Statistical Methods used for Higgs Boson Searches

Roger Wolf

16. June 2016

INSTITUTE OF EXPERIMENTAL PARTICLE PHYSICS (IEKP) – PHYSICS FACULTY

Schedule for today

- What is the meaning of the degrees of freedom of the χ^2 function?
- What is the relation between the likelihood function and the χ^2 estimate?

3. *p*-value, significance and limit setting

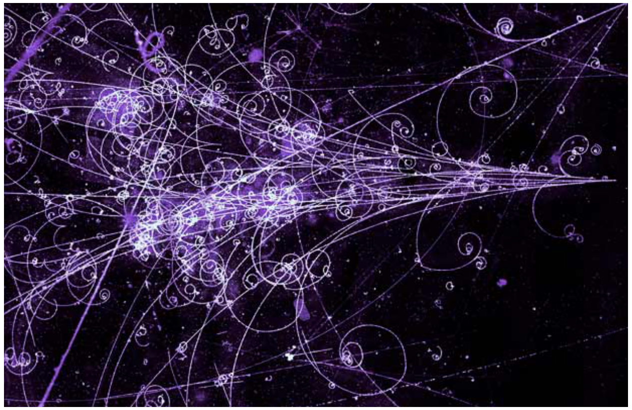
2. Parameter estimates

1. Likelihood analyses

Statistics vs. particle physics

Experiment:

- All measurements we do are derived from rate measurements.
- We record millions of trillions of particle collisions.
- Each of these collisions is independent from all the others.



Theory:

- QM wave functions are interpreted as probability density functions.
- The Matrix Element, $\mathcal{S}_{fi}$,gives the probability to find final state f for given initial state i.
- Each of the statistical processes $pdf \rightarrow ME \rightarrow hadronization \rightarrow$ $energy\ loss\ in\ material \rightarrow digitization$ are statistically independent.
- Event by event simulation using Monte Carlo integration methods.

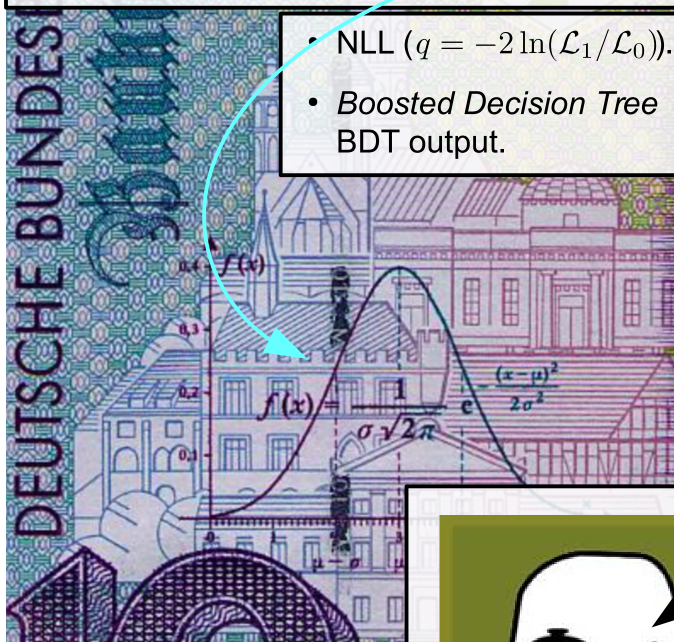
- Particle physics experiments are a perfect application for statistical methods.

Statistics vs. probability theory (stochastic)

Test statistic:

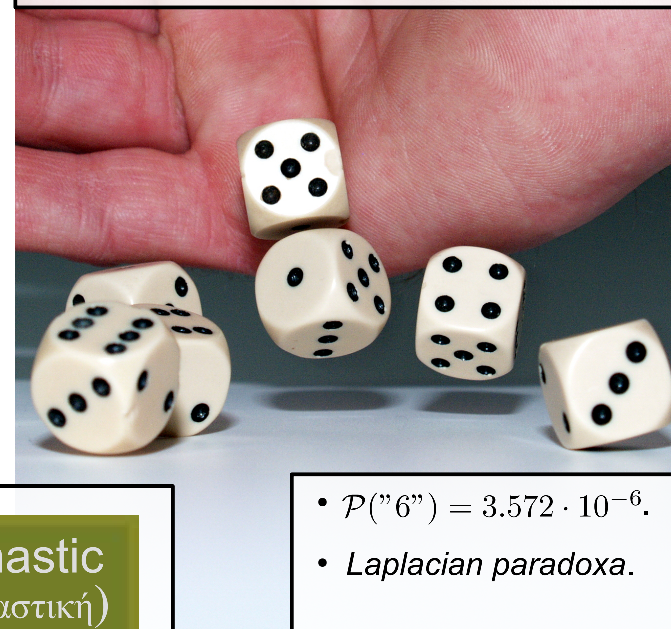
$\Omega^n \to \mathbb{R}: \qquad x \to f(x)$

- NLL ($q = -2\ln(\mathcal{L}_1/\mathcal{L}_0)$).
- *Boosted Decision Tree* BDT output.

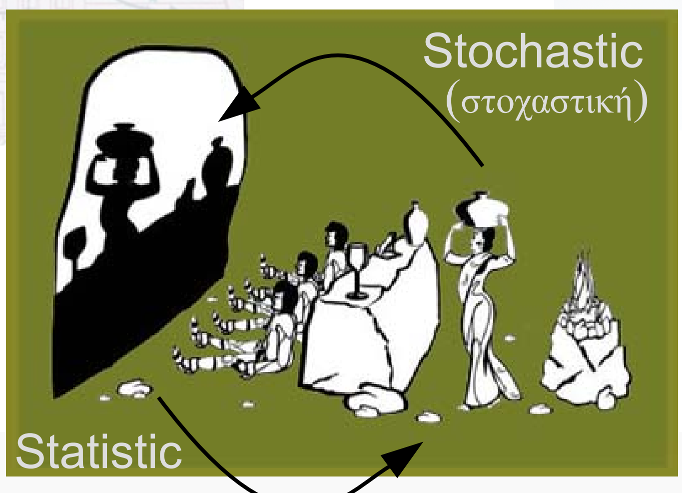


Probability (density) function:

$\Omega^n \to [0,1] \subset \mathbb{R}: \qquad x \to \mathcal{P}(x)$



- $\mathcal{P}(\text{"6"}) = 3.572 \cdot 10^{-6}$.
- *Laplacian paradoxa*.



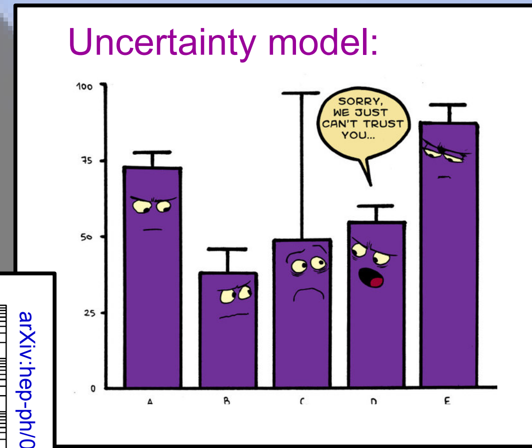
- Problem of statistics is usually *ill-defined*.
- Deduce *truth* from shadows in Platon's cave...

The case of “truth”

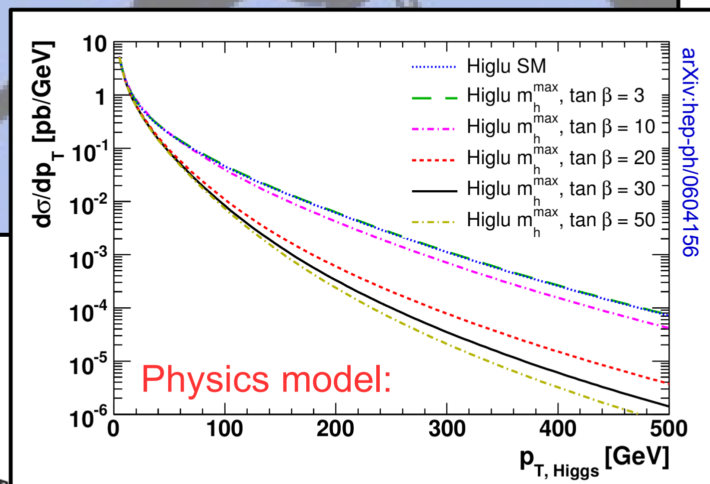
- Deduce *truth* from shadows:

  Usually phrased in form of (nested) models (=*ideas* for Platon):

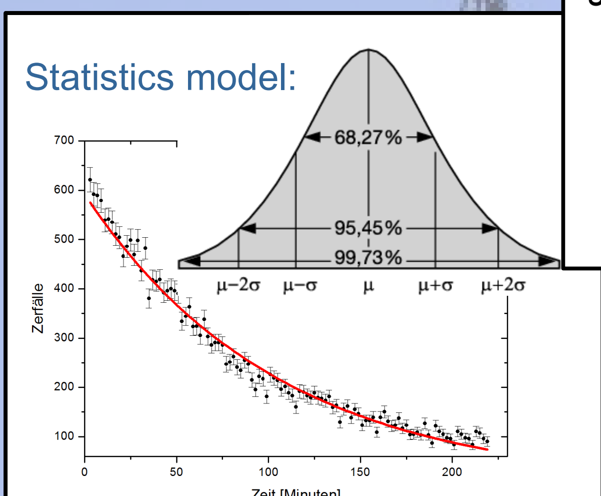
- Mathematically model = hypothesis.



Usually determined to best knowledge (not questioned)



Usually competing models/ hypotheses will be discussed here!



Usually not questioned

Probability distributions

	Expectation:	Variance:
$$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$

Probability distributions

	Expectation:	Variance:
$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$ (Gaussian distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$n \to \infty$, p fixed ↑ Central limit theorem of *de Moivre* & *Laplace*.		
$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$

Probability distributions

Distribution	Expectation:	Variance:
$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$ (*Gaussian* distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \frac{(np)^k}{k!} e^{-np}$ (*Poisson* distribution)	$\mu = np$	$\sigma^2 = \mu = np$

From Binomial to Gaussian (↑): $n \to \infty$, p fixed

Central limit theorem of *de Moivre* & *Laplace*.

From Binomial to Poisson (↓): $n \to \infty$, np fixed

Will be shown on next slide.

Binomial ↔ *Poisson* distribution

$$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}{k!} \cdot \frac{\mu^k}{n^k} \cdot \frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^k}$$

$$= \frac{1\cdot(1-\frac{1}{n})(1-\frac{2}{n})\cdot\ldots\cdot(1-\frac{k-1}{n})}{\left(1-\frac{\mu}{n}\right)^k} \cdot \frac{\mu^k}{k!} \cdot \left(1-\frac{\mu}{n}\right)^n$$

$$= \underbrace{\frac{1}{\left(1-\frac{\mu}{n}\right)} \cdot \frac{(1-\frac{2}{n})}{\left(1-\frac{\mu}{n}\right)} \cdot \frac{(1-\frac{2}{n})}{\left(1-\frac{\mu}{n}\right)} \cdot \ldots \cdot \frac{(1-\frac{k-1}{n})}{\left(1-\frac{\mu}{n}\right)}}_{\to 1} \cdot \frac{\mu^k}{k!} \cdot \underbrace{\left(1-\frac{\mu}{n}\right)^n}_{\to e^{-\mu}}$$

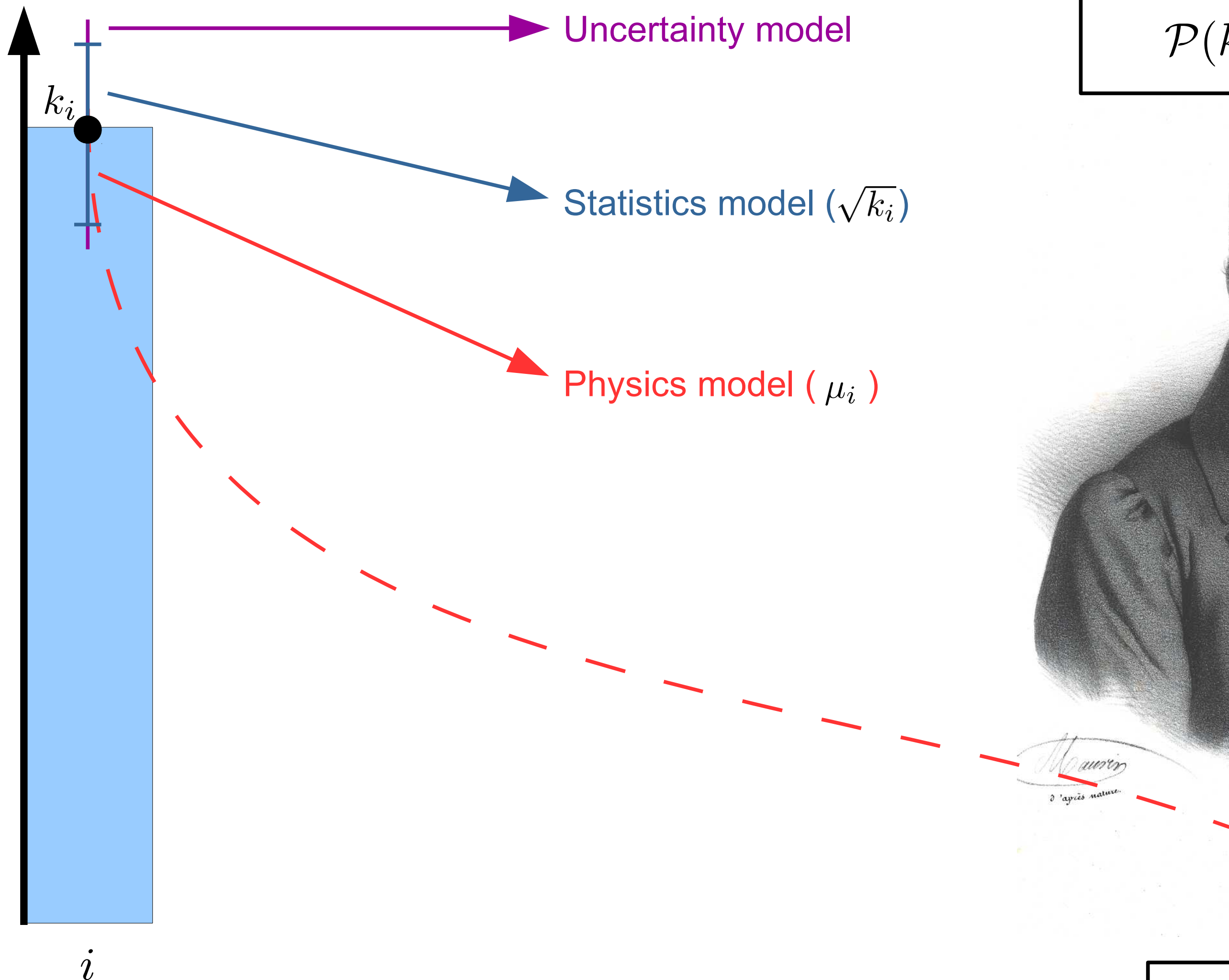
$$= \frac{\mu^k}{k!} e^{-\mu}$$

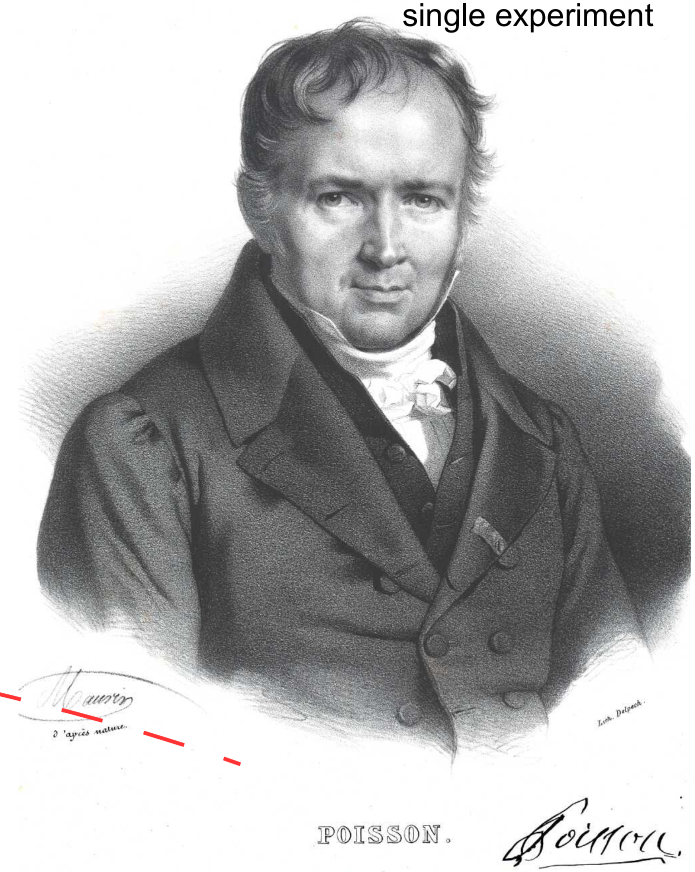
$$\mu = const, n \to \infty$$

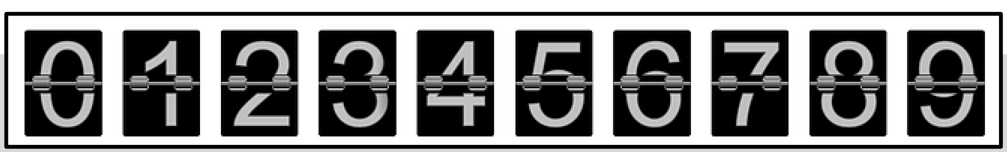
Models for counting experiments

$$\mathcal{P}(k_i, \mu_i) = \frac{\mu_i^{k_i}}{k_i!} e^{-\mu_i}$$

single experiment





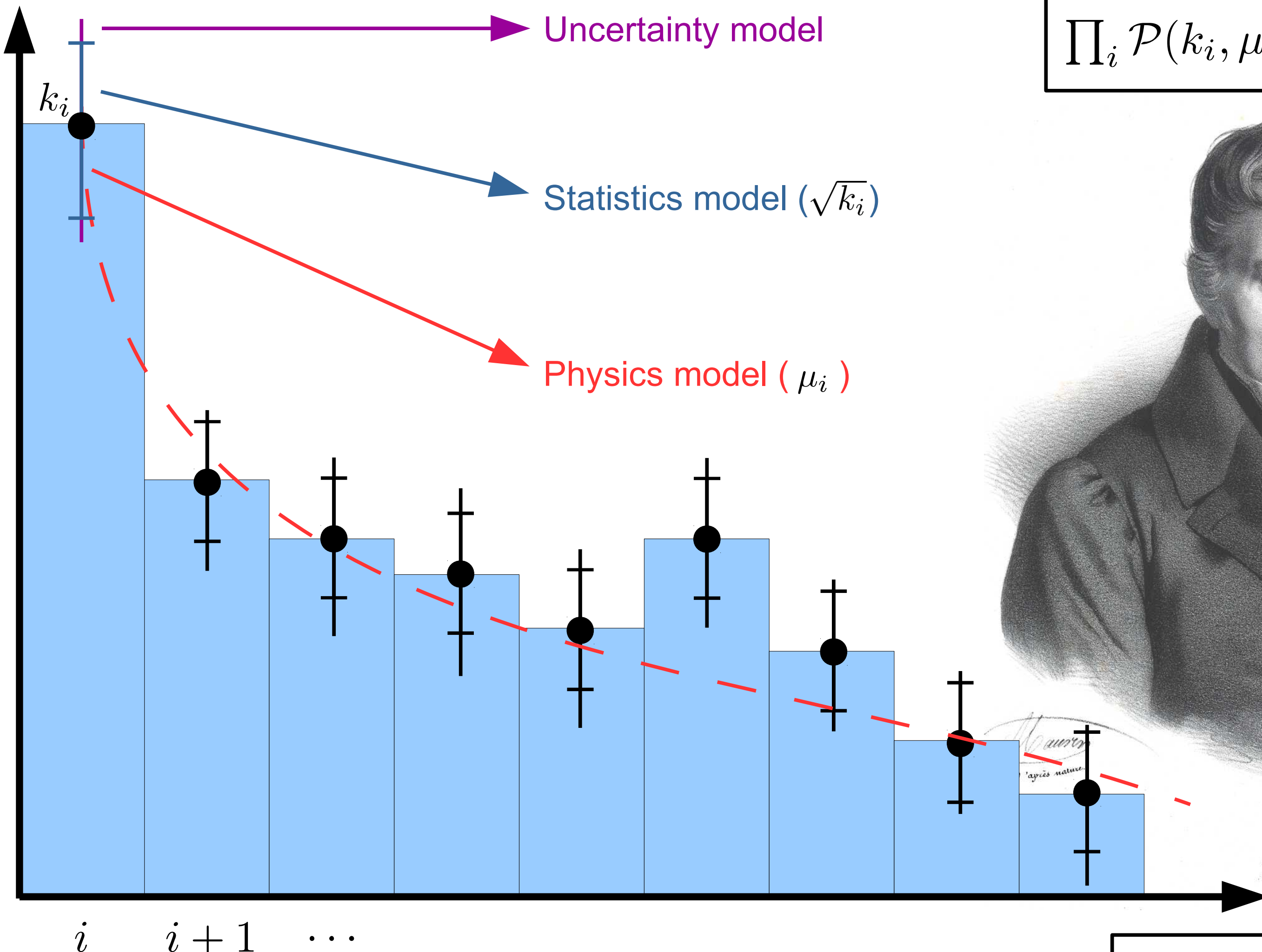


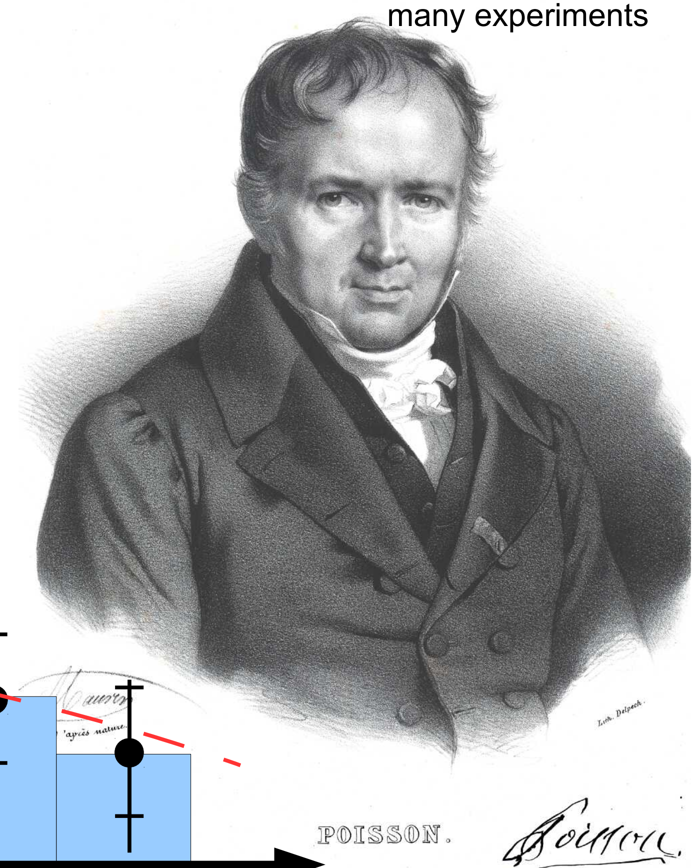
Siméon Denis Poisson
(21.07.1781 – 25.04.1840)

Models for counting experiments

$$\prod_i \mathcal{P}(k_i, \mu_i) = \prod_i \frac{\mu_i^{k_i}}{k_i!} e^{-\mu_i}$$

many experiments





Siméon Denis Poisson
(21.07.1781 – 25.04.1840)

Model building (likelihood functions)

$$\prod_i \mathcal{P}(k_i, \mu_i) = \prod_i \frac{\mu_i^{k_i}}{k_i!} e^{-\mu_i}$$

- Likeliness of a model to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\}, \{\kappa_j\})$.

$\{\kappa_j\}$ → model parameters.

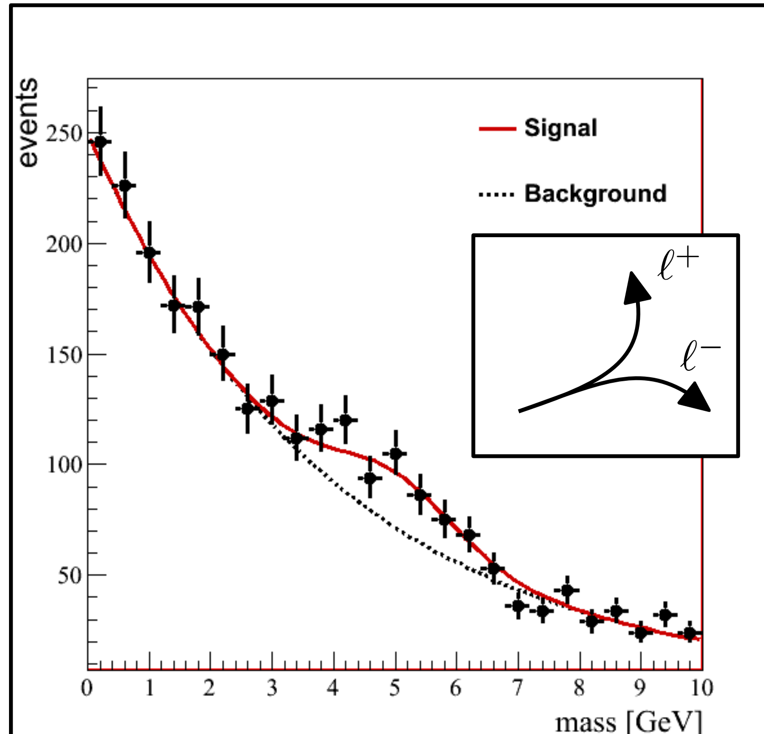
$\{k_i\}$ → measured number of events (e.g. in bins i).

- Simple example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Model building (likelihood functions)

- Likeliness of a model to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\},\{\kappa_j\})$.

$$\prod_i \mathcal{P}(k_i,\mu_i) = \prod_i \frac{\mu_i^{k_i}}{k_i!} e^{-\mu_i}$$

$\{\kappa_j\}$ → model parameters.

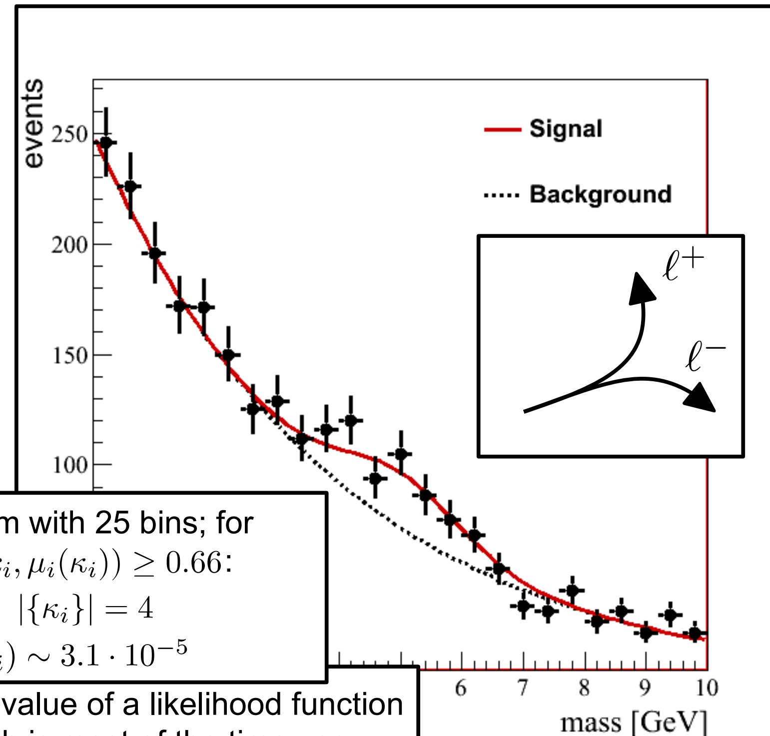
$\{k_i\}$ → measured number of events (e.g. in bins i).

- Simple example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i,\mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



EX: histogram with 25 bins; for each bin $\mathcal{P}(k_i,\mu_i(\kappa_i)) \geq 0.66$:
$|\{k_i\}| = 25 \quad |\{\kappa_i\}| = 4$
$\prod \mathcal{P}(k_i,\mu_i(\kappa_i)) \sim 3.1 \cdot 10^{-5}$

NB: a value of a likelihood function as such is most of the time very close to zero, and w/o a reference in general w/o further meaning.

Distinguishing models (*likelihood ratio*)

- Task of likelihood analyses:

  do not determine likelihood of an experimental outcome per se, but distinguish models (=hypotheses) and determine the one that explains the experimental outcome best.

Fundamental lemma of Neyman-Pearson:

when performing a test between two simple hypotheses H_1 and H_0 the *likelihood ratio test*, which rejects H_0 in favor of H_1 when

$$Q = \frac{\mathcal{L}_{H_1}(\{k_i\},\{\kappa_i\})}{\mathcal{L}_{H_0}(\{k_i\},\{\kappa_i\})} \leq \eta$$

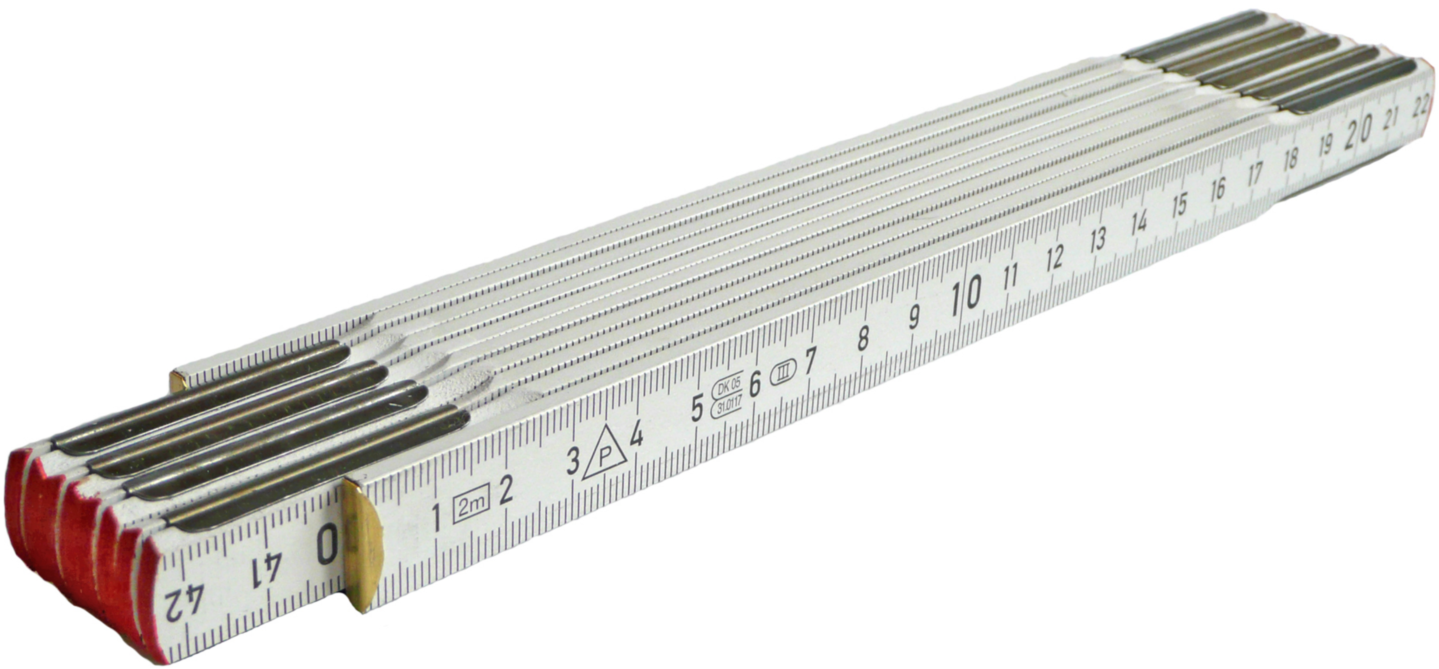
$$\mathcal{P}(Q(\{k_i\},\{\kappa_i\}) \leq \eta | H_i) = \alpha$$

is the most powerful test at significance level α for a threshold η.

- For $q = -2\ln Q$ this ratio turns into a difference ($\Delta\mathrm{NLL}$).

This is usually the *test statistic* of choice!

Parameter estimates



Distinguish best parameter (set) in discrete or continuous transformations.

Maximum likelihood fit

Karlsruhe Institute of Technology

- Each likelihood (ratio of) function(s) (with one or more parametric model part(s)) can be subject to a maximum likelihood fit (**NB:** negative log-likelihood finds its minimum where the log-likelihood is maximal...).

(see next slides)

Minimization problem as known from school.

In our example e.g. four parameters κ_i.

Parameters can be constrained or unconstrained

- Simple example:
signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i,\mu_i(\kappa_j))}$$

Product for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$

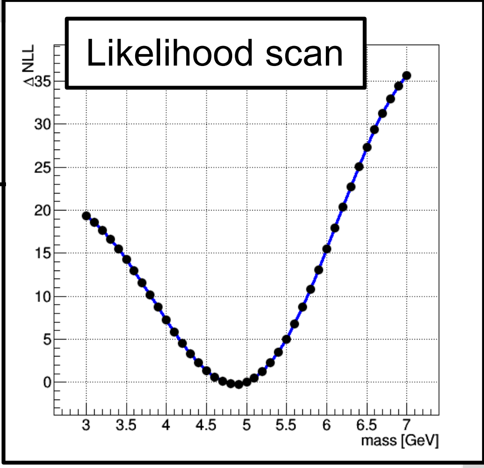
The ATLAS+CMS Higgs couplings combined fit has $\mathcal{O}(4250)$ parameters and up to seven POI's.

The CMS Tracker Alignment problem has $\mathcal{O}(50'000)$ parameters and several thousand POI's.

Parameter(s) of interest (POI)

NB: this is a likelihood ratio on its own.

NB: I've also made the scan based on a likelihood ratio.

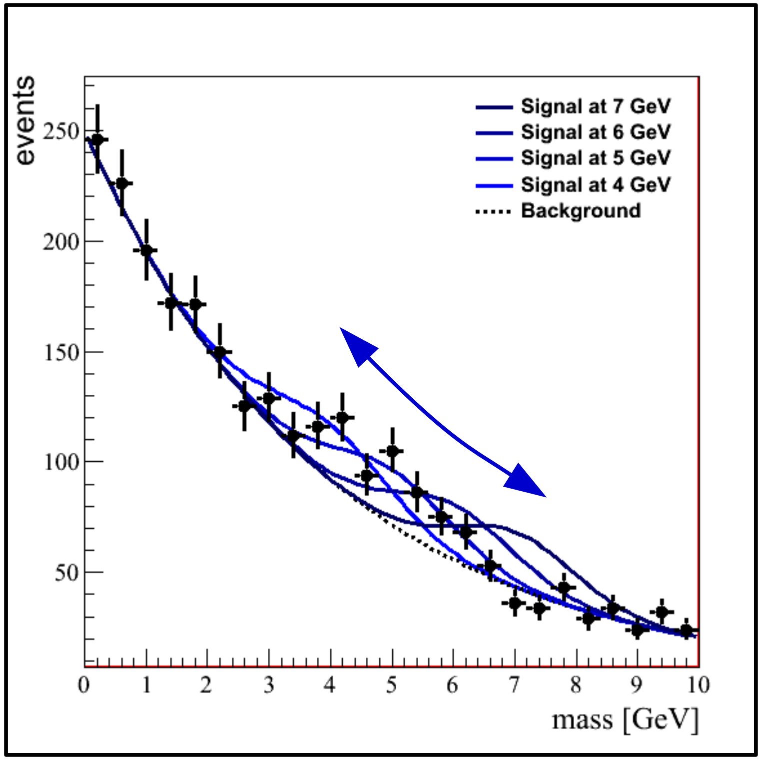


- In a maximum likelihood fit each case/problem defines its own *parameter(s) of interest* (POI's):
  - POI could be the mass (κ_3).

- Simple example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}_{\text{Product for each bin (Poisson).}}$$

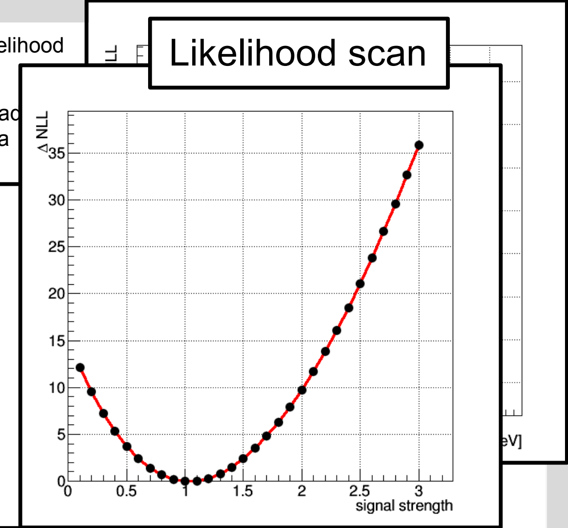
$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Parameter(s) of interest (POI)

NB: this is a likelihood ratio on its own

NB: I've also mad[e a] scan based on a likelihood ratio.



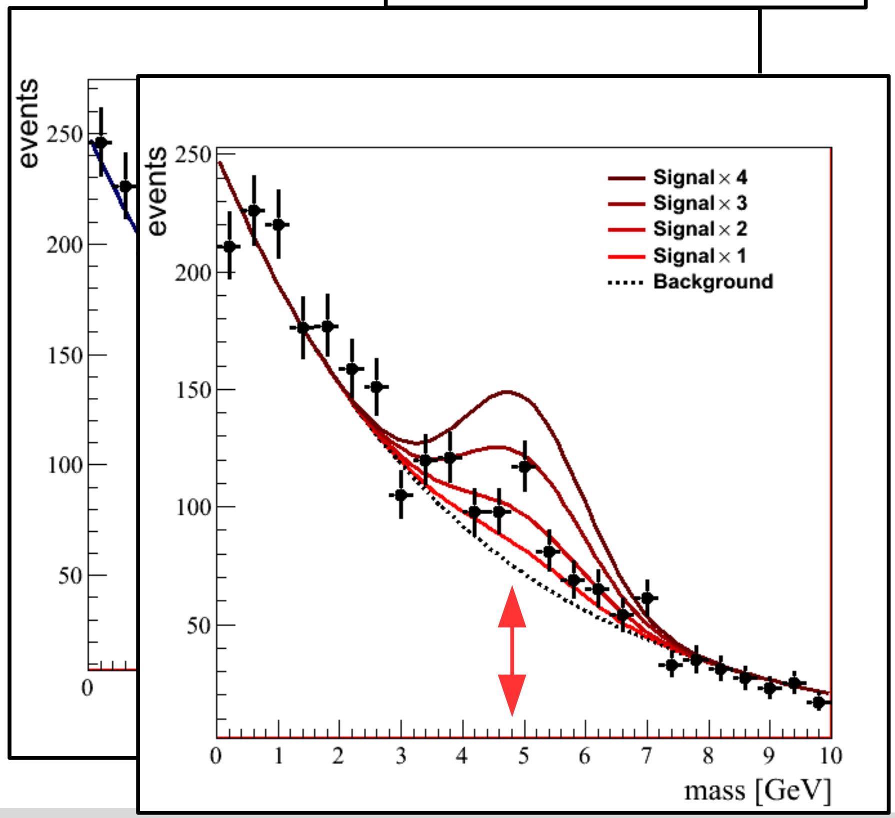
- In a maximum likelihood fit each case/problem defines its own *parameter(s) of interest* (POI's):
  - POI could be the mass (κ_3).
  - In our case POI usually is the signal strength (κ_2) (for a fixed value for κ_3).

- Simple example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i,\mu_i(\kappa_j))}$$

Product for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Incorporation of systematic uncertainties

- Systematic uncertainties are usually incorporated in form of *nuisance parameters*:
  - E.g. background normalization κ_0 not precisely known, but with uncertainty $\sigma(\kappa_0)$:

$$\mu_i(\kappa_j) = \kappa_0 \mathcal{P}'(x, 1, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}$$

$\sigma(\kappa_0)$ → uncertainty

$x, 1$ → possible values of single "measurements" (integrated out)

κ_0 → expected value/best knowledge

- Simple example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$

Incorporation of systematic uncertainties

- Systematic uncertainties are usually incorporated in form of *nuisance parameters*:
  - E.g. background normalization κ_0 not precisely known, but with uncertainty $\sigma(\kappa_0)$:

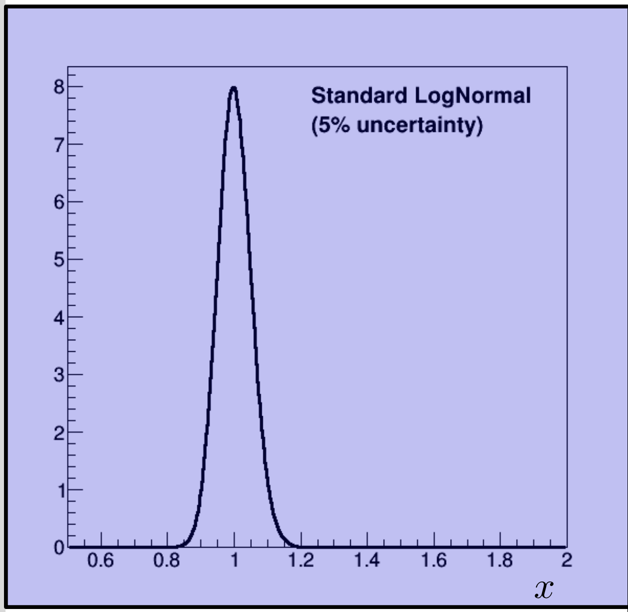
$$\mu_i(\kappa_j) = \kappa_0 \mathcal{P}'(x, 1, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}$$

$\sigma(\kappa_0)$ → uncertainty

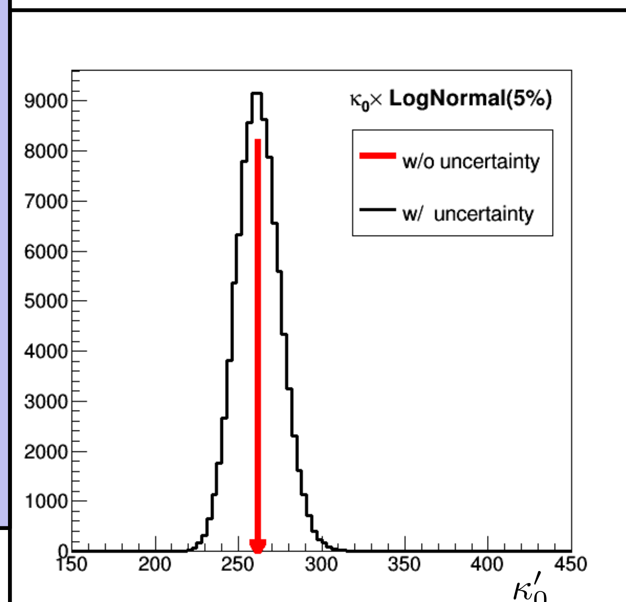
1 → possible values of single "measurements" (integrated out)

κ_0 → expected value/best knowledge

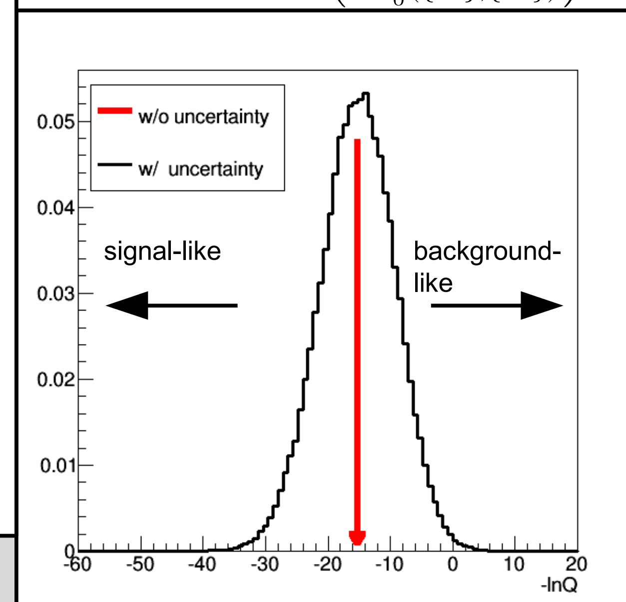
Probability density function ($\mathcal{P}$)



Effect on BG normalization



$$-\ln Q = -\ln\left(\frac{\mathcal{L}_{H_1}(\{k_i\},\{\kappa_i\})}{\mathcal{L}_{H_0}(\{k_i\},\{\kappa_i\})}\right)$$



Relations between probability distributions

$$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

$$\left(-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$$

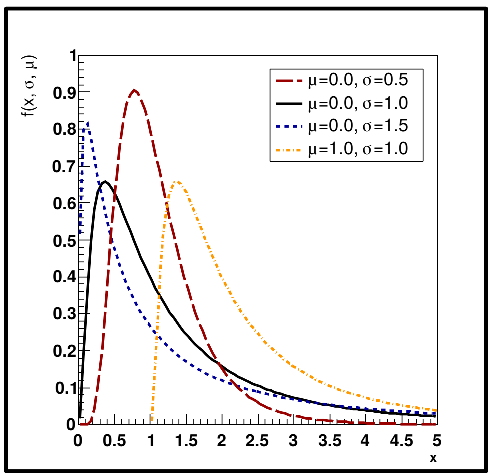
Central Limit Theorem:

Random variable variable made up of a sum of many single measurements.

Log-normal ← log ← **Gaussian** → log → **χ^2 Distribution**

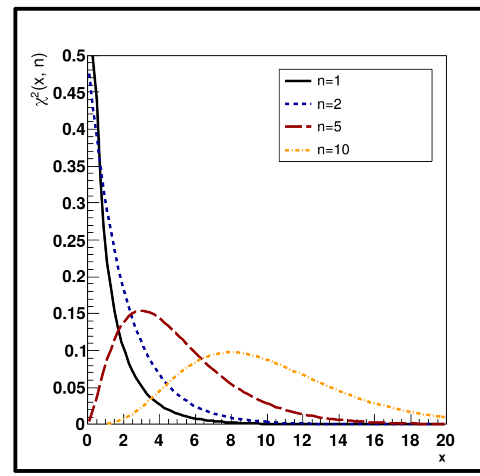
Random variable variable made up of a product of many single measurements.

What does the parameter k correspond to in the χ^2 distributions?



Binomial → $n \to \infty, p = cont$ → **Gaussian**

Binomial → $n \to \infty, np = cont$ → **Poisson**



Look for something that is very rare very often.

Relations between probability distributions

$$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

$$\left(-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$$

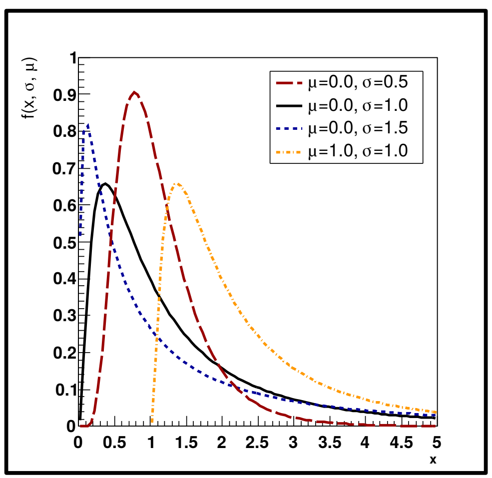
Central Limit Theorem:

Random variable variable made up of a sum of many single measurements.

Log-normal ← log ← **Gaussian** → log → **χ^2 Distribution**

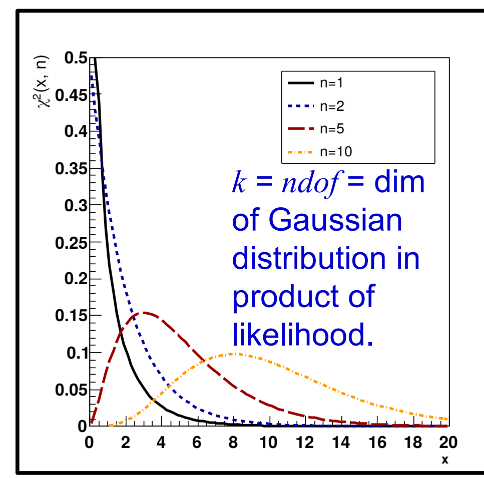
Random variable variable made up of a product of many single measurements.

What does the parameter k correspond to in the χ^2 distributions?



Binomial → $n \to \infty, p = cont$ → **Gaussian**

Binomial → $n \to \infty, np = cont$ → **Poisson**



Look for something that is very rare very often.

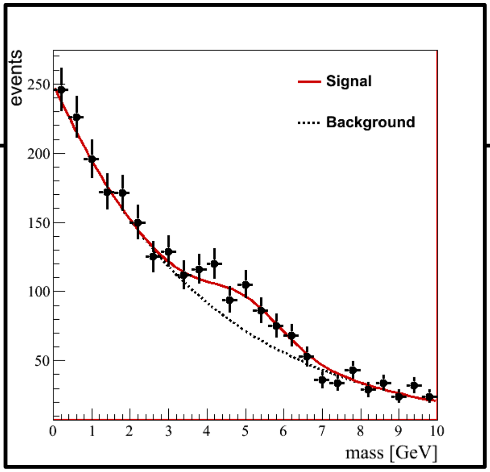
Example: saturated model

- Example of a likelihood ratio:

$$q_\lambda = -2\ln\left(\frac{\mathcal{L}(\text{data}|_{\text{test}})}{\mathcal{L}(\text{data}|_{\text{saturated}})}\right)$$

Model to be tested.

Model w/ as many parameters, λ_j, as measurements.



e.g. one shape for each bin.

- Special case: (i) histogram; (ii) no further nuisance parameters; (iii) uncertainties normal distributed:

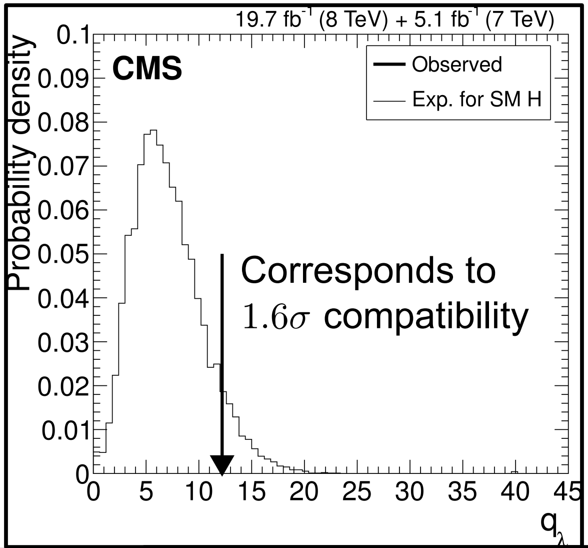
$$\mathcal{L}(\text{data}|_{\text{test}}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} e^{-(d_i-\lambda_i)^2/2\sigma_i}$$

$$\mathcal{L}(\text{data}|_{\text{saturated}}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}}$$

$$q_\lambda = -2\ln\left(\frac{\mathcal{L}(\text{data}|_{\text{test}})}{\mathcal{L}(\text{data}|_{\text{saturated}})}\right) = \sum_i \frac{(d_i-\lambda_i)^2}{\sigma_i}$$

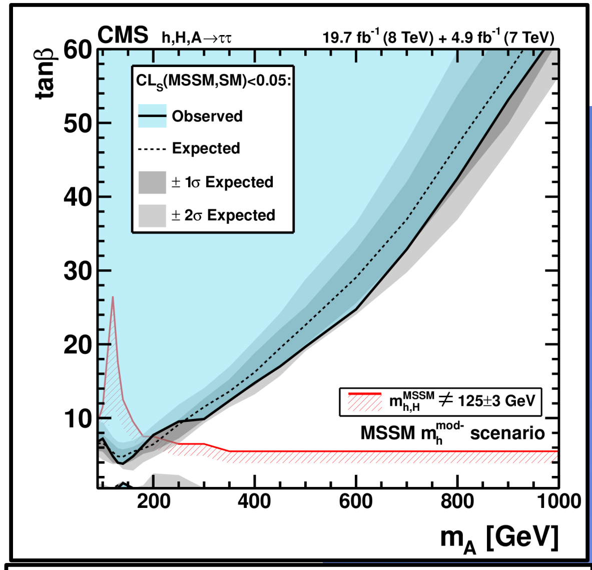
Generalization of the χ^2 test.

- General case: (i) many histograms; (ii) many nuisance parameters:



CL of interest: $\int_{q_{\text{obs}}}^{+\infty} \mathcal{P}_{\text{test}}$

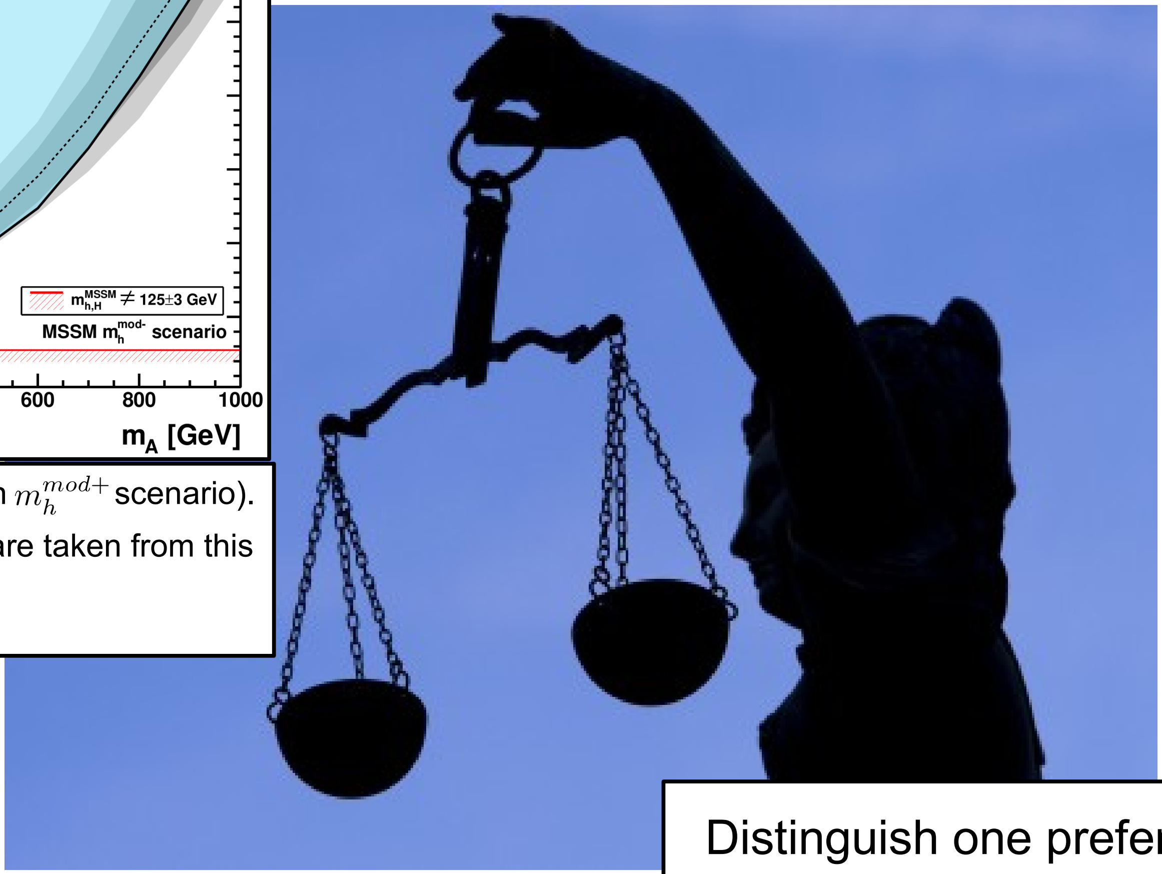
Hypothesis testing



Full exclusion (here in m_h^{mod+} scenario).

All further examples are taken from this very publication:

PRL 106 (2011) 231801



Distinguish one preferred hypothesis (H_0) against alternative hypotheses, in general in discrete but in special cases also in continuous transformations.

Example: test statistics (LEP ~2000)

- Test signal (H_1, for fixed mass, m, and fixed signal strength, μ) vs. background-only (H_0).

pdf's for nuisance parameters modified according to Bayes theorem.

$$\mathcal{L}(n|\,b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$\mathcal{L}(n|\,\mu s(\kappa_j)+b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j)+b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(\,n|\mu s+b)}{\mathcal{L}(\,n|b)}\right), \quad 0 \leq \mu$$

nuisance parameters $\tilde{\kappa}_j$ integrated out before evaluation of q_μ (→marginalization).

Example: test statistics (Tevatron ~2005)

- Test signal (H_1, for fixed mass, m, and fixed signal strength, μ) vs. background-only (H_0).

pdf's for nuisance parameters modified according to Bayes theorem.

$$\mathcal{L}(n|\,b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$\mathcal{L}(n|\,\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(n|\mu s(\hat{\kappa}_\mu)+b(\hat{\kappa}_\mu))}{\mathcal{L}(n|b(\hat{\kappa}_{\mu=0}))}\right), \quad 0 \leq \mu$$

→ profiling.

nominator maximized for given μ before marginalization. Denominator for $\mu = 0$. Better estimates of nuisance parameters w/ reduced uncertainties.

Example: test statistics (LHC ~2010)

- Test signal (H_1, for fixed mass, m, and fixed signal strength, μ) vs. background-only (H_0).

profile likelihood (→ *Feldman-Cousins* test statistic).

$$\mathcal{L}(n|\,b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j))$$

$$\mathcal{L}(n|\,\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j))$$

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(n|\mu s(\hat{\kappa}_\mu)+b(\hat{\kappa}_\mu))}{\mathcal{L}(n|\hat{\mu} s(\hat{\kappa}_{\hat{\mu}})+b(\hat{\kappa}_{\hat{\mu}}))}\right), \quad 0 \leq \hat{\mu} \leq \mu$$

→ profiling.

nominator maximized for given μ before marginalization. For the denominator a global maximum is searched for at $\hat{\mu}$. In addition allows use of asymptotic formulas (→ no more toys needed!(*)).

(*) will not be discussed further here.

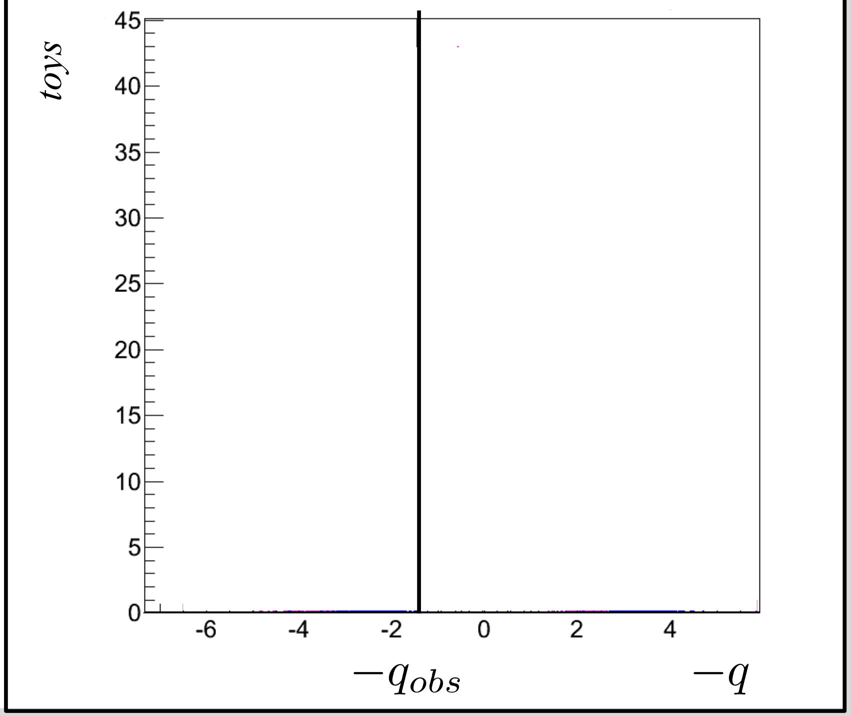
Test statistic in life

- From the evaluation of the test statistic on data always obtain a plain value q_{obs} (in our discussion: $q_{obs} < 0$ – signal-like; $q_{obs} > 0$ – background-like).
- $\rightarrow$ True outcome of the experiment (nuisance parameters estimated to best knowledge, no uncertainties involved here)!

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$

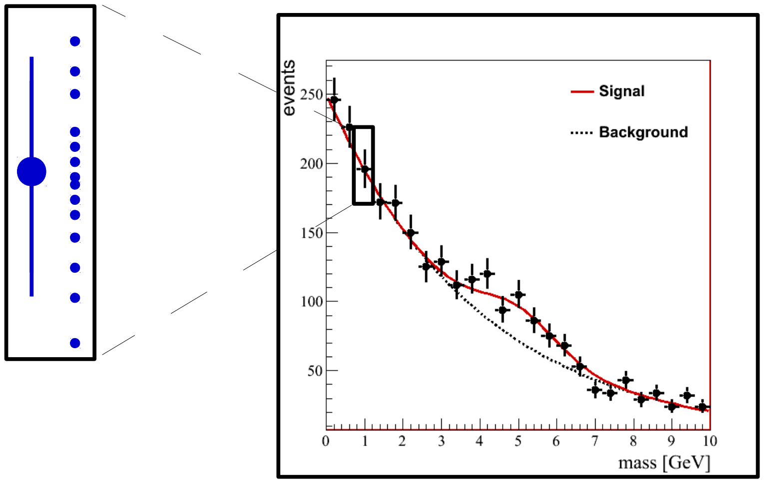


Meaning and interpretation of the test statistic

- How compatible is q_{obs} with H_0 or H_1? For this evaluate the test statistic on large number of toy experiments based on H_0 or H_1 .

- Determine *toy dataset*.
- Determine *toy values* for all uncertainties.
- Determine value of $-q$ for each toy.
- Proceed as often as possible; do this for H_0 & H_1.

$\mu_i(\hat{\kappa}_j)$

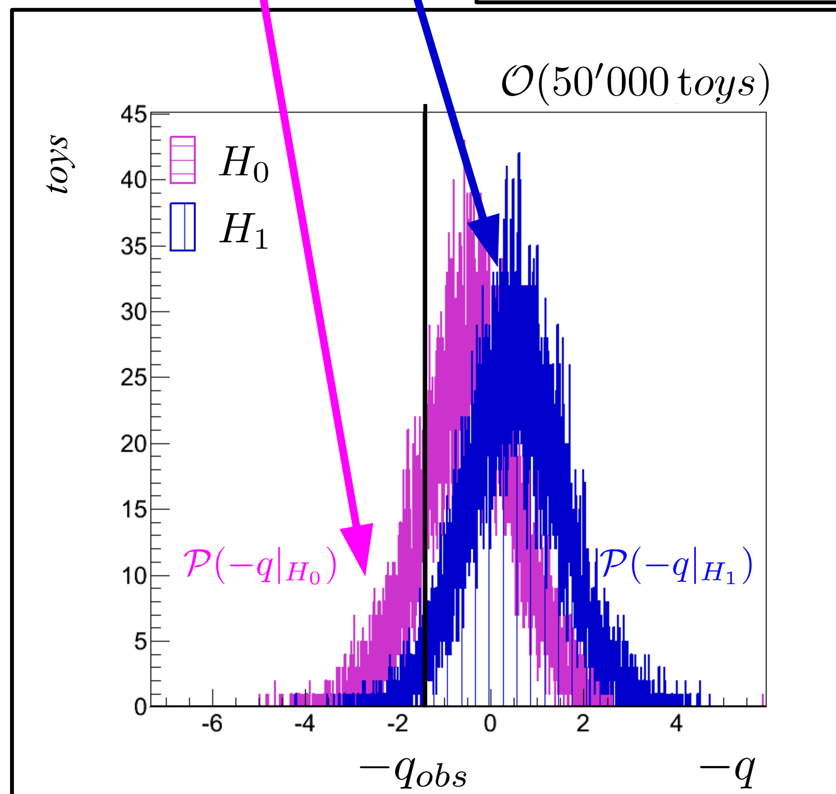


$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}$$

background signal



Confidence levels (*CL*)

- The association to one or the other hypothesis can be performed up to a given confidence level α .

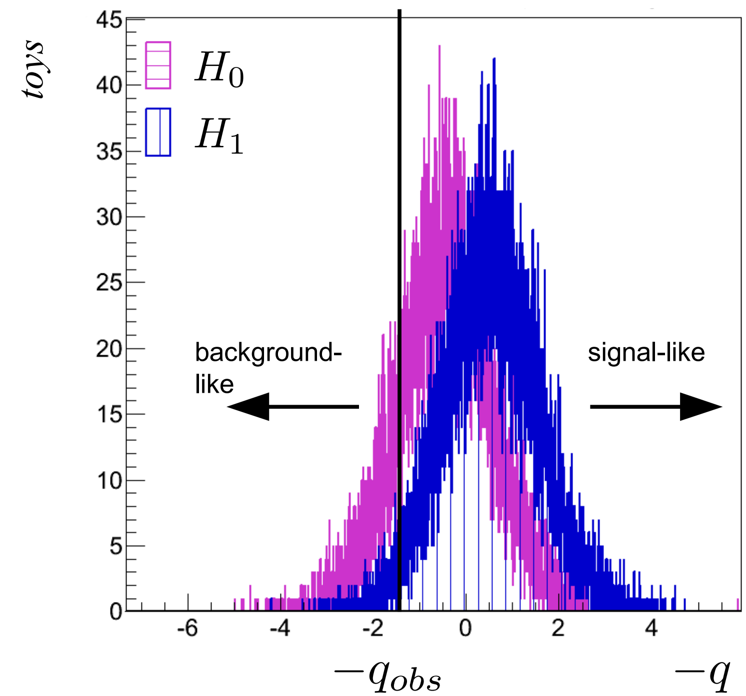
$$(1 - CL_b) = \int_{-\infty}^{q_{\text{obs}}} \mathcal{P}_b \qquad (p\text{-value})$$

$$CL_{s+b} = \int_{q_{\text{obs}}}^{+\infty} \mathcal{P}_{s+b} \qquad (CL_{s+b} \text{ confidence})$$

$$CL_b = \int_{q_{\text{obs}}}^{+\infty} \mathcal{P}_b \qquad (CL_b \text{ confidence})$$

$$CL_s = \frac{CL_{s+b}}{CL_b} \qquad (CL_s \text{ confidence})$$

Attention: in all plots $-q$ is shown.



The p-value

Challenging the H_0 hypothesis

- The association to one or the other hypothesis can be performed up to a given confidence level α.

Probability to obtain values of q, which are at least as signal-like as q_{obs}. If p-value is small H_0 can be excluded. (*)

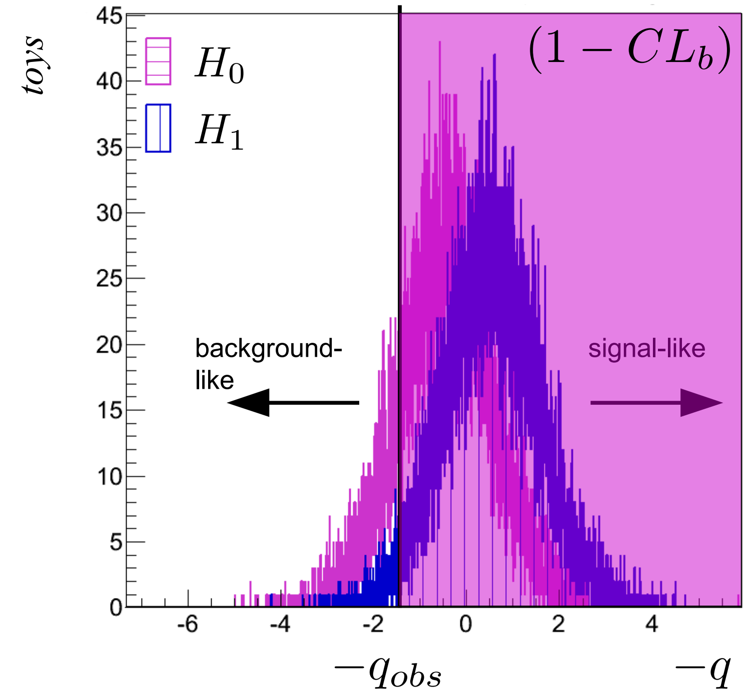
$$(1-CL_b) = \int_{-\infty}^{q_{\mathrm{obs}}} \mathcal{P}_b \qquad (p\text{-value})$$

$$CL_{s+b} = \int_{q_{\mathrm{obs}}}^{+\infty} \mathcal{P}_{s+b} \qquad (CL_{s+b}\ \text{confidence})$$

$$CL_b = \int_{q_{\mathrm{obs}}}^{+\infty} \mathcal{P}_b \qquad (CL_b\ \text{confidence})$$

$$CL_s = \frac{CL_{s+b}}{CL_b} \qquad (CL_s\ \text{confidence})$$

Attention: in all plots $-q$ is shown.

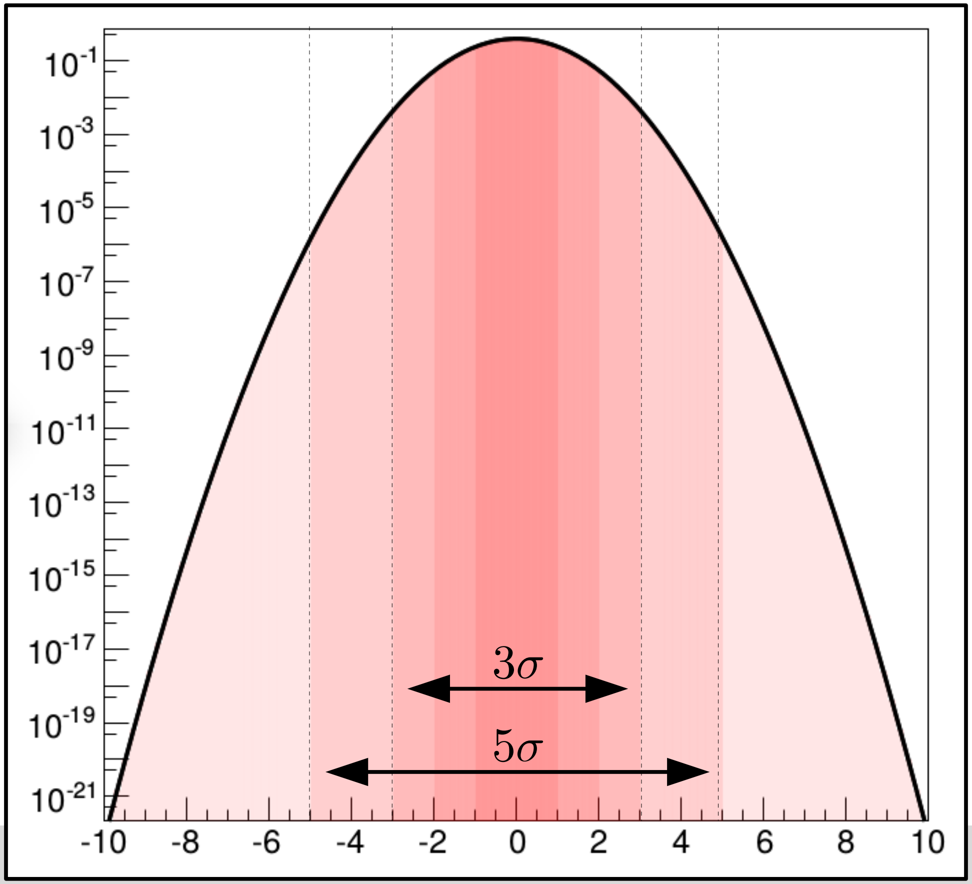


(*) Imagine data show a peak. What is the prob. that this is due to an upward fluctuation of the expectation from H_0.

Significance

Challenging the H_0 hypothesis

- If the measurement is normal distributed q is distributed according to a χ^2 distribution (cf. slide 21f).
- The resulting χ^2 probability is then equivalent to a Gaussian confidence interval in terms of standard deviations σ.



p-values:

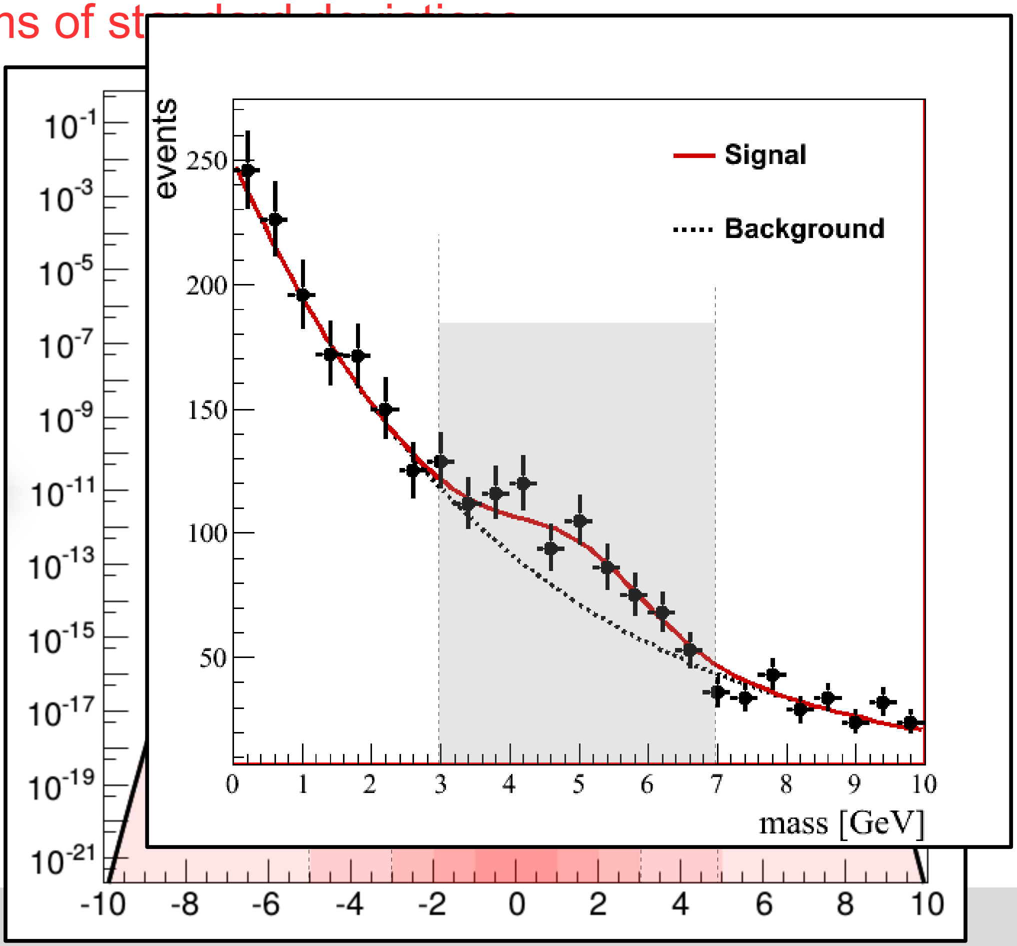
$$\mathcal{P}(q \geq 3\sigma | H_0) = 1 \cdot 10^{-3}$$

$$\mathcal{P}(q \geq 5\sigma | H_0) = 2 \cdot 10^{-5}$$

Significance (in practice)

Challenging the H_0 hypothesis

- If the measurement is normal distributed q is distributed according to a χ^2 distribution (cf. slide 21f).
- The resulting χ^2 probability is then equivalent to a Gaussian confidence interval in terms of st



Deviation from expectation for H_0 .

$$\mathcal{S} = \frac{n_{\mathrm{obs}} - n_b}{\sqrt{n_b}}$$

Poisson uncert. for H_0 .

Excluding parameters

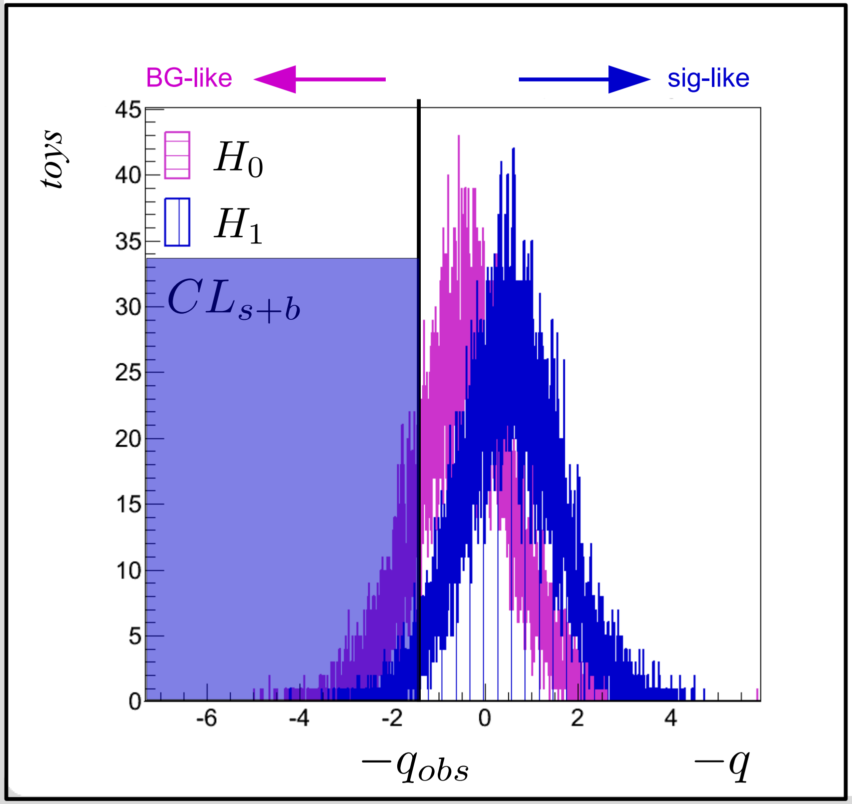
Challenging the H_1 hypothesis

- Sorry, don't see any signal. Up to what size should I definitely have seen it?
- Usually q depends on POI: $q = -2\ln\left(\frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0}\right)$

varies

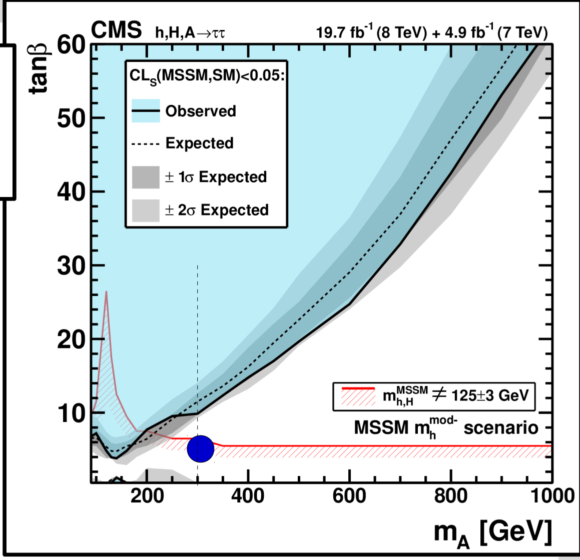
fixed

fixed

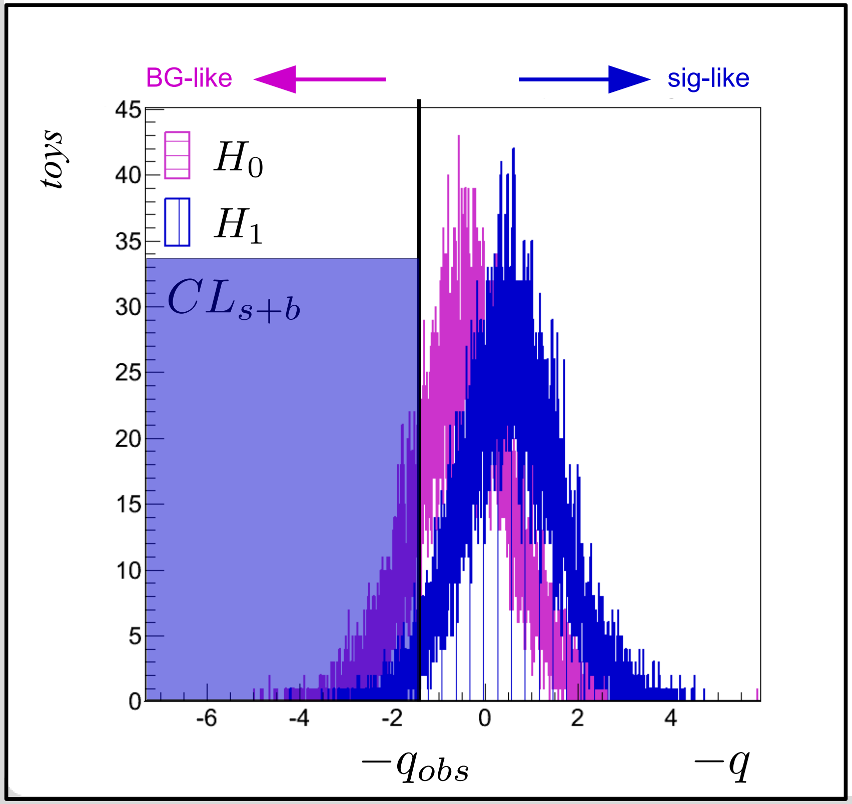


Excluding parameters

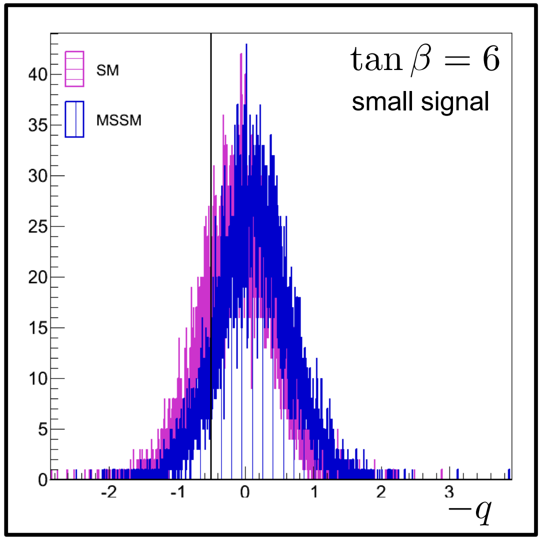
Challenging the H_1 hypothesis



- Usually q depends on POI: $q = -2\ln\left(\frac{\mathcal{L}(\mathrm{obs})|H_1}{\mathcal{L}(\mathrm{obs})|H_0}\right)$

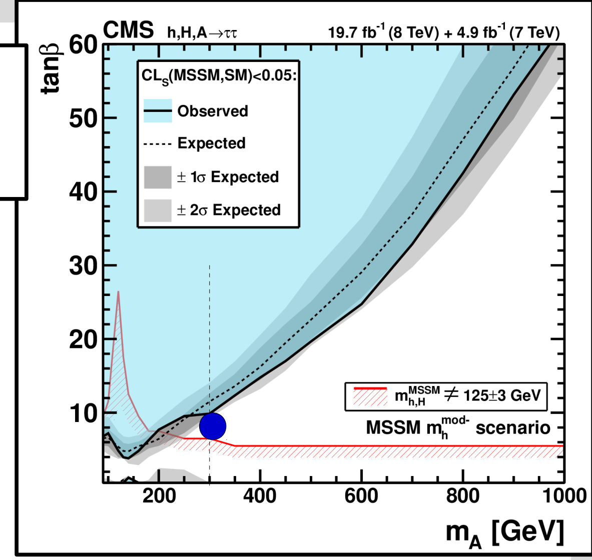


$m_A = 300\,\mathrm{GeV}$

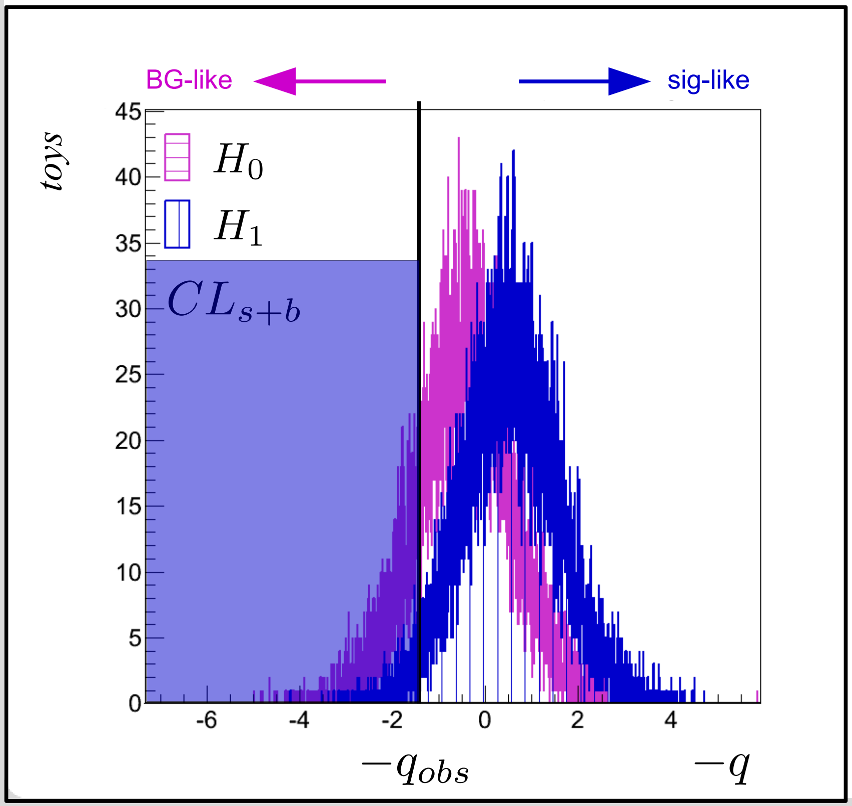


Excluding parameters

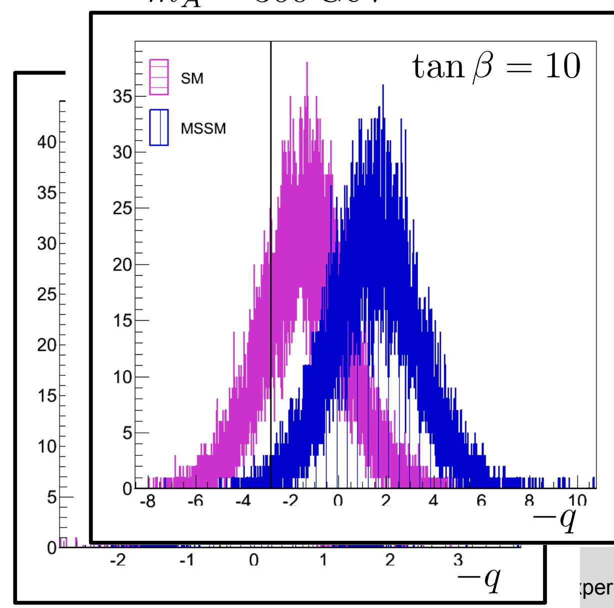
Challenging the H_1 hypothesis



- Usually q depends on POI: $q = -2\ln\left(\frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0}\right)$

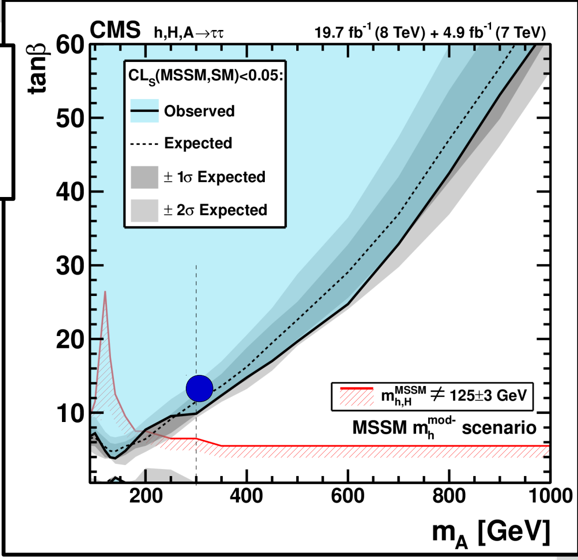


$m_A = 300\,\text{GeV}$

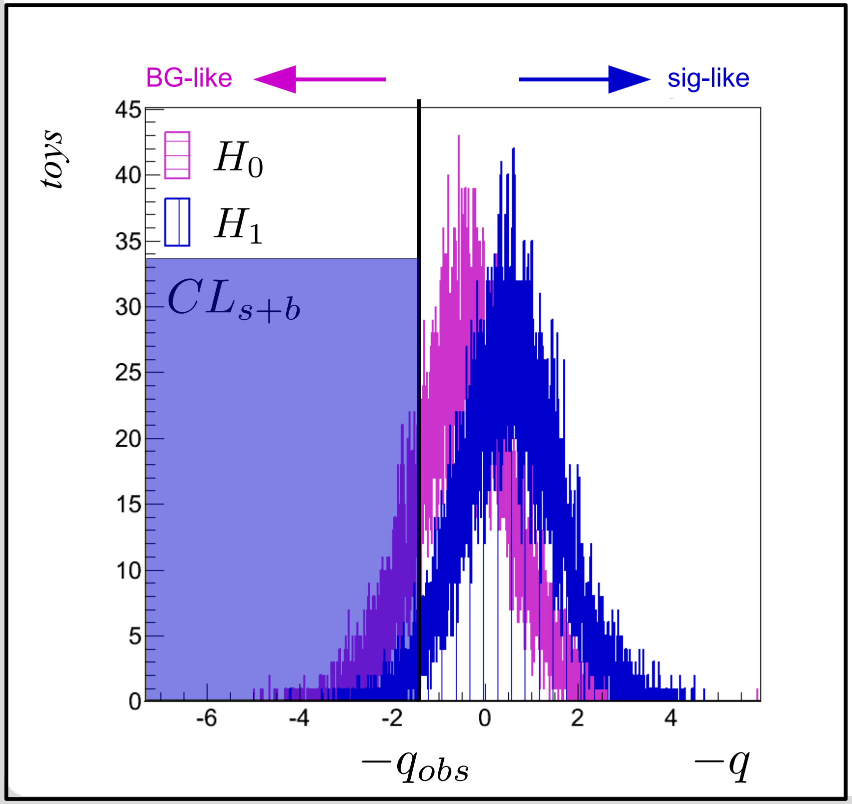


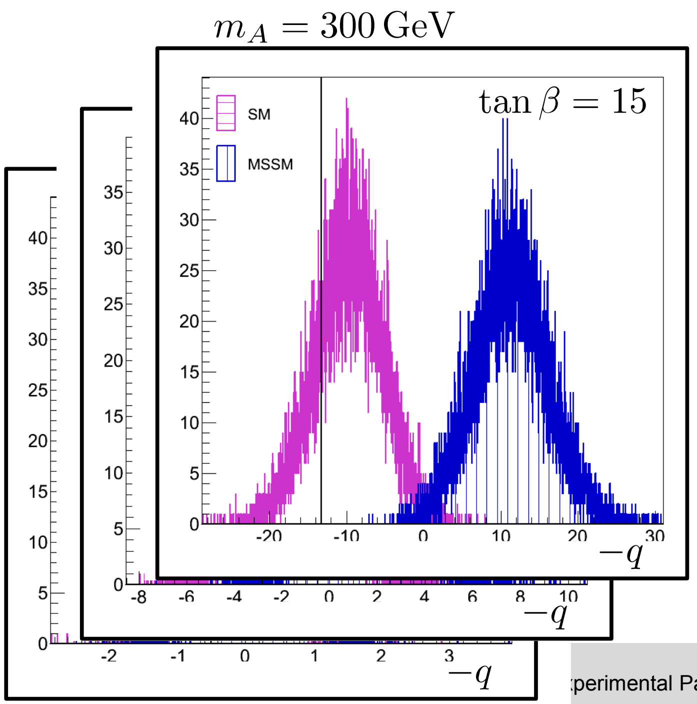
Excluding parameters

Challenging the H_1 hypothesis



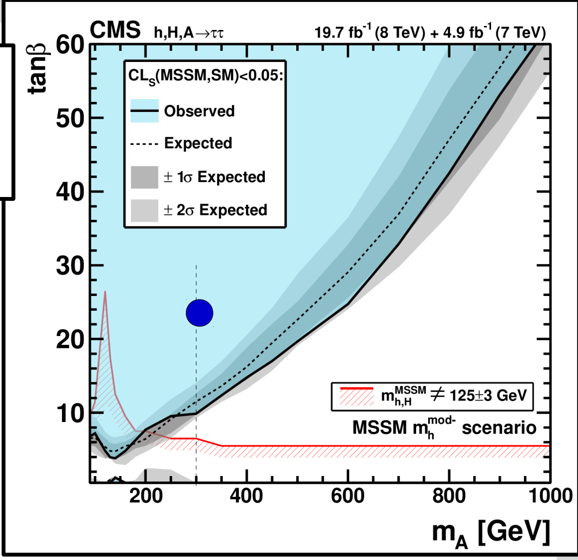
- Usually q depends on POI: $q = -2\ln\left(\frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0}\right)$



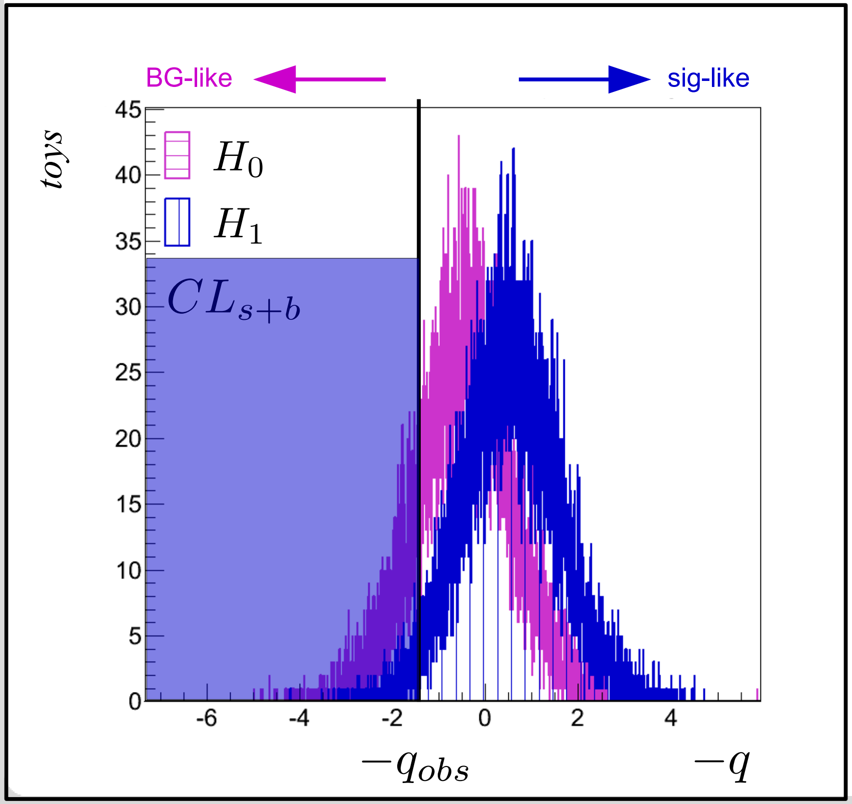


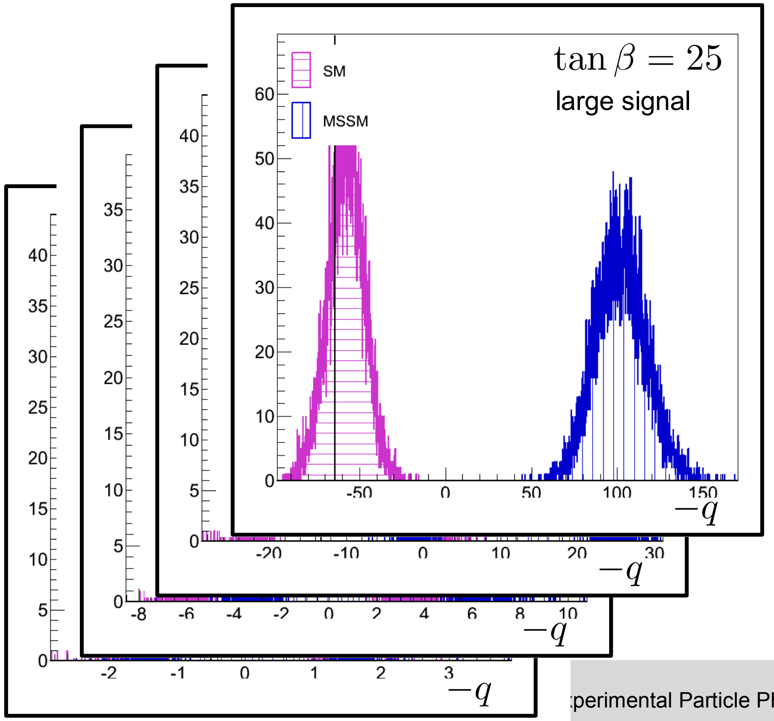
Excluding parameters

Challenging the H_1 hypothesis



- Usually q depends on POI: $q = -2\ln\left(\frac{\mathcal{L}(\mathrm{obs})|H_1}{\mathcal{L}(\mathrm{obs})|H_0}\right)$

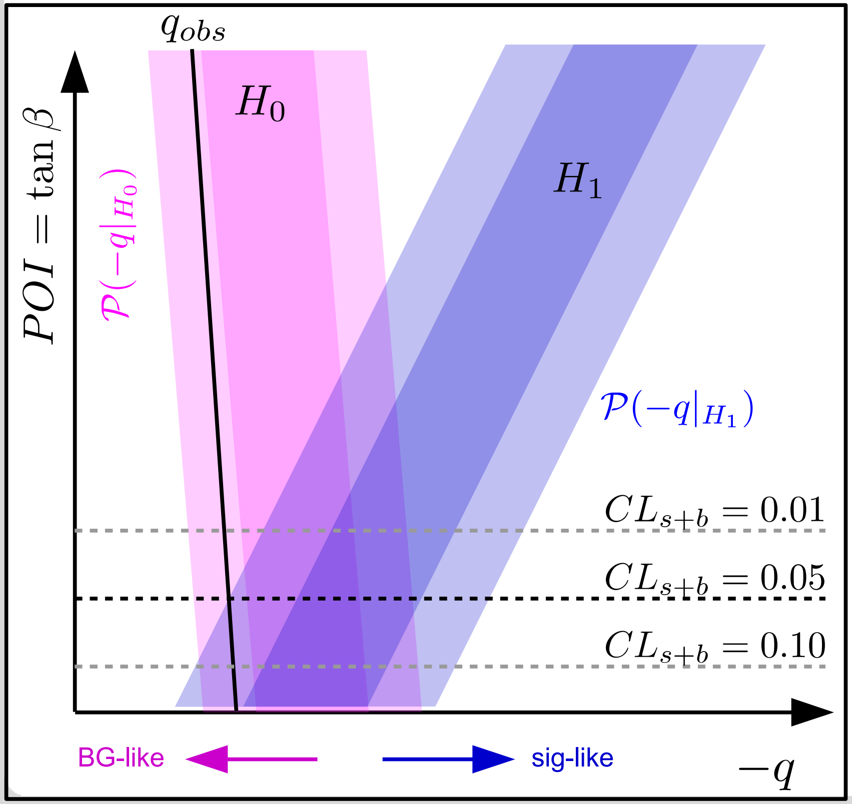




Observed exclusion

Challenging the H_1 hypothesis

- Traditionally we determine 95% CL exclusions on the POI ($\alpha = 0.05$).
- To be conservative choose probability α that q is more BG-like than q_{obs} low (→ safer exclusion).



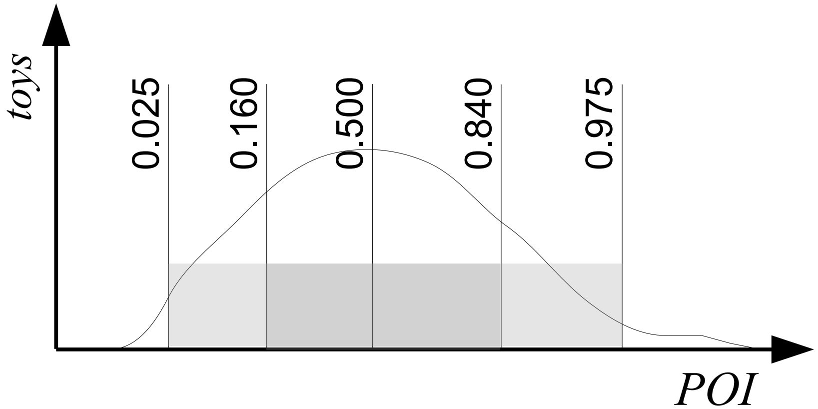
- $\mathcal{P}(-q|_{H_0})$ and $\mathcal{P}(-q|_{H_1})$ move apart from each other with increasing POI.
- The more separated $\mathcal{P}(-q|_{H_0})$ and $\mathcal{P}(-q|_{H_1})$ are the clearer H_0 and H_1 can be distinguished.
- For 95% CL identify value of POI for which:
  $$CL_{s+b} = \int_{q_{obs}}^{+\infty} \mathcal{P}_{s+b} = 0.05$$
  for this value $q|H_1$ would have been more signal-like than q_{obs} with 95% probability.
- There is still a 5% chance that we exclude by mistake.

Expected exclusion

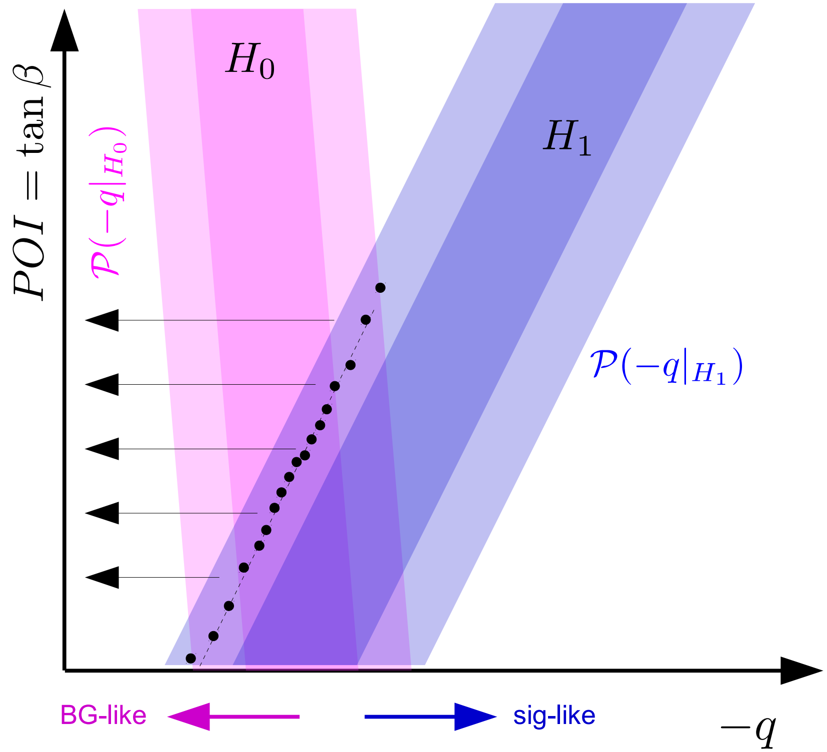
Challenging the H_1 hypothesis

- To obtain expected limit mimic calculation of observed; base it on toy datasets.
- Use fact that $\mathcal{P}(-q|_{H_0})$ and $\mathcal{P}(-q|_{H_1})$ do not depend on toys (i.e. schematic plot on the left does not change).

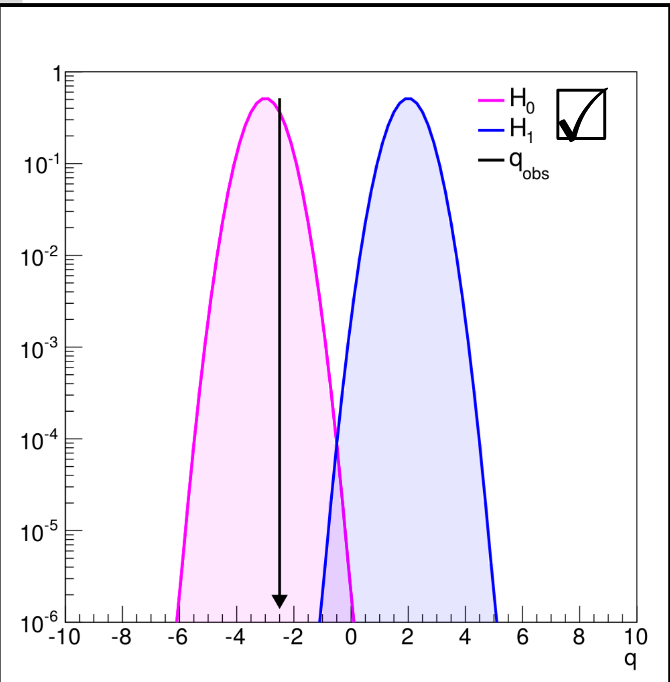
Throw toys under H_0 hypothesis; determine distribution of 95% CL limits on POI:



Obtain quantiles for expected exclusion from this distribution (expected limit = median).

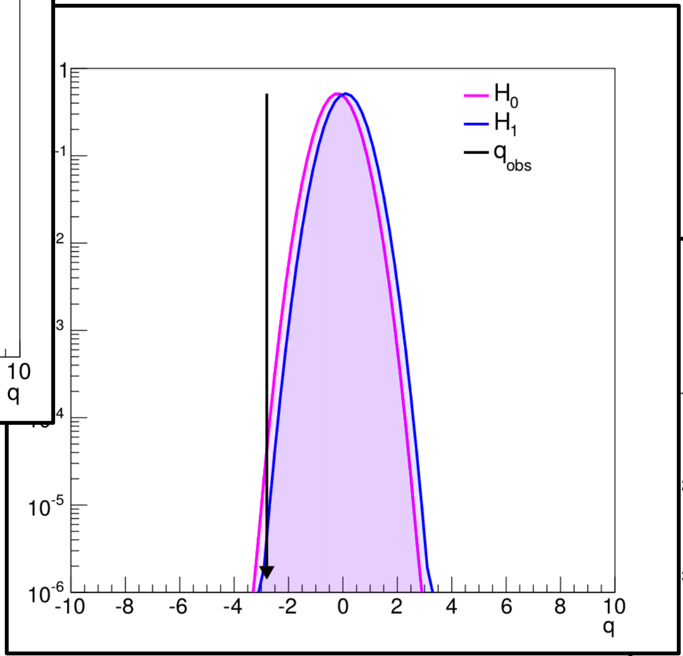


Interpretation issues (increasing pathology)

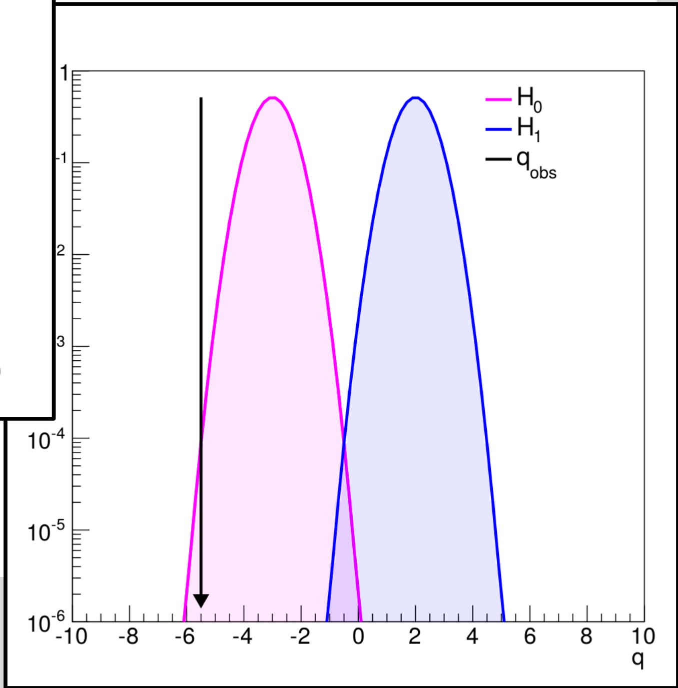


- q_{obs} compatible with BG hypothesis.
- q_{obs} incompatible with signal hypothesis.

- Signal and BG hypothesis cannot be distinguished.
- Should this outcome lead to an exclusion of the signal hypothesis?

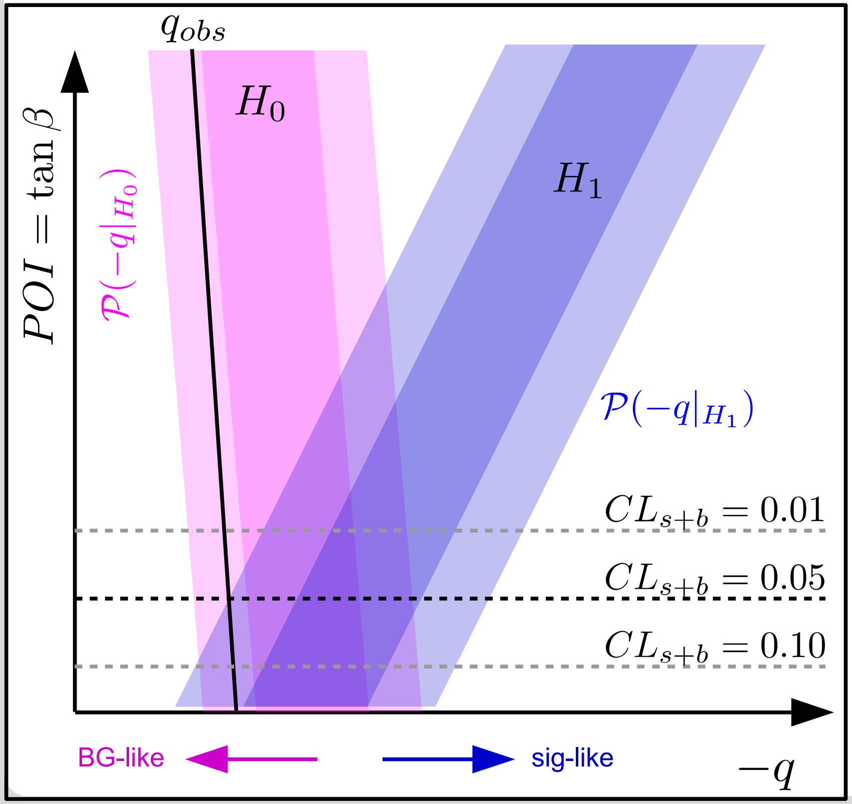


- q_{obs} incompatible both with signal and BG hypothesis.
- Should this outcome lead to an exclusion of the signal hypothesis?



Modified frequentist exclusion method (CLs)

- In particle physics we set more conservative limits, following the CLs method:



- $CL_{s+b} = \int_{q_{obs}}^{+\infty} \mathcal{P}_{s+b}$
  $CL_b = \int_{q_{obs}}^{\infty} \mathcal{P}_b$
- Identify value of POI for which:
  $CL_s = \frac{CL_{s+b}}{CL_b} = 0.05$
- If H_0 and H_1 become indistinguishable:
  $CL_{s+b} < CL_s \to 1$

Judgment call

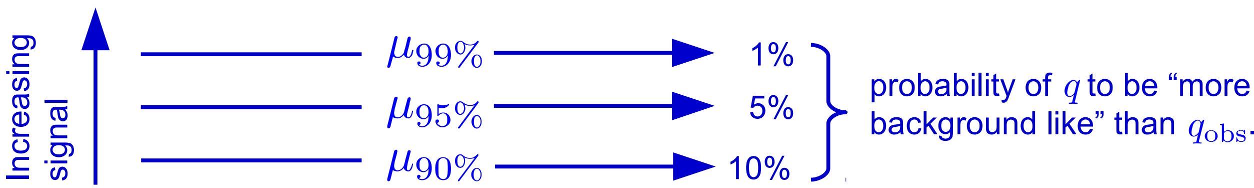
Karlsruhe Institute of Technology

- Assume our POI is the signal strength μ of a new signal: does the 90% CL upper limit on μ correspond to a higher or a lower value than the 95% CL limit?

Judgment call

- Assume our POI is the signal strength μ of a new signal: does the 90% CL upper limit on μ correspond to a higher or a lower value than the 95% CL limit?

  It's lower:



Concluding remarks

- Reviewed all statistical tools necessary to search for the Higgs boson signal ($\rightarrow$ as a small signal above a known background):
- Limits: usual way to 'challenge' signal hypothesis (H_1).
- *p*-values: usual way to 'challenge' background hypothesis (H_0).
- Under the assumption that the test statistic q is χ^2 distributed *p*-values can be translated into Gaussian confidence intervals σ.
- In particle physics we call an observation with $\geq 3\sigma$ an evidence.
- We call an observation with $\geq 5\sigma$ a discovery.

During the next lectures we will see 1:1 life examples of all methods that have been presented here.

Backup