; \quad \sigma_N = \sqrt{b};$$

- the strategy is : use N (= $N^{exp}$) to distinguish between case [1.] and [2.];
- since $\mathcal{P}$ is > 0 for any N in both cases, the procedure is to define a CL **a priori**, and accept the hypothesis [1.] or [2.] only if it falls in the **predefined** interval;
- modern (LHC) evolution : define a parameter, usually called "$\mu$" :

$$\mathcal{P}(N|b+\mu s) = \frac{e^{-(b+\mu s)}(b+\mu s)^N}{N!}; \quad \langle N \rangle = b+\mu s; \quad \sigma_N = \sqrt{b+\mu s};$$

clearly, $\mu$ = 0 is bckgd only, while $\mu$ = 1 means discovery; sometimes results are presented as limits on "$\mu$" [*e.g. exclude* $\mu$ = 0 *means "discovery"*].

# limits : discovery, exclusion

- the "rule" on the CL usually accepted by experiments is:

> ➢ DISCOVERY : $\mathcal{P}$(b only) $\leq 2.86\times10^{-7}$, [called also "5σ" [(1)]];
>
> ➢ EXCLUSION : $\mathcal{P}$(s+b) $\leq 5\times10^{-2}$ ; [called also "95% CL"];

- a priori, the integrated luminosity $\mathcal{L}_{int}$ for discovery / exclusion can be computed :

  ➢ $\mathcal{L}_{disc}$ : $\mathcal{L}_{int}$ min, such that 50% of the experiments[(2)] (i.e. an experiment in 50% of the times) had $\mathcal{P}$(b only) ≤ $\mathcal{P}_{disc}$;

  ➢ $\mathcal{L}_{excl}$ : $\mathcal{L}_{int}$ min, such that 50% of the experiments[(2)] (i.e. an experiment in 50% of the times) had $\mathcal{P}$(s+b) ≤ $\mathcal{P}_{excl}$;

NB : this rule corresponds to the median *["an experiment, in 50% of the times..."]*, and is slightly different from the average *["an experiment, with exactly the expected number of events ..."]*.

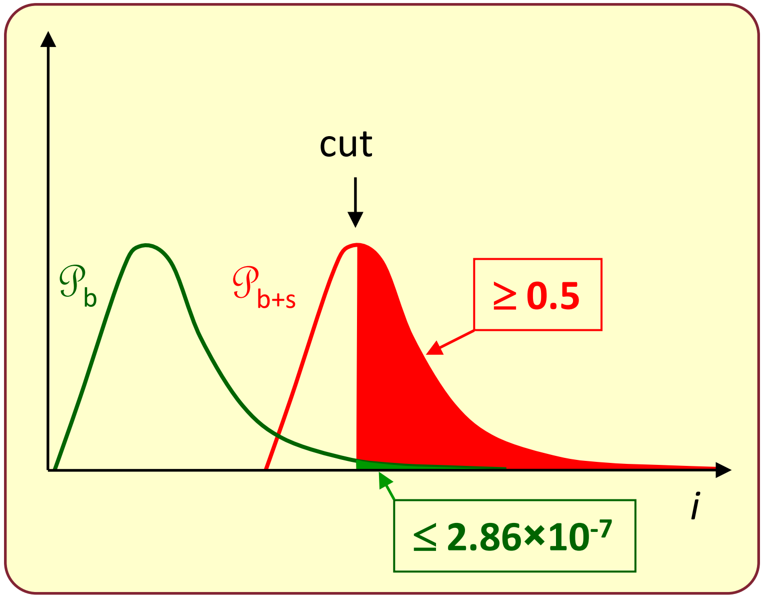
---

(1) this probability corresponds to 5σ for a gaussian pdf only; but the experimentalists use (*always*) the cut in probability and (*sometimes*) call it "5σ";

(2) for combined studies an "experiment" at LEP [LHC] results from the data of all 4 [2] collaborations; in this case $\mathcal{L}_{int} \to 4\ [2] \times \mathcal{L}_{int}$.

> *"A parameter is said to be excluded at xx% confidence level [say 95%] if the parameter itself would yield more evidence than that observed in the data at least 95% of the time in a [pseudo-] set of repeated experiments, all equivalent to the one under consideration."*
> *[CMS web dixit]*

# limits : Luminosity of discovery, exclusion




- The values of $\mathcal{L}_{disc}$ and $\mathcal{L}_{excl}$ come from the previous equations; compute $\mathcal{L}_{disc}$ ($\mathcal{L}_{excl}$ is similar) :

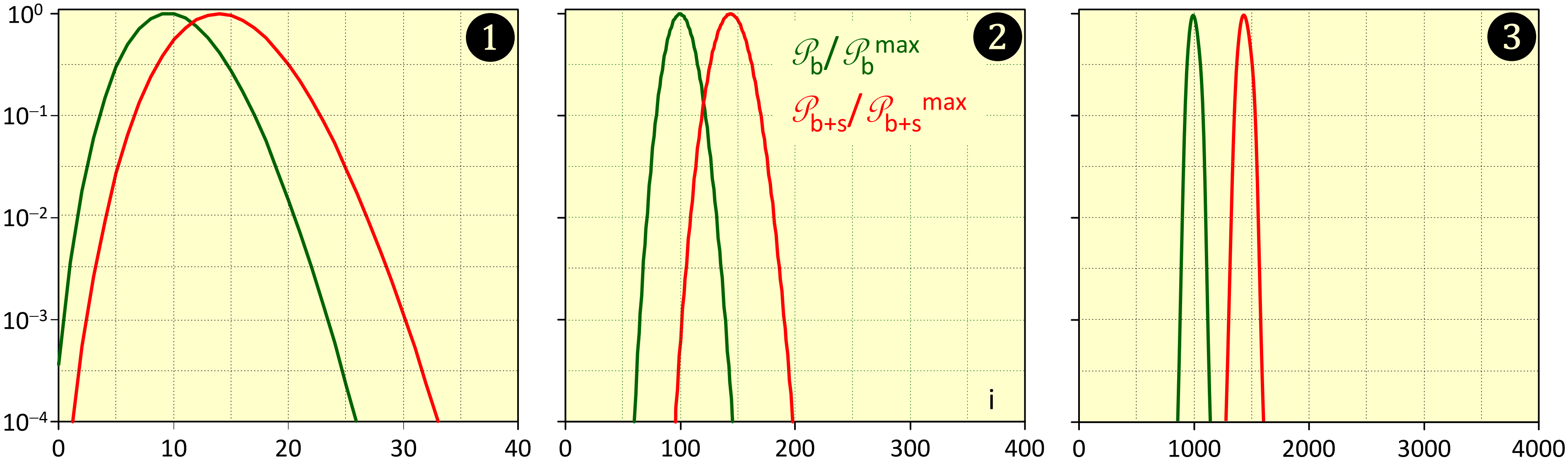
$$"\blacklozenge" = e^{-b} \times \sum_{i=N}^{\infty} \frac{(b)^i}{i!} \le \mathcal{P}(5\sigma) = 2.86 \times 10^{-7};$$

$$"\blacklozenge" = e^{-(b+s)} \times \sum_{i=N}^{\infty} \frac{(b+s)^i}{i!} \ge 0.5;$$

$$b = \mathcal{L}_{disc}\varepsilon_B\sigma_B; \quad s = \mathcal{L}_{disc}\varepsilon_S\sigma_S.$$

- assume increasing luminosity ($\mathcal{L}_{int}$ = $\mathcal{L}_{disc}$ [$\mathcal{L}_{excl}$]) and constant $\varepsilon_s$, $\varepsilon_b$, $\sigma_s$, $\sigma_b$;
- assume to start with small $\mathcal{L}_{int}$ : the two distributions overlap a lot, no N satisfies the system (i.e. the **green tail** above the **median** is too large);
- when $\mathcal{L}_{int}$ increases, the two distributions are more and more distinct (overlap $\propto 1/\sqrt{\mathcal{L}_{int}}$);
- for a given value of $\mathcal{L}_{int}$, it exists a number of events N, such that the cuts at **$2.86 \times 10^{-7}$** (**0.5**) in the **bckgd** (**signal**) cumulative coincide; this value of $\mathcal{L}_{int}$ correspond to $\mathcal{L}_{disc}$;
- this is the luminosity when, if the signal exists, 50% of the experiments have (at least) $5\sigma$ incompatibility with the hypothesis of bckgd only.

# limits : Luminosity increase




back to our example:

- b=100, s=44, b+s=144
- show the Poisson distributions for bckgd only and for bckgd+signal

[*notice: log-scale and normalization, x-axis=$N_{events}$, y-axis=$\mathcal{P}_{b,s}/\mathcal{P}_{b,s}^{max}$*]

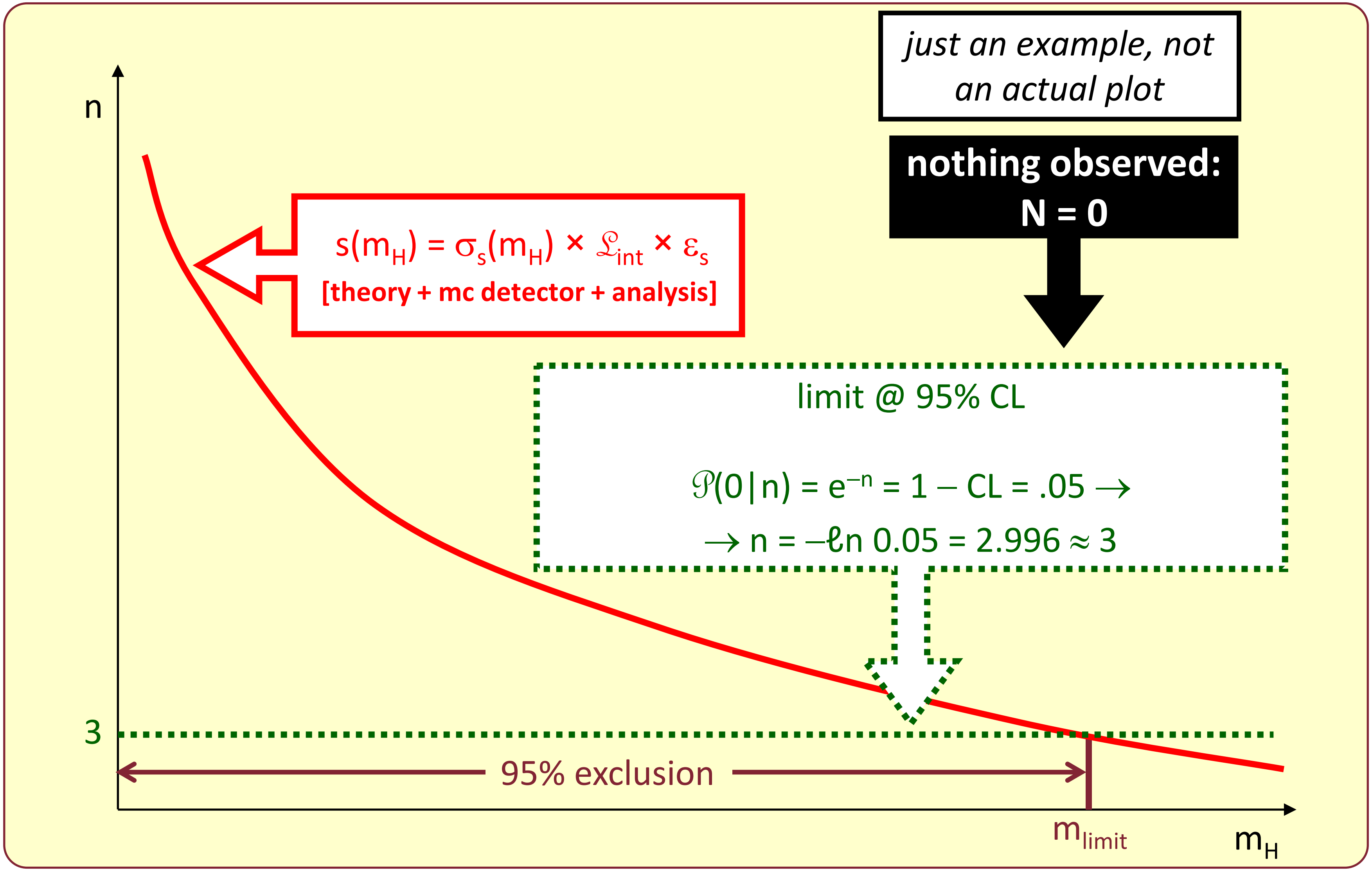
Q in the average case, ok for the 5σ rule ?

A no !!! because b+s (= 144) is at 4.4 σ from b (= 100) → $\mathcal{L}_{int}$ is not sufficient.

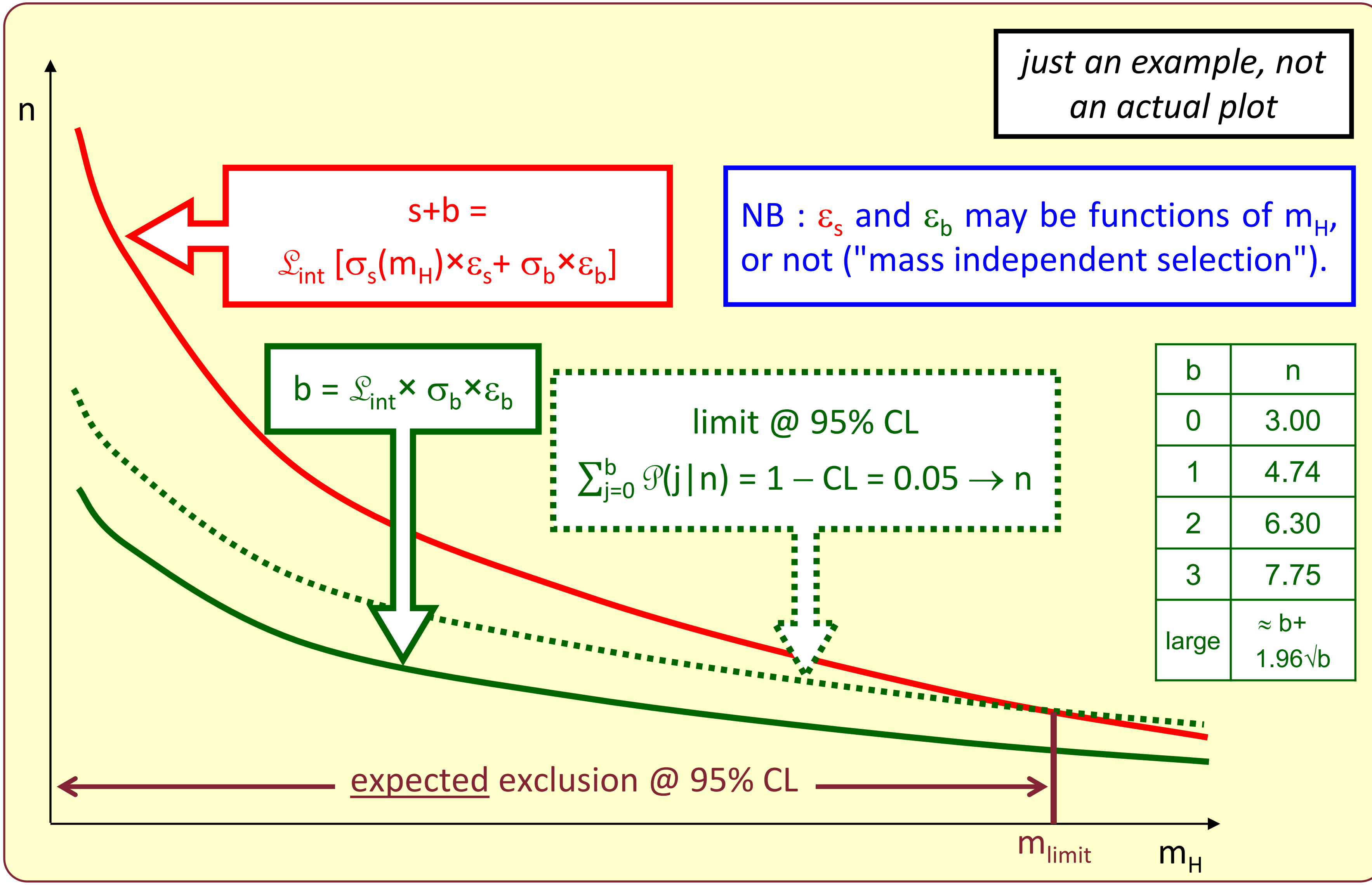
In a real data-taking run:

- at the beginning $\mathcal{L}_{int}$ is small, e.g. b=10, s=4.4, b+s=14.4 (plot n. 1);
- then our previous $\mathcal{L}_{int}$ (plot n. 2);
- finally a further increase of 10 in $\mathcal{L}_{int}$ (b=1000, s=440, b+s=1440, plot n. 3);
- in case 3, the 5σ rule is more than satisfied: ok ! (but long & expensive).

# limits : ex. $m_H$ (b=0, N=0)

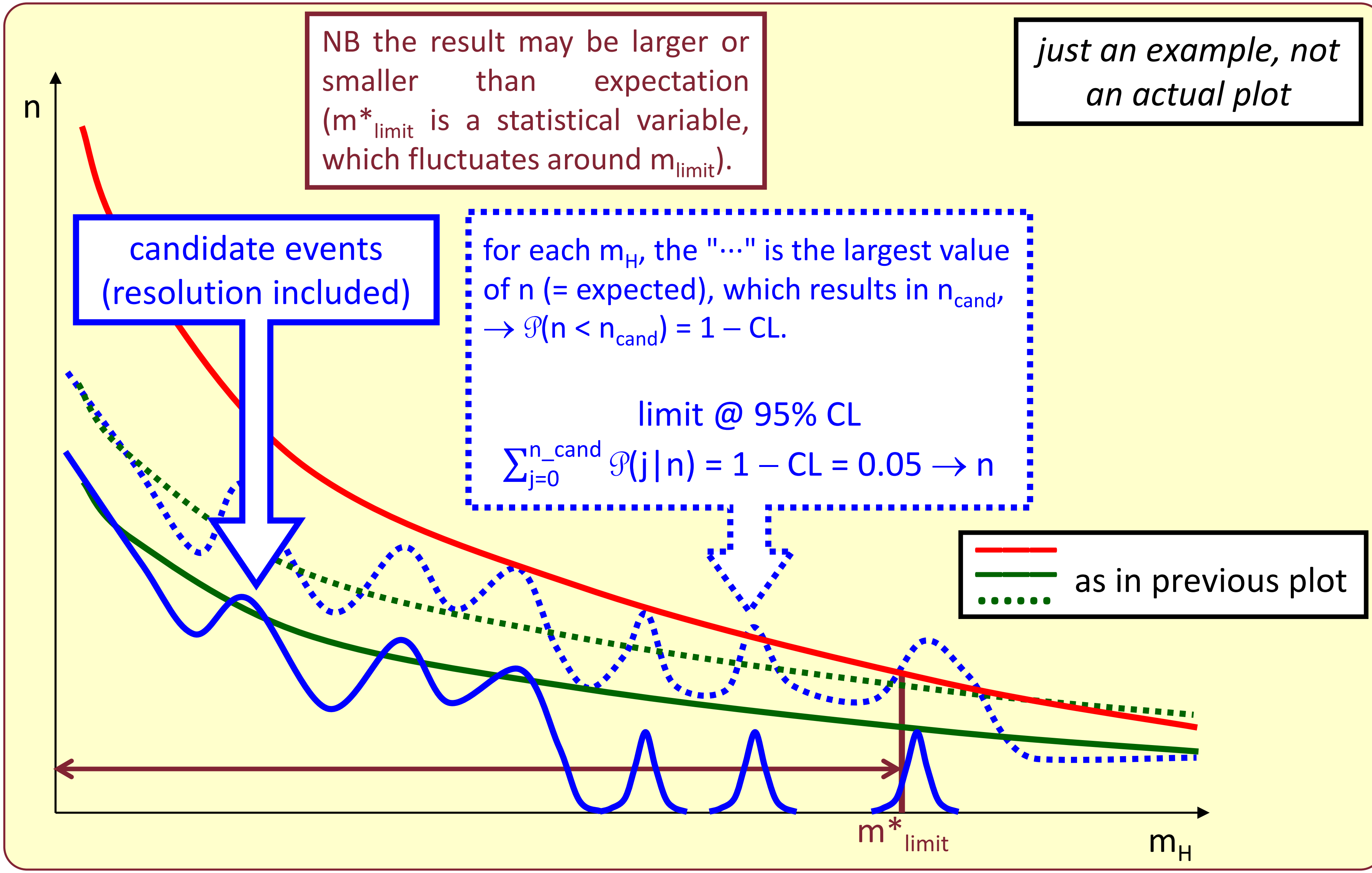


# limits : ex. $m_H$ (a priori, b>0)




| b | n |
|---|---|
| 0 | 3.00 |
| 1 | 4.74 |
| 2 | 6.30 |
| 3 | 7.75 |
| large | ≈ b+ 1.96√b |

# limits : ex. $m_H$ (a posteriori, b>0)



# maximum likelihood: definition

- A random variable x follows a pdf $f(x\,|\theta_k)$;
- the pdf f is a function of some parameters $\theta_k$ (k = 1,…,M), sometimes unknown;
- assume a repeated measurement (N times) of x :

  $x_j$ ( j = 1,…,N);
- define the likelihood $\Lambda$ and its logarithm $\ell n(\Lambda)$ [*see box*].

$$\Lambda \quad = \prod_{j=1}^{N} f(x_j\,|\,\theta_k);$$

$$\ell n(\Lambda) = \sum_{j=1}^{N} \ell n[f(x_j\,|\,\theta_k)].$$

Example : observe N decays with (unknown) lifetime $\tau$, measuring the lives $t_j$, j = 1,…,N.

$$\Lambda = \prod_{j=1}^{N} f(t_j\,|\,\tau) = \prod_{j=1}^{N} \frac{1}{\tau} e^{-t_j/\tau} = \frac{1}{\tau^N} e^{-\sum t_j/\tau};$$

$$\ell n(\Lambda) = \sum_{j=1}^{N} \ell n[\frac{1}{\tau} e^{-t_j/\tau}] = -N\ell n(\tau) - \frac{1}{\tau}\sum_{j=1}^{N} t_j.$$

then look for the value $\tau^*$, which maximizes $\Lambda$ (or $\ell n\Lambda$).

$\tau^*$ is the **max.lik. estimate** of $\tau$.

$$\frac{\partial \ell n(\Lambda)}{\partial \tau} = 0 = -\frac{N}{\tau^*} + \frac{1}{\tau^{*2}} \sum_{j=1}^{N} t_j \Rightarrow$$

$$\tau^* = \frac{1}{N} \sum_{j=1}^{N} t_j = <t>.$$

# maximum likelihood: parameter estimate

the m.l. method has the following important **asymptotic** properties [*no proof, see the references*]:

- consistent (1) ;
- no-bias (2);
- result $\theta^*$ distributed around $\theta_{true}$, with a variance given by the Cramér-Frechet-Rao limit [*see*];

(1) $N \to \infty \Rightarrow \theta^* \to \theta^{true}$

(2) $\text{Bias}(\theta^* \mid \theta^{true}) = E[\theta^*] - \theta^{true} = 0$

- "invariant" for a change of parameters, [i.e. the m.l. estimate of a quantity, function of the parameters, is given by the function of the estimates, e.g. $(\theta^2)^* = (\theta^*)^2$];
- such values are also no-bias;
- popular wisdom : "*the m.l. method is like a Rolls-Royce: expensive, but the best*".

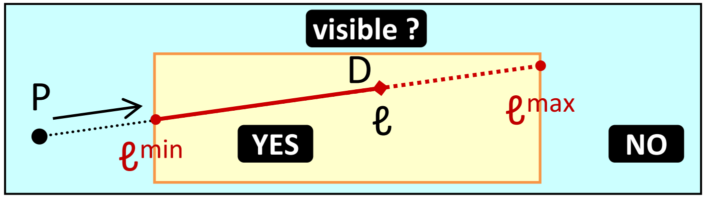
an (in)famous example, not discussed here, is the ***uniform distribution***.

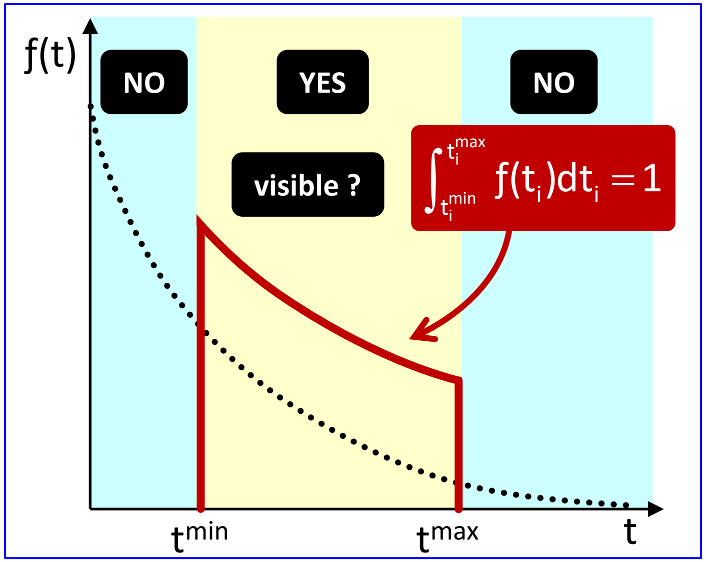
*NB. "asymptotically" means : the considered property is valid in the limit $N_{meas} \to \infty$; if N is finite, the property is NOT valid anymore; sometimes the physicists show some "lack of rigor" (say).*

# maximum likelihood: another example

A famous problem.

We observe a limited region of space (□), with N decays (D) of particles, coming from a point P, possibly external. In all events we measure $\vec{p}$, m, $\ell$, $\ell^{min}$, $\ell^{max}$ (minimum and maximum observable lengths), different in every event. Find $\tau$.




Solution

Get t $(=m\ell/|\vec{p}|)$, $t^{min,max}$ $(=m\ell^{min,max}/|\vec{p}|)$. However, $t^{min}$ and $t^{max}$ (and the pdf), are different for each event [*see figure*].

Then:

$$\int_{t_i^{min}}^{t_i^{max}} f(t)dt = 1 \to f(t) = \begin{cases} 0 & , t < t^{min} \\ \dfrac{e^{-t/\tau}/\tau}{e^{-t^{min}/\tau} - e^{-t^{max}/\tau}} & , t^{min} \le t \le t^{max} \\ 0 & , t > t^{max} \end{cases}$$

$$\ell n\Lambda = \sum_i \left[ -\ell n\tau - \frac{t_i}{\tau} - \ell n\left( e^{-t_i^{min}/\tau} - e^{-t_i^{max}/\tau} \right) \right];$$

$$\frac{\partial \ell n\Lambda}{\partial \tau} = 0 = -\frac{N}{\tau} + \frac{1}{\tau^2} \sum_i \left( t_i - \frac{t_i^{min} e^{-t_i^{min}/\tau} - t_i^{max} e^{-t_i^{max}/\tau}}{e^{-t_i^{min}/\tau} - e^{-t_i^{max}/\tau}} \right);$$

$$t_i^{max} = \infty \to N\tau = \sum_i \left( t_i - t_i^{min} \right) \to \tau = \frac{1}{N} \sum_i \left( t_i - t_i^{min} \right).$$

otherwise, if $t_i^{max} < \infty \to$ numerical solution.

# maximum likelihood: $m_H$ at LEP

Our problem: use the full LEP statistics for the search of the Higgs boson. Define:

- "channel c", c=1,...,C : (one experiment) × (one $\sqrt{s}$) × (one final state) [e.g. (L3) – ($\sqrt{s}$ = 204 GeV) – ($e^+e^- \to HZ \to b\bar{b}\mu^+\mu^-$)] (actually C > 100);
- "$m = m_H$, test mass" : the mass under study ("the hypothesis"), which must be accepted/rejected (a grid in mass, with interval ~ mass resolution);
- for each c(hannel) and each $m_H$, (in principle) a different analysis → sets of $\{\sigma_S, \sigma_B, \varepsilon_S, \varepsilon_B, \mathcal{L}\}_{c,m}$ [$s_{c,m} = \mathcal{L}\, \varepsilon_S\, \sigma_S$, $b_{c,m} = \mathcal{L}\, \varepsilon_B\, \sigma_B$, $b_{c,m} + s_{c,m} = n_{c,m}$, all $f(m_H)$];
- for each c and each $m_H$ → a set of $N_c$ candidates (= events surviving the cuts);
- call $\vec{x}_{jc}$ the variables (e.g. 4-momenta of particles) of "jc" [event j, channel c];
- [actually "jc" is a candidate for a range of $m_H$;]
- the mc samples (both signal and bckgd) allow to define $f^S_{c,m}(\vec{x})$ and $f^B_{c,m}(\vec{x})$, the pdf for signal and bckgd of all the variables, after cuts and fits;
- other variables (e.g. reconstructed masses, secondary vertex probability, ...) are properly computed for each "jc";
- for each $m_H$, define the total number of candidates $M_m \equiv \sum_c N_{c,m}$;
- notice that, generally speaking, all variables are correlated [e.g. $m_j = m_{jm} = m_j(m_H)$, because efficiency, cuts and fits do depend on $m_H$].

<u>**Then, start the statistical analysis...**</u>

# maximum likelihood: hypothesis test

- The likelihood function [PDG] is the product of the pdf for each event, calculated for the observed values;
- for searches, it is the Poisson probability for observing N events times the pdf of each single event [*see box*];
- since there are two hypotheses (b only and b+s), there are two pdf's and therefore two likelihoods;
- both are functions of the parameter(s) of the phenomenon under study (e.g. $m_H$);
- the **likelihood ratio Q** is a powerful (**<u>the most powerful</u>**) test between two hypotheses, mutually exclusive;
- the term "-2 ℓn …" is there only for convenience [both for computing and because $-2\ell n(\Lambda) \to \chi^2$ for n large];

- in the box [*see previous slide*]:
  - "c=1,…C" refers to different channels;
  - $f^{s,b}$ are the pdf (usually from mc) of the kinematical variables $\vec{x}$ for event $j_c$ :

given $f_c^s(\vec{x}),\ f_c^b(\vec{x}),\ f_c^{b+s}(\vec{x}) = \dfrac{s_c f^s(\vec{x}_c) + b_c f^b(\vec{x}_c)}{s_c + b_c}$ :

$$\Lambda_s = \Lambda_s(m_H) = \prod_{c=1}^{C} \left\{ \begin{array}{l} \mathcal{P}_{\text{poisson}}(N_c \mid b_c + s_c) \times \\ \prod_{j_c=1}^{N_c} \left[ f_c^{b+s}(\vec{x}_{j_c}) \right] \end{array} \right\};$$

$$\Lambda_b = \Lambda_b(m_H) = \prod_{c=1}^{C} \left\{ \begin{array}{l} \mathcal{P}_{\text{poisson}}(N_c \mid b_c) \times \\ \prod_{j_c=1}^{N_c} \left[ f_c^{b}(\vec{x}_{j_c}) \right] \end{array} \right\};$$

$$-2\ell nQ = -2\ell n\left(\frac{\Lambda_S}{\Lambda_B}\right) = 2\left(\ell n\Lambda_B - \ell n\Lambda_S\right).$$

# maximum likelihood: $m_H$ at LEP - formulæ

$$\Lambda_S = \prod_{c=1}^{C} \left\{ \frac{e^{-(s_c+b_c)} (s_c + b_c)^{N_c}}{N_c!} \times \prod_{j=1}^{N_c} \left[ \frac{s_c f^S(\vec{x}_{jc}) + b_c f^B(\vec{x}_{jc})}{s_c + b_c} \right] \right\} =$$

$$= \prod_{c=1}^{C} \left\{ \frac{e^{-(s_c+b_c)}}{N_c!} \times \prod_{j=1}^{N_c} \left[ s_c f^S(\vec{x}_{jc}) + b_c f^B(\vec{x}_{jc}) \right] \right\};$$

$$\Lambda_B = \prod_{c=1}^{C} \left\{ \frac{e^{-b_c} \times b_c^{\,N_c}}{N_c!} \times \prod_{j=1}^{N_c} f^B(\vec{x}_{jc}) \right\} =$$

$$= \prod_{c=1}^{C} \left\{ \frac{e^{-b_c}}{N_c!} \times \prod_{j=1}^{N_c} b_c f^B(\vec{x}_{jc}) \right\};$$

… and therefore →

[once again, remember that everything is an implicit function of the test mass $m_H$].

$$-2\ell nQ = -2\ell n\left(\frac{\Lambda_S}{\Lambda_B}\right) = -2\ell n\left( \frac{\prod_{c=1}^{C} \left\{ \frac{e^{-(s_c+b_c)}}{N_c!} \times \prod_{j=1}^{N_c} \left[ s_c f^S(\vec{x}_{jc}) + b_c f^B(\vec{x}_{jc}) \right] \right\}}{\prod_{c=1}^{C} \left\{ \frac{e^{-b_c}}{N_c!} \times \prod_{j=1}^{N_c} b_c f^B(\vec{x}_{jc}) \right\}} \right) =$$

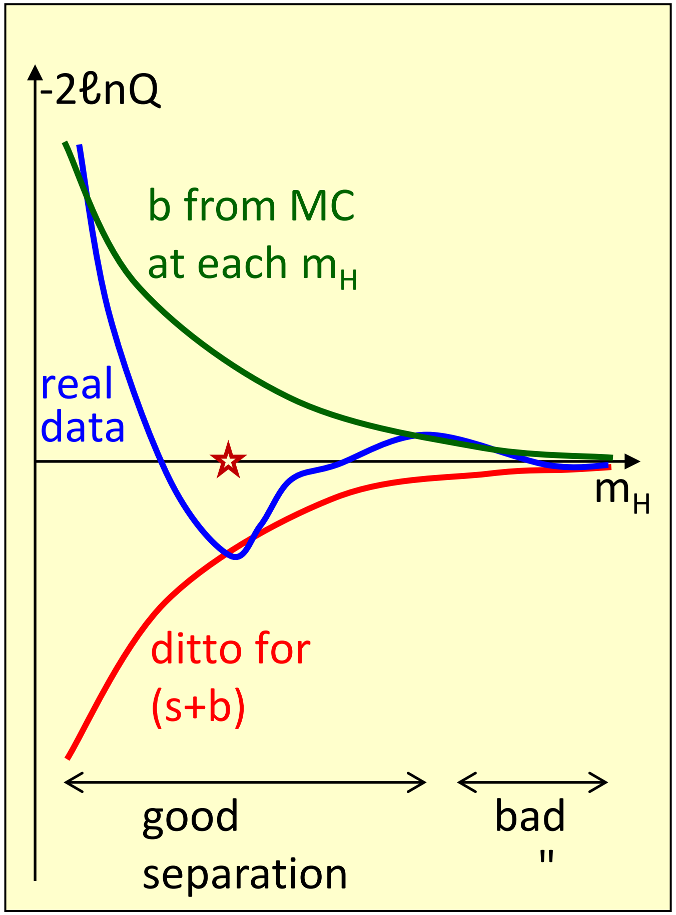
$$= 2\sum_{c=1}^{C} s_c - 2\sum_{c=1}^{C} \left[ \sum_{j=1}^{N_c} \ell n\left( 1 + \frac{s_c f^S(\vec{x}_{jc})}{b_c f^B(\vec{x}_{jc})} \right) \right].$$

# interpretation of results: discovery plot

- the likelihood is expected to be larger when the correct pdf is used;
- then the result of the test can be easily guessed (and translated into $\chi^2$):

  $-2\ \ell n\ Q = -2\ \ell n(\Lambda_s/\Lambda_b) \approx \chi_s^2 - \chi_b^2$

| | b true | (s+b) true |
|---|---|---|
| $\Lambda_b$ | +, large | +, small |
| $\Lambda_s$ | +, small | +, large |
| $\Lambda_s/\Lambda_b$ | << 1 | >> 1 |
| $\ell n(\Lambda_s/\Lambda_b)$ | –, large | +, large |
| $-2\ell nQ$ | +, large | –, large |




the plot is a little cartoon of an ideal situation (e.g. Higgs search at LEP2), that never happened :

- the cross-section decreases when $m_H$ increases → for large $m_H$ no discovery.
- ~~look the blue line → discovery !!!~~

unfortunately, for the H at LEP **it did NOT happen**

# interpretation of results: parameter μ

- put : $\sigma^{exp} = \sigma^b + \mu\, \sigma^s$
  [$\sigma^{b,s}$ include $\varepsilon^{b,s} \to n = b + \mu\, s$];
- plot : horizontal : $m_H$.
  vertical : $\mu$ [$=(\sigma^{exp}-\sigma^b)/\sigma^s$];

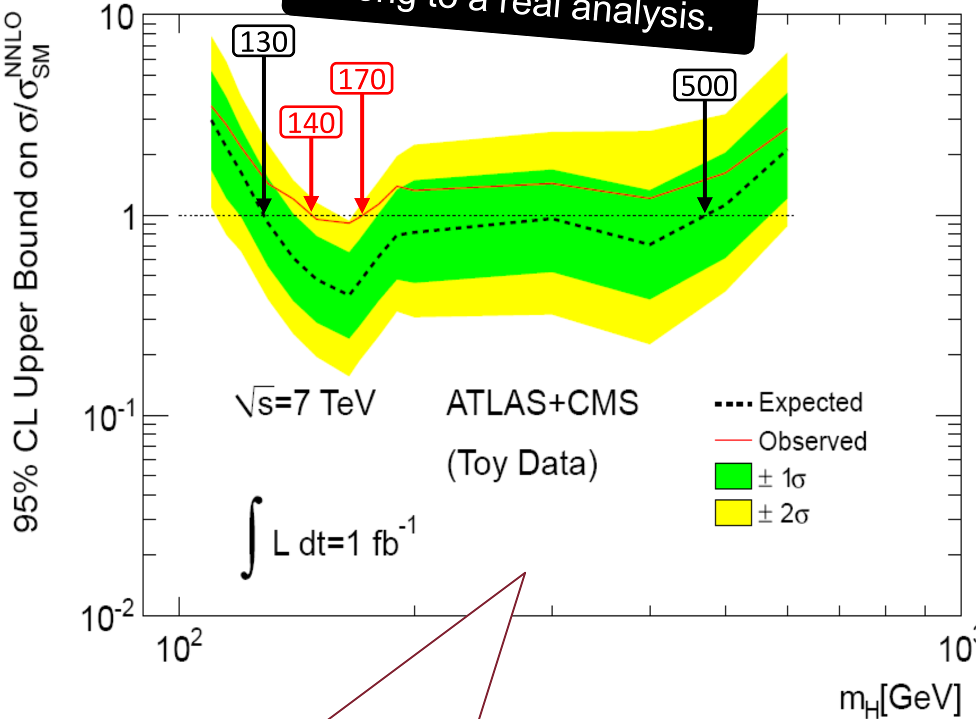
➤ the lines show, with a given $\mathscr{L}_{int}$ and analysis, the expected limit (--), and the actual observed limit (–), i.e. the μ value excluded at 95% CL;

➤ the band ◆ shows the fluctuations at ±1σ (≈ 68%) of the "bckgd only" hypothesis; the band ◆ is the same for ±2σ (≈ 95%).

---

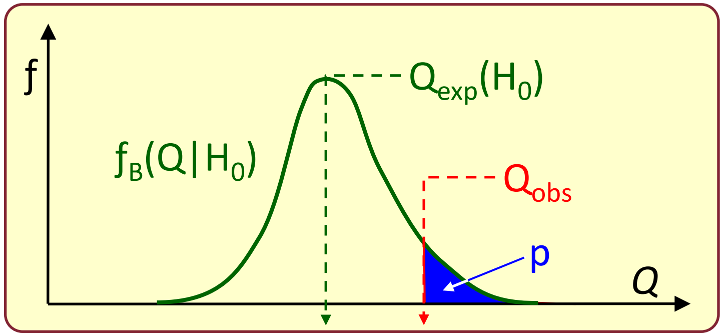
- the case $\mu \neq 0,1$ has no well-defined physical meaning (= *a theory identical to the SM, but with a scaled cross section*);
- if the lines are at $\mu > 1$, the distance respect to $\mu$=1 reflects the $\mathscr{L}_{int}$ necessary to get the limit in the SM.

the plot comes from a toy simulation, it does NOT belong to a real analysis.




in this hypothetical case, the region 140 < $m_H$ < 170 GeV is excluded at 95% CL, while the expected limit was 130÷500 GeV (either bad luck or hint of discovery).

# interpretation of results: p-value



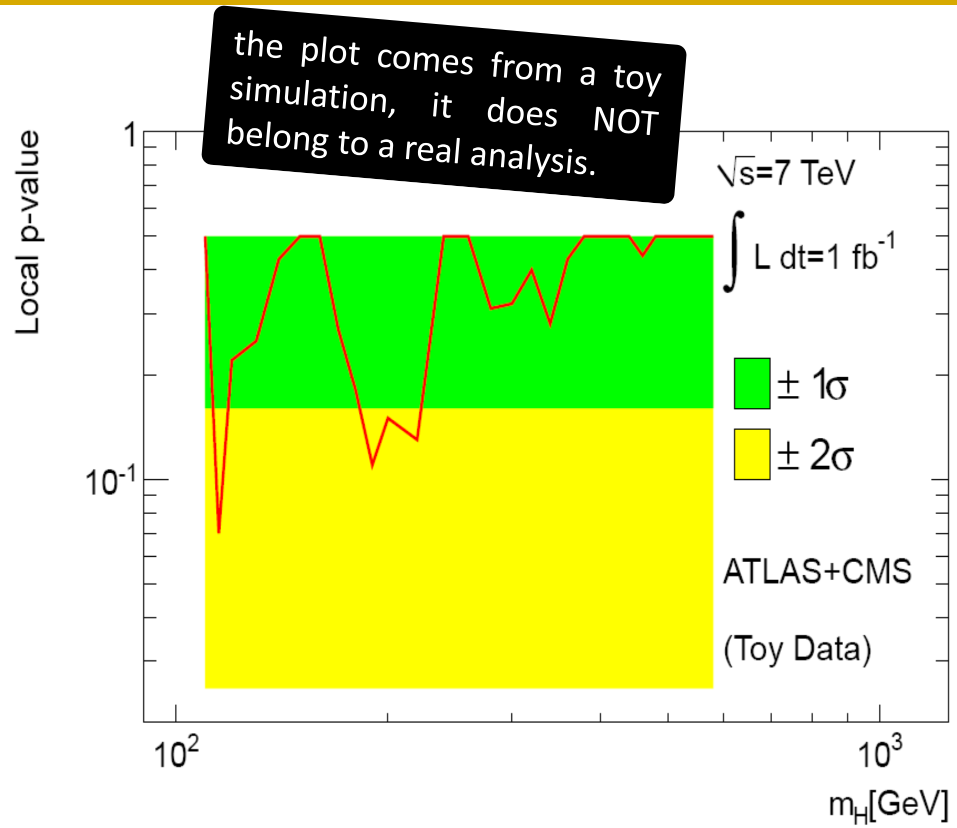

$$p \equiv \int_{x_{obs}}^{\infty} f(x\,|\,H_0)dx$$

- the "p-value" is the probability to get the same result or another less probable, in the hypothesis of bckgd only.
- x = "statistics" (e.g. likelihood ratio);
- $H_0$ = "null hypothesis" (i.e. bckgd only);

i.e.

p small → $H_0$ NOT probable

→ discovery !!!

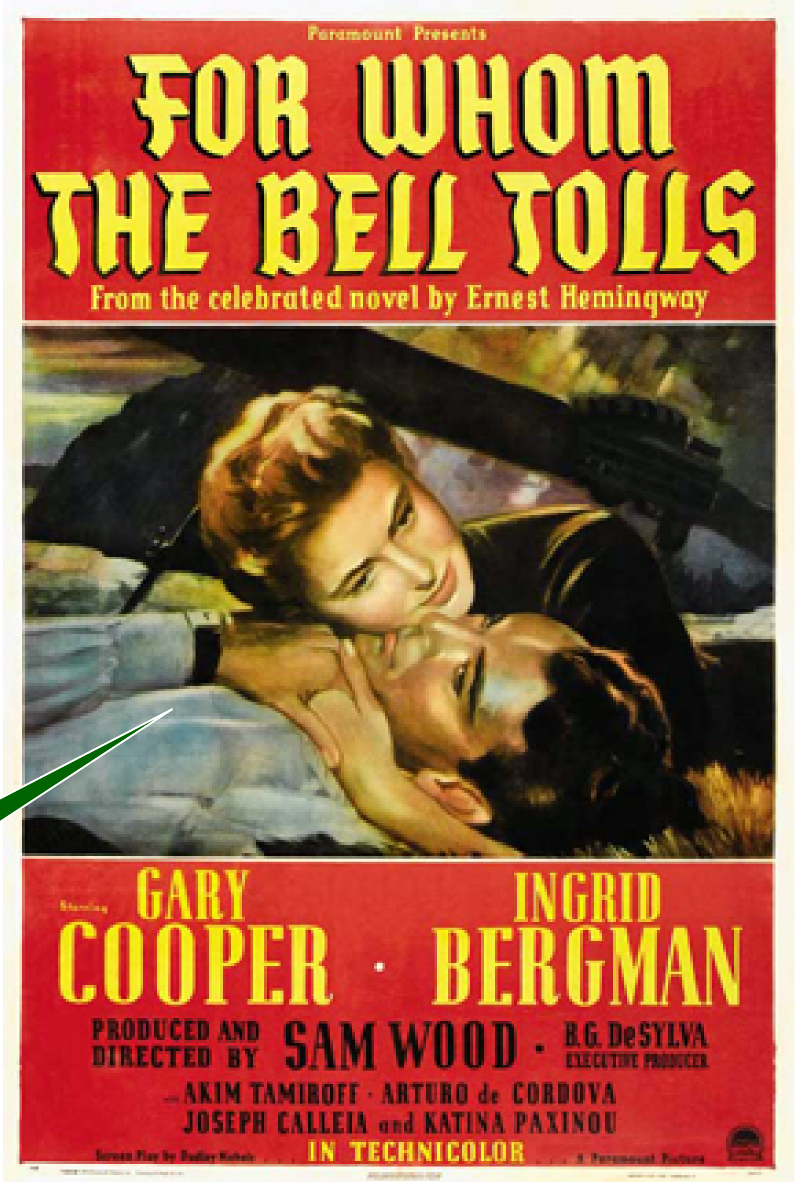



- vertical : *p*-value;
- horizontal : $m_H$.
- the band ♦ (♦) shows the fluctuations at 1σ (2σ).

NB the discovery corresponds to the red line below 5σ (or $2.86\times10^{-7}$), not shown in this fake plot.

# References

1. classic textbook : Eadie et al., Statistical methods in experimental physics;
2. modern textbook : Cowan, Statistical data analysis;
3. simple, for experimentalists : Cranmer, Practical Statistics for the LHC, [arXiv:1503.07622v1]; (also CERN Academic Training, Feb 2-5, 2009);
4. bayesian : G.D'Agostini, YR CERN-99-03.
5. statistical procedure for Higgs @ LEP2 : A.L.Read, J. Phys. G: Nucl. Part. Phys. 28 (2002) 2693;
6. mass limits : R.Cousins, Am.J.Phys., 63 (5), 398 (1995).
7. [PDG explains everything, but very concise]




bells are related to dramatic events even outside particle physics

Sapienza Università di Roma

# End of chapter 4