A boost to Higgs Physics: new regimes at high energy

Silvia Biondi
University & INFN of Bologna
silvia.biondi@cern.ch / silvia.biondi@bo.infn.it

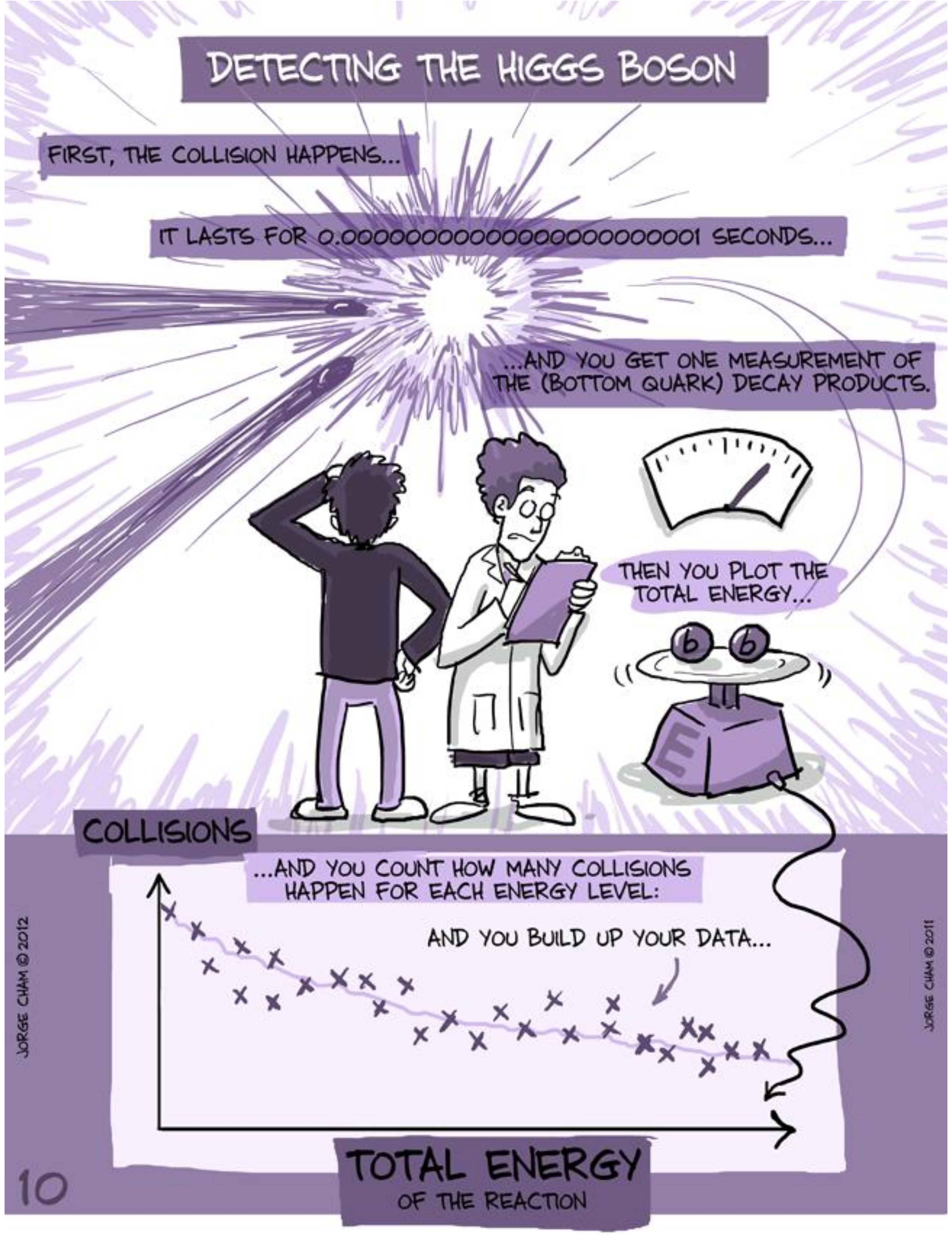
Corso di Dottorato - 2019/2020

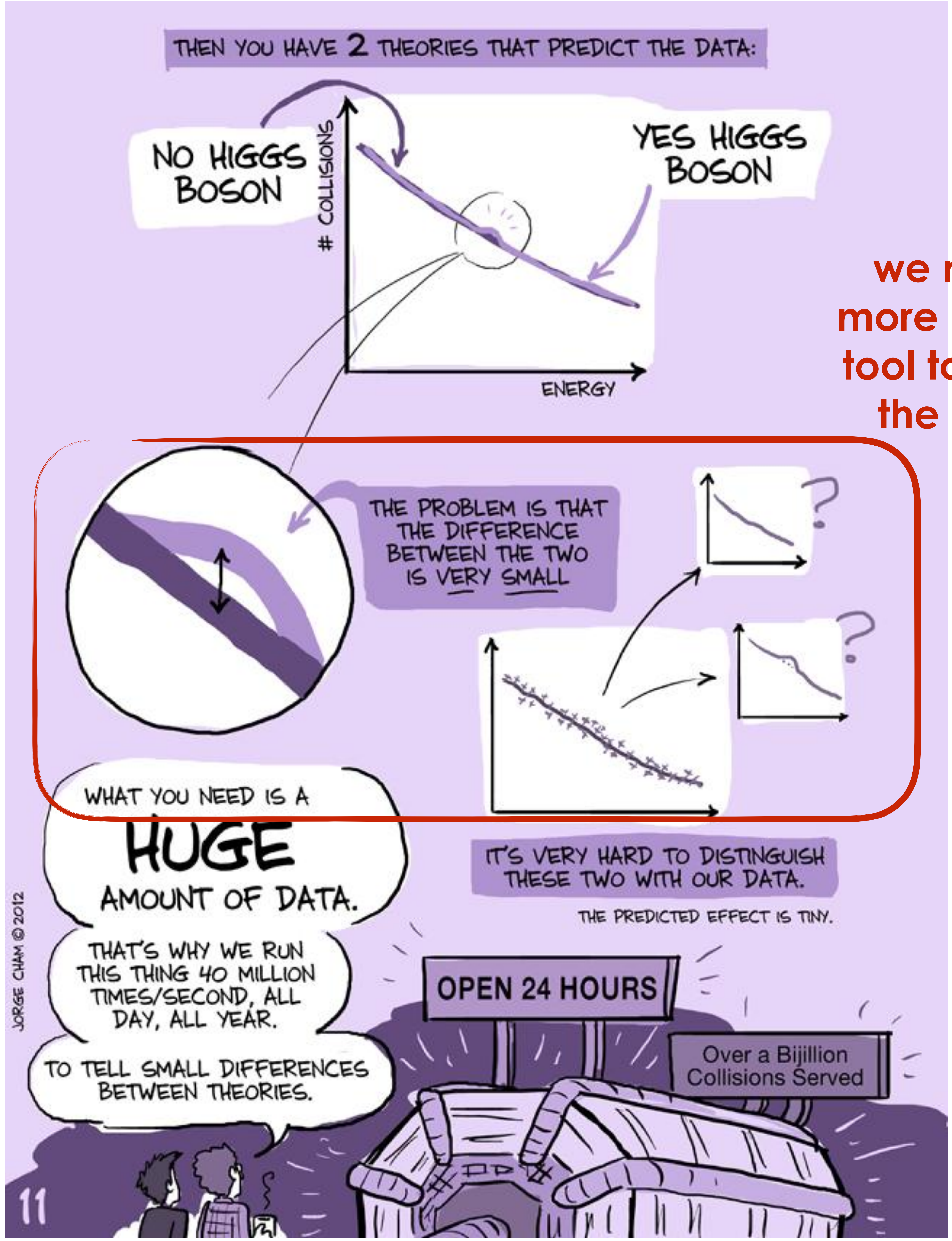
Course outline

- Theory reminder
- Higgs boson production and decay modes
- Higgs boson discovery by ATLAS and CMS
- Higgs boson mass measurement by ATLAS and CMS
- Overview of ATLAS and CMS analyses about Higgs
- Signal/background discrimination techniques
  - boosted regimes
    - tagging, large-radius jets substructure, re-clustering
  - multivariate analysis and deep neural network
- **Signal extraction techniques**
  - **likelihood and test statistic**
  - **CLs method**
- ttH analysis: an example

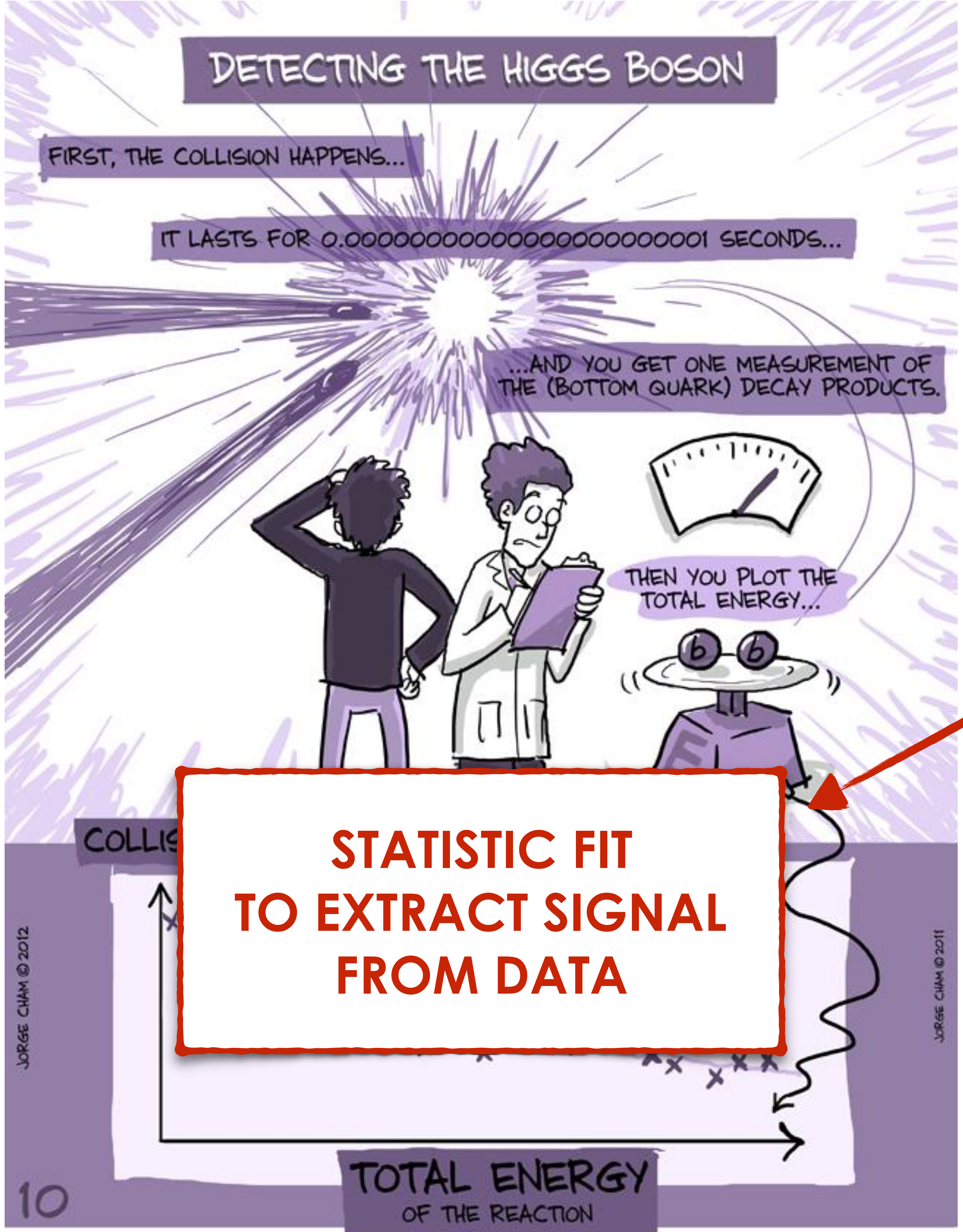
Signal extraction techniques

http://www.phdcomics.com/

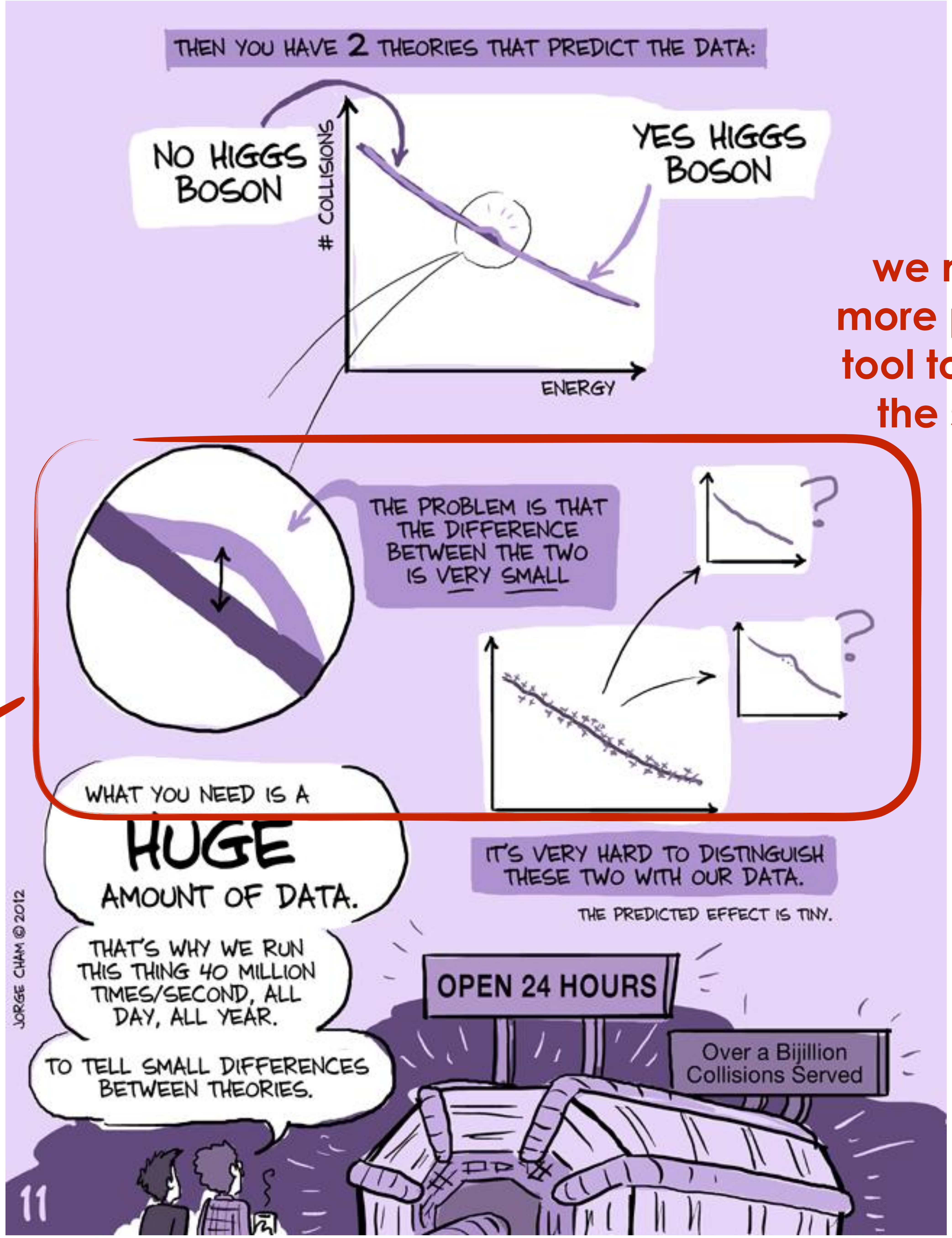




we need a more powerful tool to extract the signal!



STATISTIC FIT TO EXTRACT SIGNAL FROM DATA



we need a more powerful tool to extract the signal!

Statistical tests

Some notions

- goal of a statistical test is to make a statement about how well the observed data stand in agreement with given predicted probabilities (a **hypothesis**);

- **null hypothesis, H_0**: hypothesis under consideration, specifying a probability density f(x) of a random variable x
  - **simple** if it determines f(x) uniquely
  - **composite** if the pdf is defined but not the values of at least one free parameter θ, f(x,θ).

- Statement about H_0 validity often involves a comparison with some **alternative hypotheses** H_1, H_2, …

Statistical tests

Some notions

- goal of a statistical test is to make a statement about how well the observed data stand in agreement with given predicted probabilities (a **hypothesis**);

- **null hypothesis, H_0**: hypothesis under consideration, specifying a probability density f(x) of a random variable x
  - **simple** if it determines f(x) uniquely
  - **composite** if the pdf is defined but not the values of at least one free parameter θ, f(x,θ).

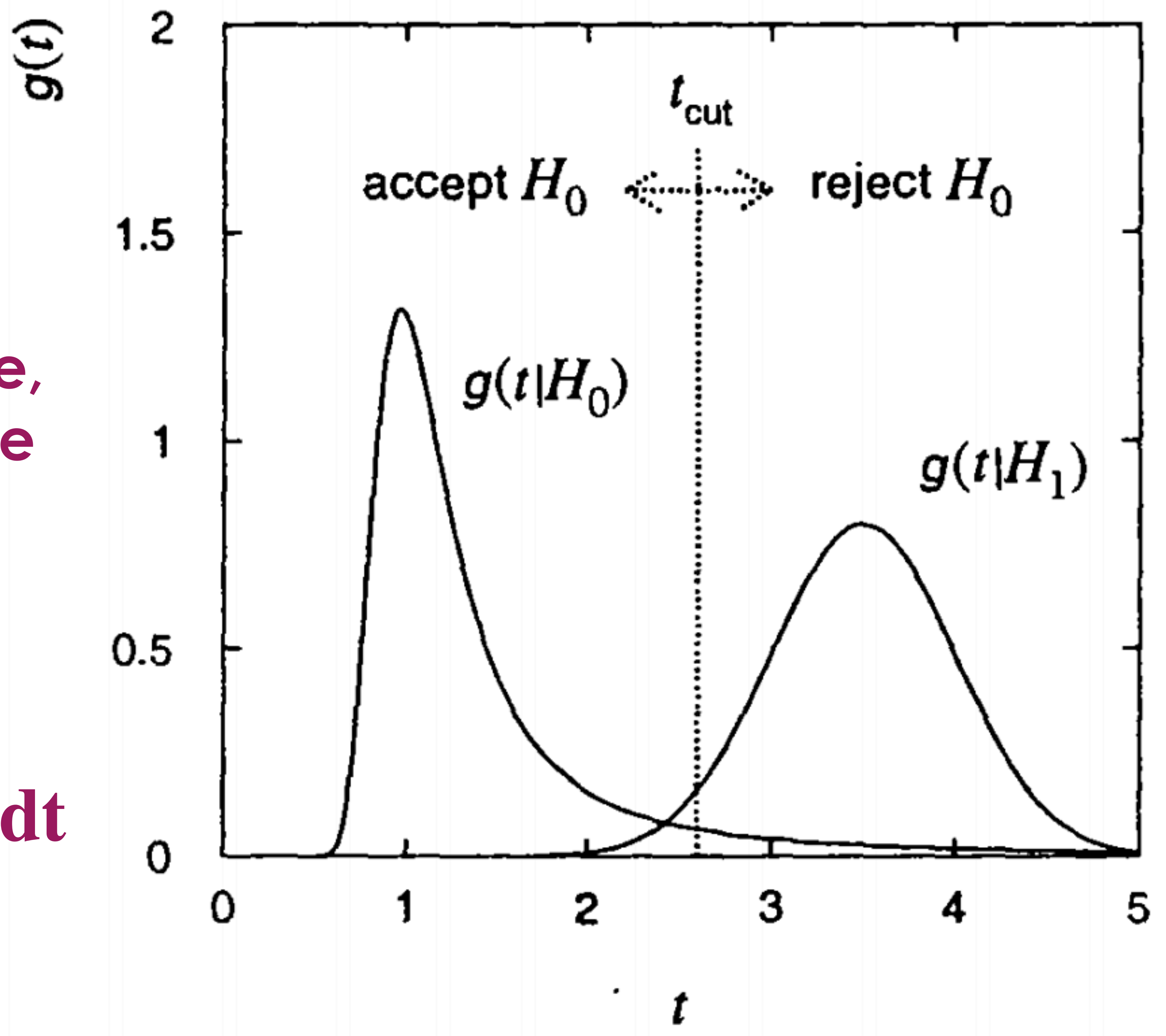
- Statement about H_0 validity often involves a comparison with some **alternative hypotheses** H_1, H_2, …

An example

- data: *n* measured values $\mathbf{x} = (x_1, x_2, \ldots, x_n)$;
- set of hypotheses: H_0, H_1; → $f(x|H_0)$, $f(x|H_1)$
- construction of a function of the measured variables called a **test statistic t(x)**, to investigate the agreement between the observed data and a given hypothesis; → $g(t|H_0)$ if H_0 is true, $g(t|H_1)$ if H_1 is true
- statement about the compatibility between the data and the various hypotheses in terms of a decision to accept or reject a given null hypothesis H_0, by defining a **critical region** for t:
  - chosen such that the probability for t to be observed there, under assumption of the hypothesis H_0, is some value a, called the **significance level** of the test; →

$$\alpha = \int_{t_{cut}}^{\infty} g(t|H_0)\,dt$$

- accepting (or not rejecting) the hypothesis H_0, if $t<t_{cut}$.



Statistical tests

How do we chose the critical region?

- depending on the **efficiency and purity of the selected events** desired in the analysis;
- **multidimensional test statistics** chosen and simple hypothesis **H_0** tested, to select events of a given type; **$t = (t_1, \ldots t_m)$**
- simple alternative hypothesis **H_1**;
- **also background events**, so that the signal purity in the selected sample will in general be less than 100%.

- **Neyman-Pearson lemma** states that the acceptance region giving the highest power (and hence the highest signal purity) for a given significance level a (or selection efficiency ε = 1 - a) is the region of t-space such that:

$$\frac{g(t|H_0)}{g(t|H_1)} > c$$

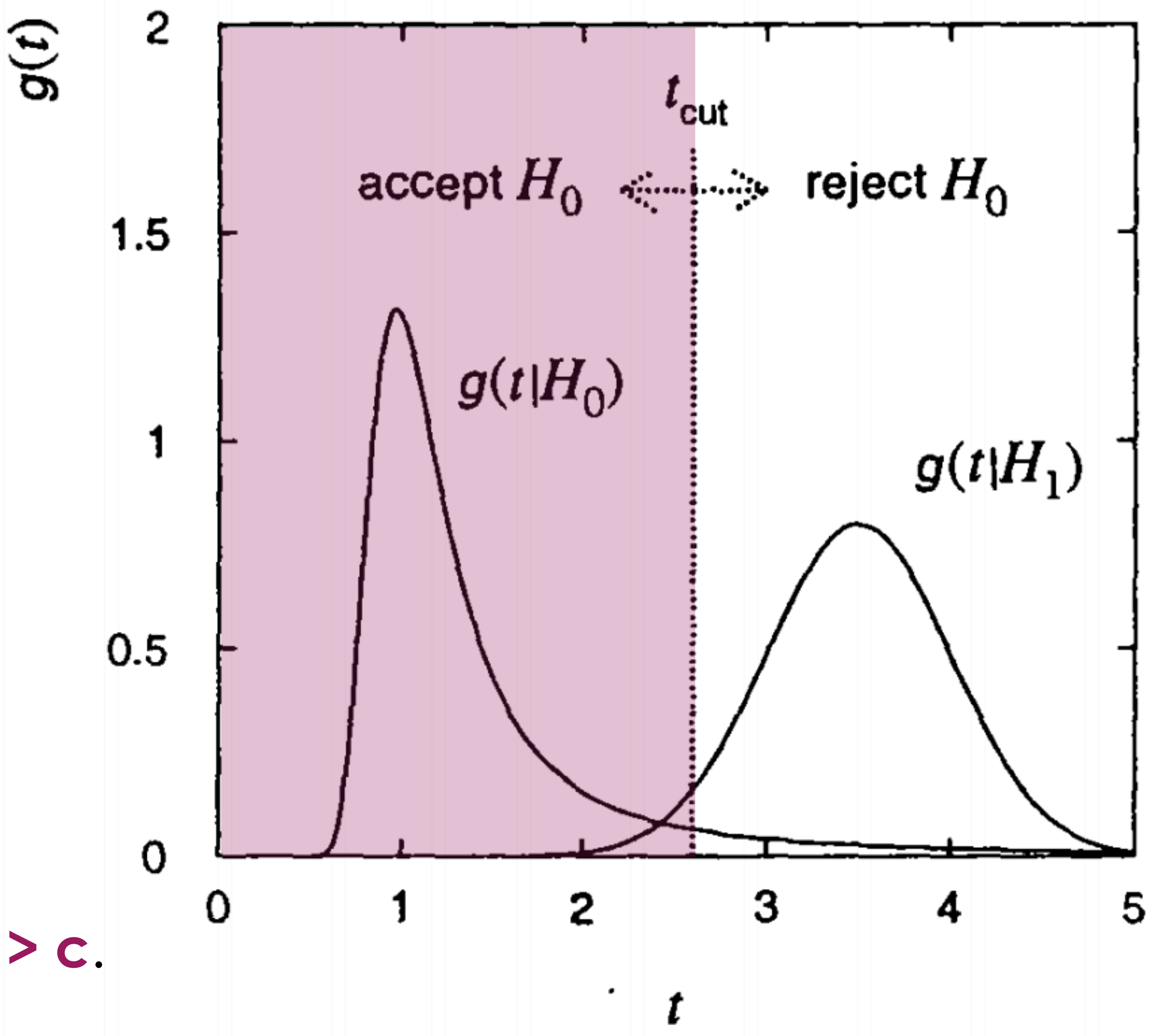
c = constant which is determined by the desired efficiency

equivalent to a test using a one-dimensional statistic given by the ratio

$$r = \frac{g(t|H_0)}{g(t|H_1)}$$

Lilkelihood ratio for simple hypotheses H_0 and H_1

The corresponding acceptance region is given by **r > c**.



Significance of an observed signal

Goodness-of-fit

- to judge whether a discrepancy between data and expectation is sufficiently significant to merit a **claim for a new discovery**;
- given by stating the so-called **P-value**, i.e. the probability P, under assumption of the hypothesis in question H_0, of obtaining a result as compatible or less with H_0 than the one actually observed:
  - also called the **observed significance level** of the test;
  - the fraction of times one would obtain data as compatible with H_0 or less so if the experiment (i.e. 20 coin tosses) were repeated many times under similar circumstances.

Significance of an observed signal

Goodness-of-fit

- to judge whether a discrepancy between data and expectation is sufficiently significant to merit a **claim for a new discovery**;
- given by stating the so-called **P-value**, i.e. the probability P, under assumption of the hypothesis in question H_0, of obtaining a result as compatible or less with H_0 than the one actually observed:
  - also called the **observed significance level** of the test;
  - the fraction of times one would obtain data as compatible with H_0 or less so if the experiment (i.e. 20 coin tosses) were repeated many times under similar circumstances.

An example

- special type of signal event, the number n_s of which can be treated as a Poisson variable with mean ν_s;
- a certain number of background events n_b, can be treated as a Poisson variable with mean ν_b;
- total number of events found is $\mathbf{n} = \mathbf{n_s} + \mathbf{n_b}$, which is a Poisson variable with mean $\nu = \nu_s + \nu_b$;

Significance of an observed signal

Goodness-of-fit

- to judge whether a discrepancy between data and expectation is sufficiently significant to merit a **claim for a new discovery**;
- given by stating the so-called **P-value**, i.e. the probability P, under assumption of the hypothesis in question H_0, of obtaining a result as compatible or less with H_0 than the one actually observed:
  - also called the **observed significance level** of the test;
  - the fraction of times one would obtain data as compatible with H_0 or less so if the experiment (i.e. 20 coin tosses) were repeated many times under similar circumstances.

An example

- special type of signal event, the number n_s of which can be treated as a Poisson variable with mean ν_s;
- a certain number of background events n_b, can be treated as a Poisson variable with mean ν_n;
- total number of events found is $n = n_s + n_b$, which is a Poisson variable with mean $\nu = \nu_s + \nu_b$;

- probability to observe n events:
- from experiment, n_{obs} events found;
- to quantify our degree of confidence in the discovery of a new effect ($\nu_s \neq 0$) we can compute **how likely it is to find n_{obs} events or more from background alone**:

$$f(n;\nu_s,\nu_b) = \frac{(\nu_s+\nu_b)^n}{n!}e^{-(\nu_s+\nu_b)}$$

$$P(n \geq n_{obs}) = \sum_{n=n_{obs}}^{\infty} f(n;\nu_s=0,\nu_b)$$

$$= 1 - \sum_{n=0}^{n_{obs}-1} f(n;\nu_s=0,\nu_b)$$

$$= 1 - \sum_{n=0}^{n_{obs}-1} \frac{\nu_b^n}{n!}e^{-\nu_b}$$

Significance of an observed signal

Goodness-of-fit

- to judge whether a discrepancy between data and expectation is sufficiently significant to merit a **claim for a new discovery**;
- given by stating the so-called **P-value**, i.e. the probability P, under assumption of the hypothesis in question H_0, of obtaining a result as compatible or less with H_0 than the one actually observed:
  - also called the **observed significance level** of the test;
  - the fraction of times one would obtain data as compatible with H_0 or less so if the experiment (i.e. 20 coin tosses) were repeated many times under similar circumstances.

An example

- special type of signal event, the number n_s of which can be treated as a Poisson variable with mean ν_s;
- a certain number of background events n_b, can be treated as a Poisson variable with mean ν_b;
- total number of events found is $\mathbf{n = n_s + n_b}$, which is a Poisson variable with mean $\nu = \nu_s + \nu_b$;

- probability to observe n events:
- from experiment, n_{obs} events found;
- to quantify our degree of confidence in the discovery of a new effect ($\nu_s \neq 0$) we can compute **how likely it is to find n_{obs} events or more from background alone**:

$$\mathbf{P(n \geq n_{obs}) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{\nu_b^n}{n!} e^{-\nu_b}}$$

- if we expect $\nu_b = 0.5$ background events and we observe $\mathbf{n_{obs} = 5}$, then the P-value is **1.7 x 10^{-4}**;

- this is not the probability of the hypothesis $\nu_s = 0$.

- **It is rather the probability, under the assumption $\nu_s = 0$, of obtaining as many events as observed or more.**

Significance of an observed signal

Goodness-of-fit

- to judge whether a discrepancy between data and expectation is sufficiently significant to merit a **claim for a new discovery**;
- given by stating the so-called **P-value**, i.e. the probability P, under assumption of the hypothesis in question H_0, of obtaining a result as compatible or less with H_0 than the one actually observed:
  - also called the **observed significance level** of the test;
  - the fraction of times one would obtain data as compatible with H_0 or less so if the experiment (i.e. 20 coin tosses) were repeated many times under similar circumstances.

An example

- special type of signal event, the number n_s of which can be treated as a Poisson variable with mean ν_s;
- a certain number of background events n_b, can be treated as a Poisson variable with mean ν_b;
- total number of events found is $\mathbf{n = n_s + n_b}$, which is a Poisson variable with mean $\nu = \nu_s + \nu_b$;
- probability to observe n events:
- from experiment, n_{obs} events found;
- to quantify our degree of confidence in the discovery of a new effect ($\nu_s \neq 0$) we can compute **how likely it is to find n_{obs} events or more from background alone**:

$$\mathbf{P(n \geq n_{obs}) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{\nu_b^n}{n!} e^{-\nu_b}}$$

Danger!

- We assumed ν_b to be known without error;
- if we had $\nu_b = 0.8$ instead, the P-value would increase by almost an order of magnitude to **1.4 x 10^{-3}**!
- **important to quantify the systematic uncertainty in the background when evaluating the significance of a new effect!**

Likelihood-based test

Standard procedure to discover a new signal process

- A widely used procedure to establish discovery (or exclusion) in particle physics is based on a significance test using a **profile likelihood ratio** as a test statistic;

- necessary that the **model predictions** for data distributions **represent accurately the underlying theory being tested**:
  - any errors due to approximations (e.g. in detector modelling) should be negligible in the full parameter space;

- By including **additional parameters to the model** (accounting for systematics effects) it is possible to approach this ideal situation more closely, but resulting in a loss in sensitivity.

H_0 only background
H_1 both background and signal

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **n = (n_1, ..., n_N)** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$\mathbf{E[n_i] = \mu s_i + b_i}$$

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **n = (n_1, ..., n_N)** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process

μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **n = (n_1, ..., n_N)** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process

μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{bin\ i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{bin\ i} f_b(x; \theta_b)\, dx$$

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **n = (n_1, ..., n_N)** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{\text{bin } i} f_b(x; \theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **$n = (n_1, ..., n_N)$** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{bin\ i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{bin\ i} f_b(x; \theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

nuisance parameters
parameters describing in general unknown systematic effect, whose contributions must be fitted from the data

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **$n = (n_1, ..., n_N)$** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
$\mu = 0$ corresponding to the bkg-only hypothesis
$\mu = 1$ being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{\text{bin } i} f_b(x; \theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

nuisance parameters
parameters describing in general unknown systematic effect, whose contributions must be fitted from the data

total signal and background events
the integrals represent the probabilities for an event to be found in bin i only

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **$n = (n_1, ..., n_N)$** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{bin\ i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{bin\ i} f_b(x; \theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

nuisance parameters
parameters describing in general unknown systematic effect, whose contributions must be fitted from the data

total signal and background events
the integrals represent the probabilities for an event to be found in bin i only

* notation used in the following to denote all the NPs and the bkg contribution

$$\theta = (\theta_s, \theta_b, b_{tot})$$

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **$n = (n_1, ..., n_N)$** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{bin\ i} f_s(x; \theta_s)\, dx \qquad b_i = b_{tot} \int_{bin\ i} f_b(x; \theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

nuisance parameters
parameters describing in general unknown systematic effect, whose contributions must be fitted from the data

total signal and background events
the integrals represent the probabilities for an event to be found in bin i only

* notation used in the following to denote all the NPs and the bkg contribution $\theta = (\theta_s, \theta_b, b_{tot})$

- subsidiary measurements performed **to help constraining the NPs**;
- evaluating some chosen kinematic variables in a control region to construct a new histogram **$m = (m_1, ..., m_M)$ of M bins**.

$$E[m_i] = u_i(\theta)$$

Likelihood-based test: building the Likelihood

How to build a Likelihood

- **For each event** selected in the signal sample, a variable **x** of a certain kinematic quantity can be measured;
- values **n = (n_1, ..., n_N)** can be used to construct a **histogram of N bins**.

expectation value of n_i

$$E[n_i] = \mu s_i + b_i$$

strength of the signal process
μ = 0 corresponding to the bkg-only hypothesis
μ = 1 being the nominal (predicted by the SM) signal hypothesis

mean number of entries in the i^{th} bin from signal and bkg

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x;\theta_s)\, dx \qquad b_i = b_{tot} \int_{\text{bin } i} f_b(x;\theta_b)\, dx$$

pdfs of variable x for signal (f_s) and bkg (f_b)

nuisance parameters
parameters describing in general unknown systematic effect, whose contributions must be fitted from the data

total signal and background events
the integrals represent the probabilities for an event to be found in bin i only

* notation used in the following to denote all the NPs and the bkg contribution $\theta = (\theta_s, \theta_b, b_{tot})$

- subsidiary measurements performed **to help constraining the NPs**;
- evaluating some chosen kinematic variables in a control region to construct a new histogram **m = (m_1, ..., m_M) of M bins**.

provides information on the background normalisation parameter b_{tot}

$$E[m_i] = u_i(\theta)$$

calculable quantities depending on the parameters θ

Likelihood-based test: building the Likelihood

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

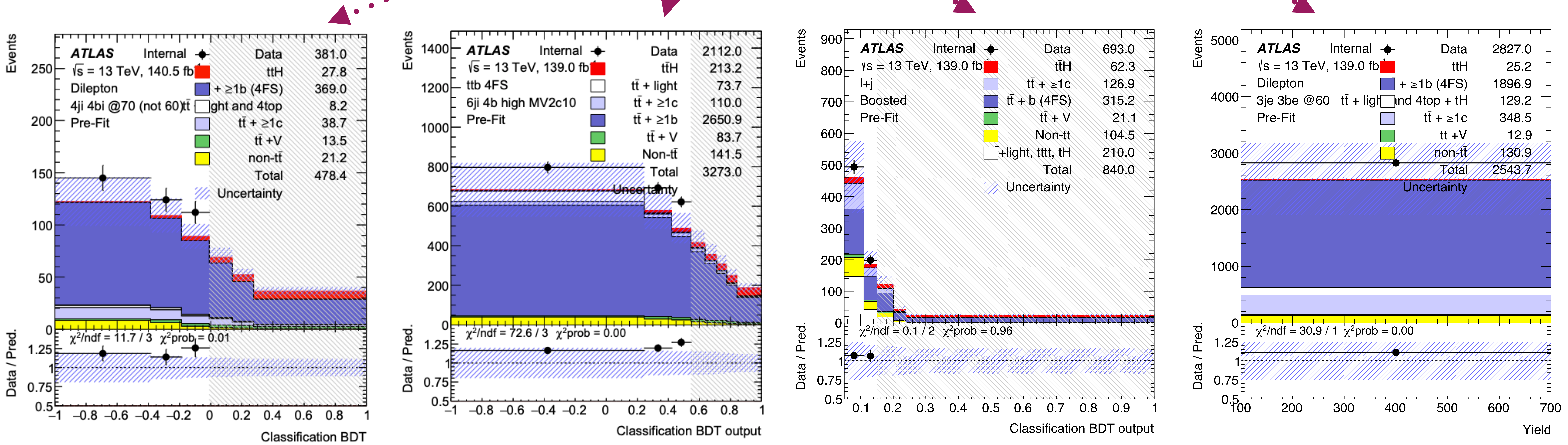
Likelihood-based test: building the Likelihood

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

example



Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

profile likelihood function

Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

profile likelihood function

value of θ

maximising $\mathscr{L}(\mu, \hat{\hat{\theta}})$ for the specified μ and representing the conditional maximum-likelihood (ML) estimator of θ (function of μ)

Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

profile likelihood function

maximised (unconditional) likelihood function

value of θ

maximising $\mathscr{L}(\mu, \hat{\hat{\theta}})$ for the specified μ and representing the conditional maximum-likelihood (ML) estimator of θ (function of μ)

Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

profile likelihood function

maximised (unconditional) likelihood function

value of θ

maximising $\mathscr{L}(\mu, \hat{\hat{\theta}})$ for the specified μ and representing the conditional maximum-likelihood (ML) estimator of θ (function of μ)

values of μ and θ

maximising $\mathscr{L}(\hat{\mu}, \hat{\theta})$ and representing the unconditional maximum-likelihood (ML) estimators

Likelihood-based test: building the Likelihood and its ratio

How to build a Likelihood

- likelihood function is defined as the **product of Poisson probabilities for all bins**:

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

Profile Likelihood ratio

- To test a hypothesised value of μ:

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

profile likelihood function

maximised (unconditional) likelihood function

value of θ

maximising $\mathscr{L}(\mu, \hat{\hat{\theta}})$ for the specified μ and representing the conditional maximum-likelihood (ML) estimator of θ (function of μ)

values of μ and θ

maximising $\mathscr{L}(\hat{\mu}, \hat{\theta})$ and representing the unconditional maximum-likelihood (ML) estimators

- $\lambda(\mu)$ **assumes values between 0 and 1 (at $\mu = \hat{\mu}$)**;
- λ **close to 1** implies **good agreement between data and the hypothesised value of μ**;
- The presence of the nuisance parameters broadens the profile likelihood as a function of μ
  - **loss of information about μ due to the systematic uncertainties**.

Likelihood-based test: test statistic

To establish an upper limit on the strength parameter μ

- upper limit is obtained by **testing μ against the alternative hypothesis consisting of lower values of μ**;
- definition of a test statistic **q_μ**:

$$q_\mu = \begin{cases} 0, & \mu < \hat{\mu} \\ -2\ln\lambda(\mu), & \mu \geq \hat{\mu}, \end{cases}$$

→ to not represent the values with $\mu < \hat{\mu}$ with less compatibility with respect to the μ obtained from the data

To clarify the formula and the results

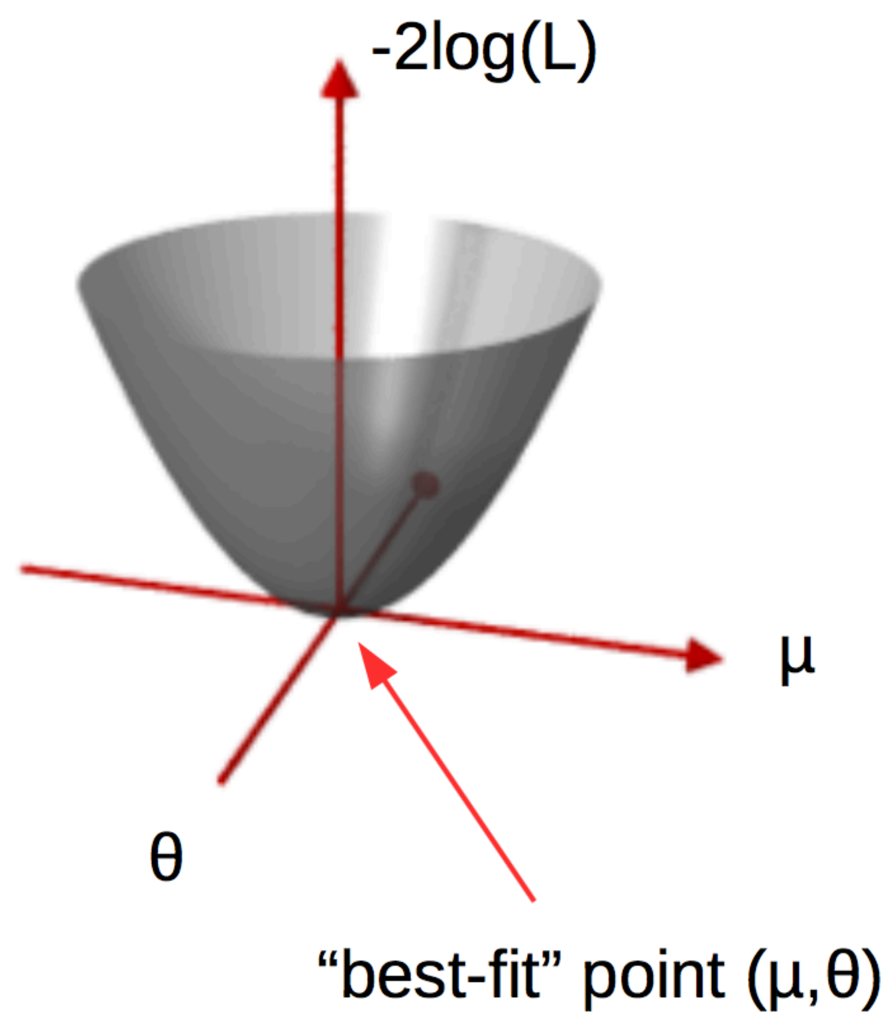
- in the conditions of the central limit theorem, given enough statistics:

$$\lambda(\mu) \approx exp(-\chi^2/2) \quad \rightarrow \quad q_\mu = \chi^2(\mu) \qquad \text{for } \mu > \hat{\mu}$$

high values of q(μ) are equivalent to high values of a χ^2

↓

incompatibility between the data and the test hypothesis



Likelihood-based test: p-value

To quantify the level of agreement between data and hypothesised μ

- definition of p-value, relative to the test statistic;
- for an observed value $\mathbf{q_{\mu,obs}}$:

$$\mathbf{p}_\mu = \int_{\mathbf{q}_{\mu,\mathbf{obs}}}^{\infty} \mathbf{f}(\mathbf{q}_\mu \,|\, \mu)\ \mathbf{dq}_\mu$$

Likelihood-based test: p-value

To quantify the level of agreement between data and hypothesised μ

- definition of p-value, relative to the test statistic;
- for an observed value $q_{\mu,obs}$:

pdf of q_μ assuming the hypothesis μ

$$\mathbf{p}_\mu = \int_{\mathbf{q}_{\mu,\mathbf{obs}}}^{\infty} \mathbf{f}(\mathbf{q}_\mu \,|\, \mu)\, \mathbf{dq}_\mu$$

Likelihood-based test: p-value

To quantify the level of agreement between data and hypothesised μ

- definition of p-value, relative to the test statistic;
- for an observed value $q_{\mu,obs}$:

$$\mathbf{p}_\mu = \int_{\mathbf{q}_{\mu,\mathrm{obs}}}^{\infty} \mathbf{f}(\mathbf{q}_\mu \,|\, \mu)\, \mathbf{d}\mathbf{q}_\mu$$

pdf of q_μ assuming the hypothesis μ

value of μ assumed in the distribution of the data

Likelihood-based test: p-value

To quantify the level of agreement between data and hypothesised μ

- definition of p-value, relative to the test statistic;
- for an observed value $\mathbf{q_{\mu,obs}}$:

$$\mathbf{p}_\mu = \int_{\mathbf{q}_{\mu,\mathbf{obs}}}^{\infty} \mathbf{f}(\mathbf{q}_\mu \,|\, \mu)\, \mathbf{dq}_\mu$$

pdf of q_μ assuming the hypothesis μ

value of μ assumed in the distribution of the data

hypothesis being tested

Likelihood-based test: p-value and significance

To quantify the level of agreement between data and hypothesised μ

- definition of p-value, relative to the test statistic;
- for an observed value $q_{\mu,obs}$:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid \mu)\, dq_\mu$$

pdf of q_μ assuming the hypothesis μ

value of μ assumed in the distribution of the data

hypothesis being tested

Convenient to define also the significance Z

- If x is a Gaussian distributed variable with mean m_x;
- $\hat{x}$ ($\hat{x} > m_x$) is defined as the value of x which has an upper-tail probability equal to the p-value.
- **Z** is defined as the **number of standard deviations of $\hat{x}$ with respect to m_x**:

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

Φ = quantile of the standard Gaussian

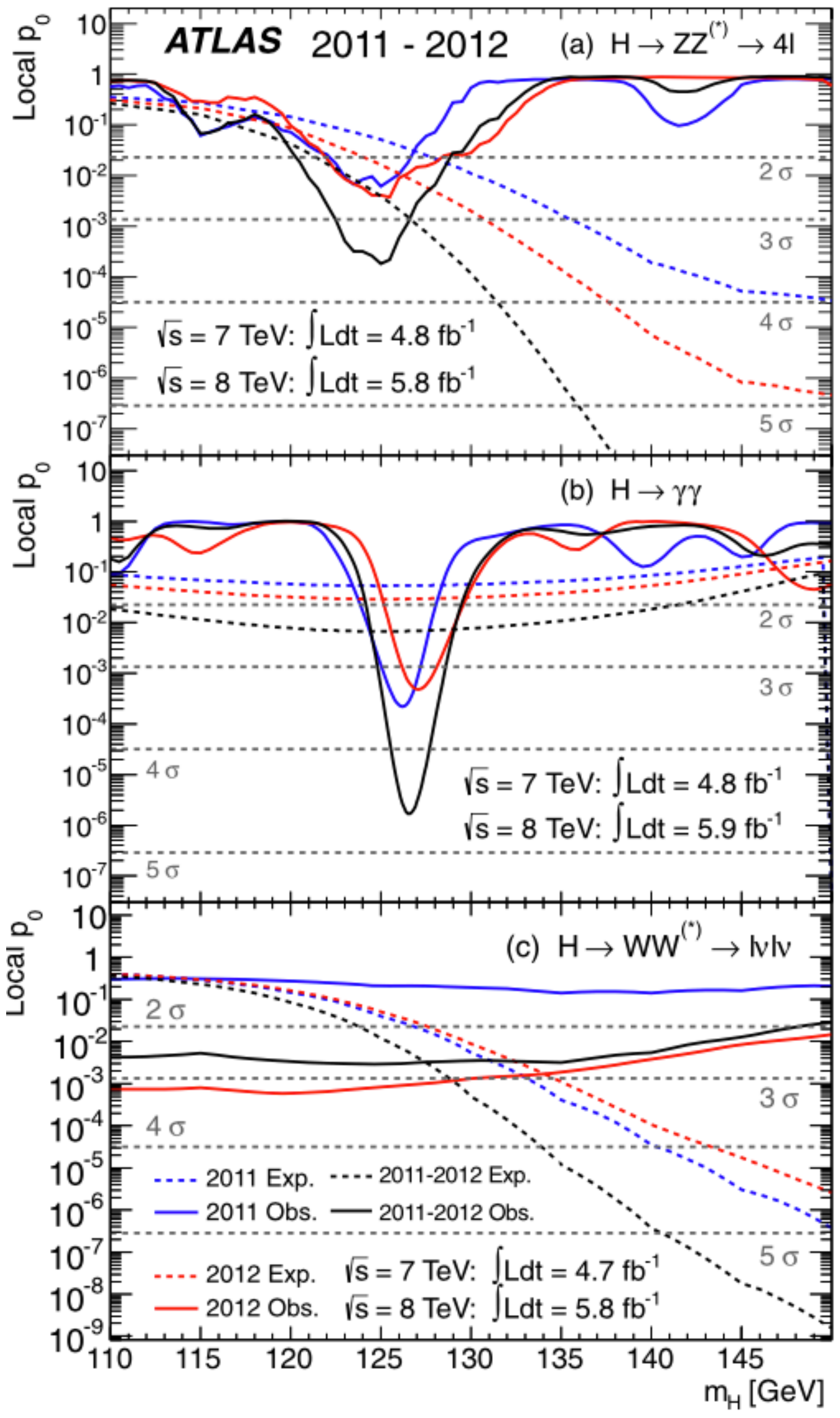
- **For a signal process such as the Higgs boson**, the appropriate level to constitute a physics discovery is a significance of at least **Z = 5**, corresponding to a **p-value = 2.87×10^{-7}**;

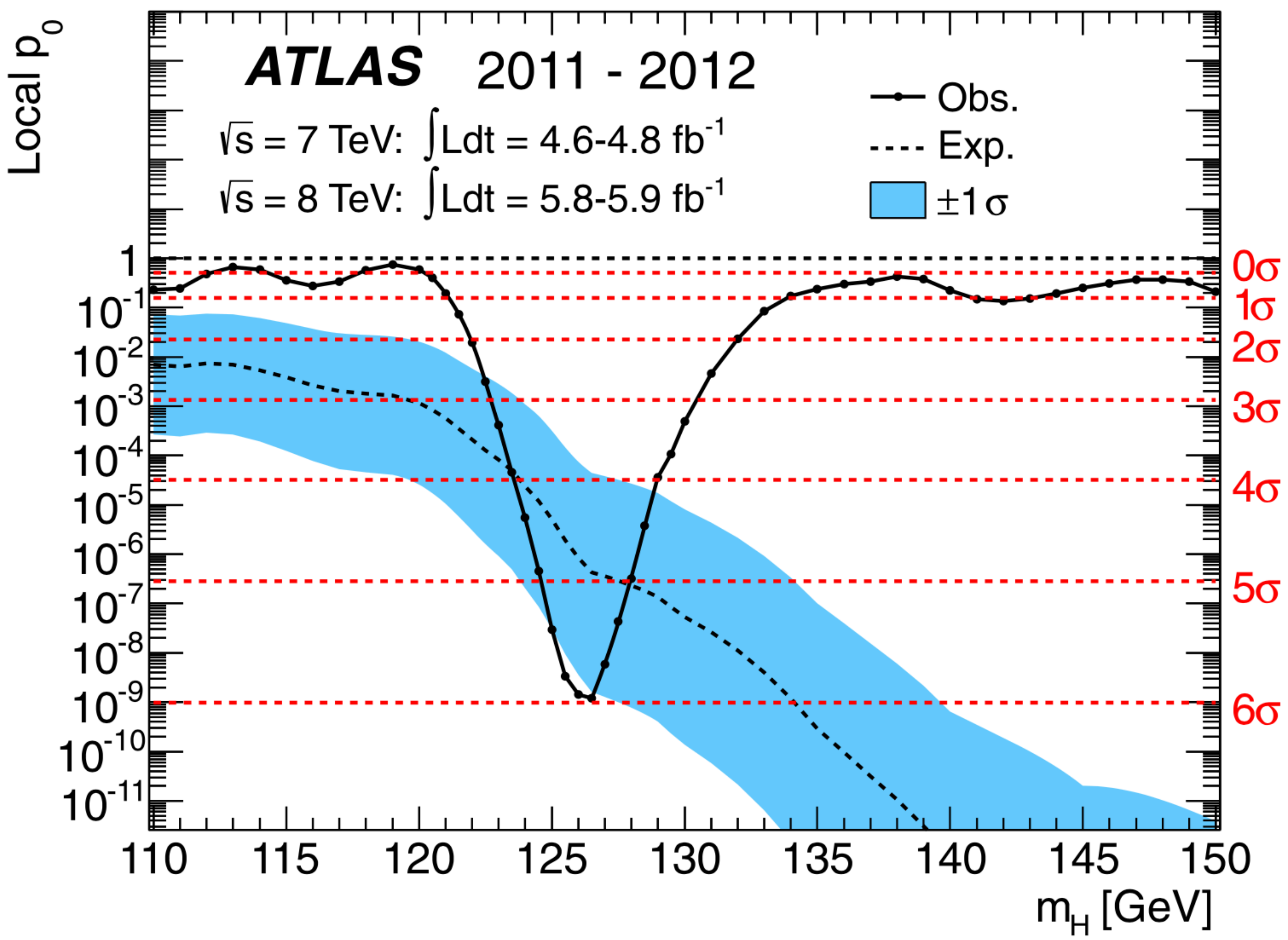
- **for excluding a signal hypothesis**, a threshold p-value of **0.05** (i.e., 95% confidence level, CL) is often used, which corresponds to **Z = 1.64**.

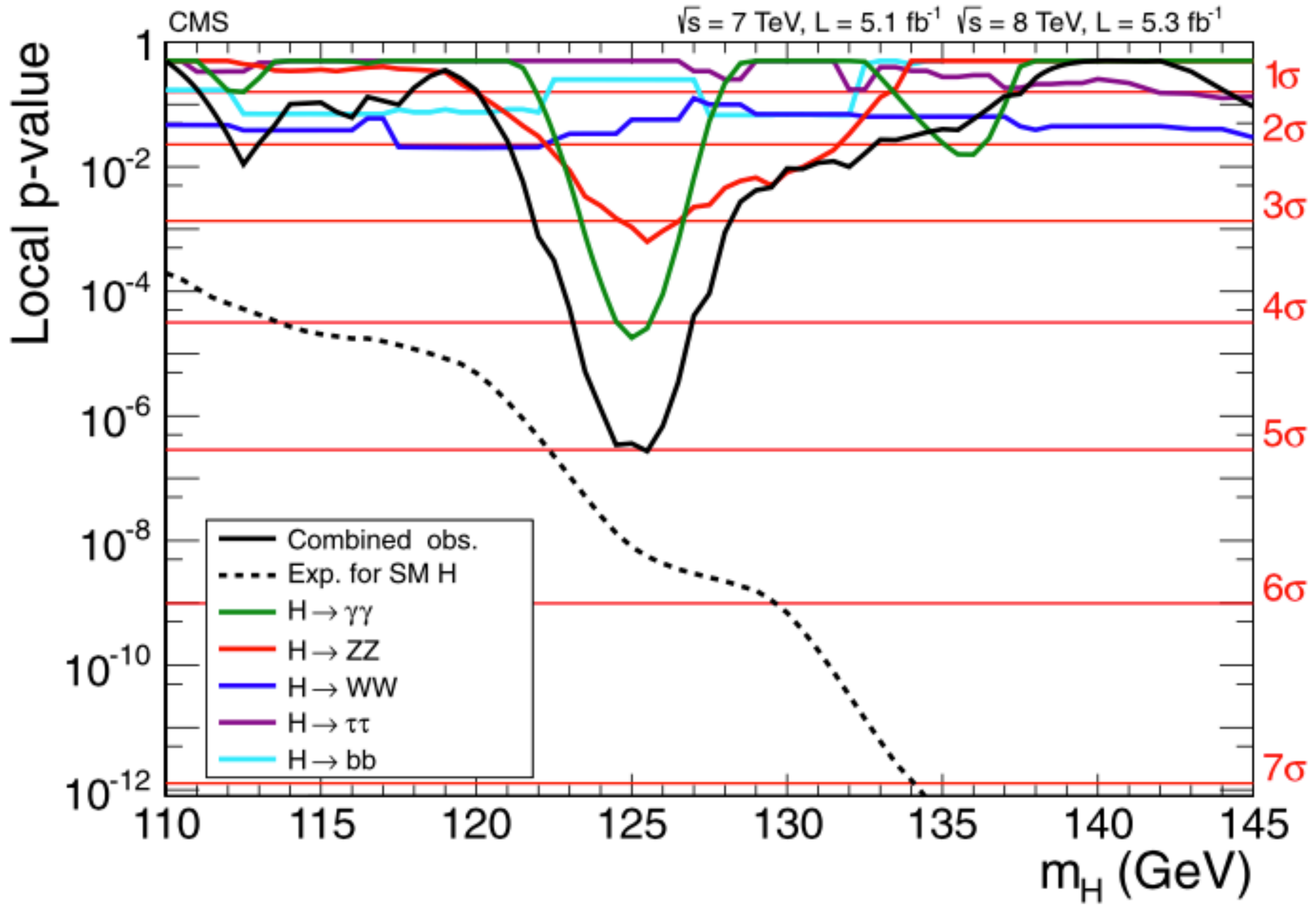
* quantile = inverse of the cumulative distribution

Likelihood-based test: p-value and significance

examples







Sensitivity of an experiment

Asimov data sets

- characterised not only in the significance obtained from a single data set, but rather in the **expected median significance to reject different values of μ**;
- estimator is evaluated by using the so called **“Asimov” data set**, replacing the ensemble of real data with a distribution obtained by a **suitable weighted addition of the MC samples**, according to the chosen test statistic.

For exclusion limit setting

sensitivity characterised by the median significance, assuming data generated using the μ = 0 hypothesis, with which one rejects a nonzero value of μ (usually μ = 1 is of greatest interest)

For discovery

sensitivity characterised by the median, under the assumption of the nominal signal model (μ = 1), with which one would reject the background-only (μ = 0) hypothesis

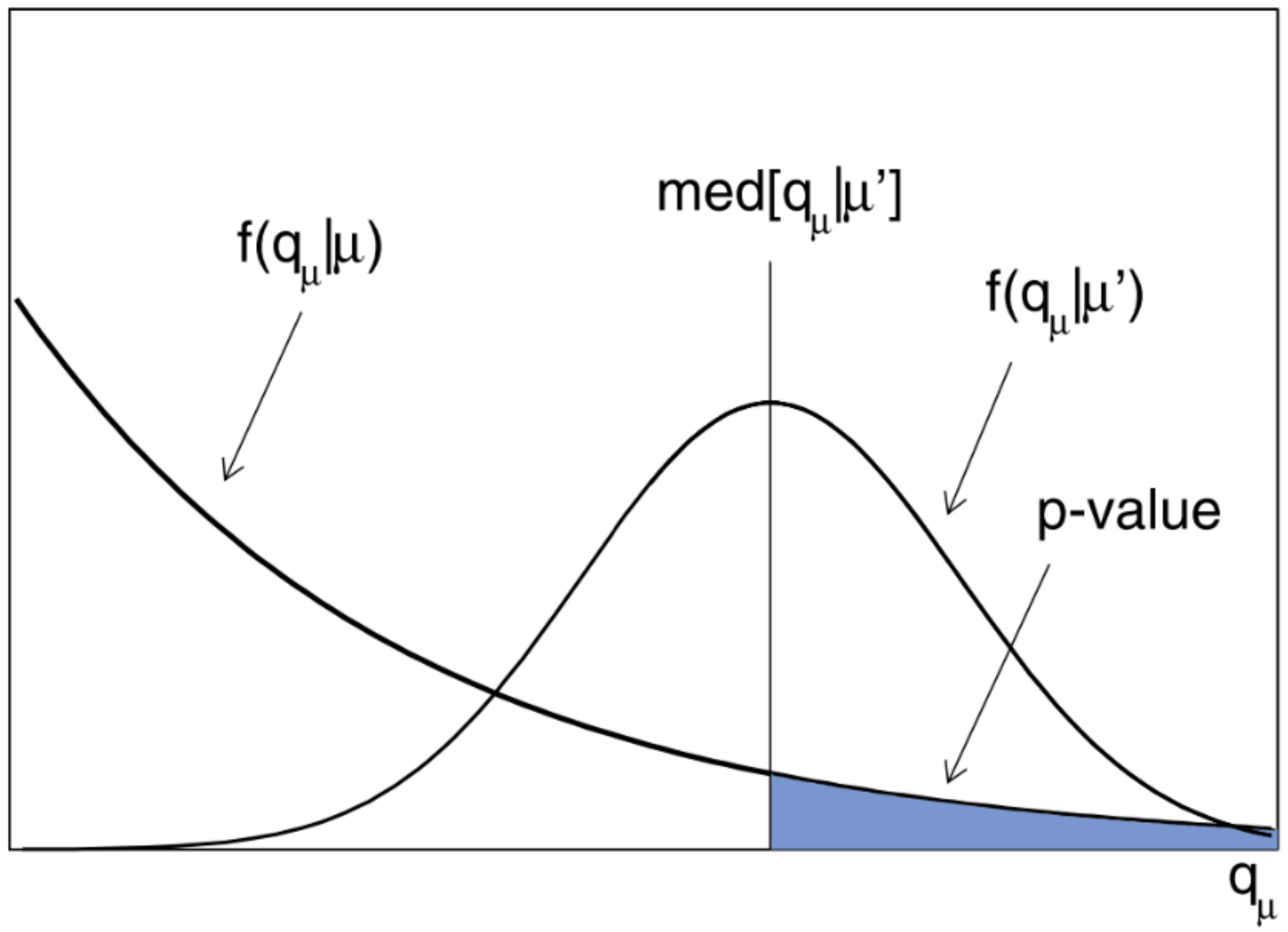
Sensitivity of an experiment

Asimov data sets

- characterised not only in the significance obtained from a single data set, but rather in the **expected median significance to reject different values of μ**;
- estimator is evaluated by using the so called **"Asimov" data set**, replacing the ensemble of real data with a distribution obtained by a **suitable weighted addition of the MC samples**, according to the chosen test statistic.

For exclusion limit setting
sensitivity characterised by the median significance, assuming data generated using the $\mu = 0$ hypothesis, with which one rejects a nonzero value of μ (usually $\mu = 1$ is of greatest interest)

For discovery
sensitivity characterised by the median, under the assumption of the nominal signal model ($\mu = 1$), with which one would reject the background-only ($\mu = 0$) hypothesis



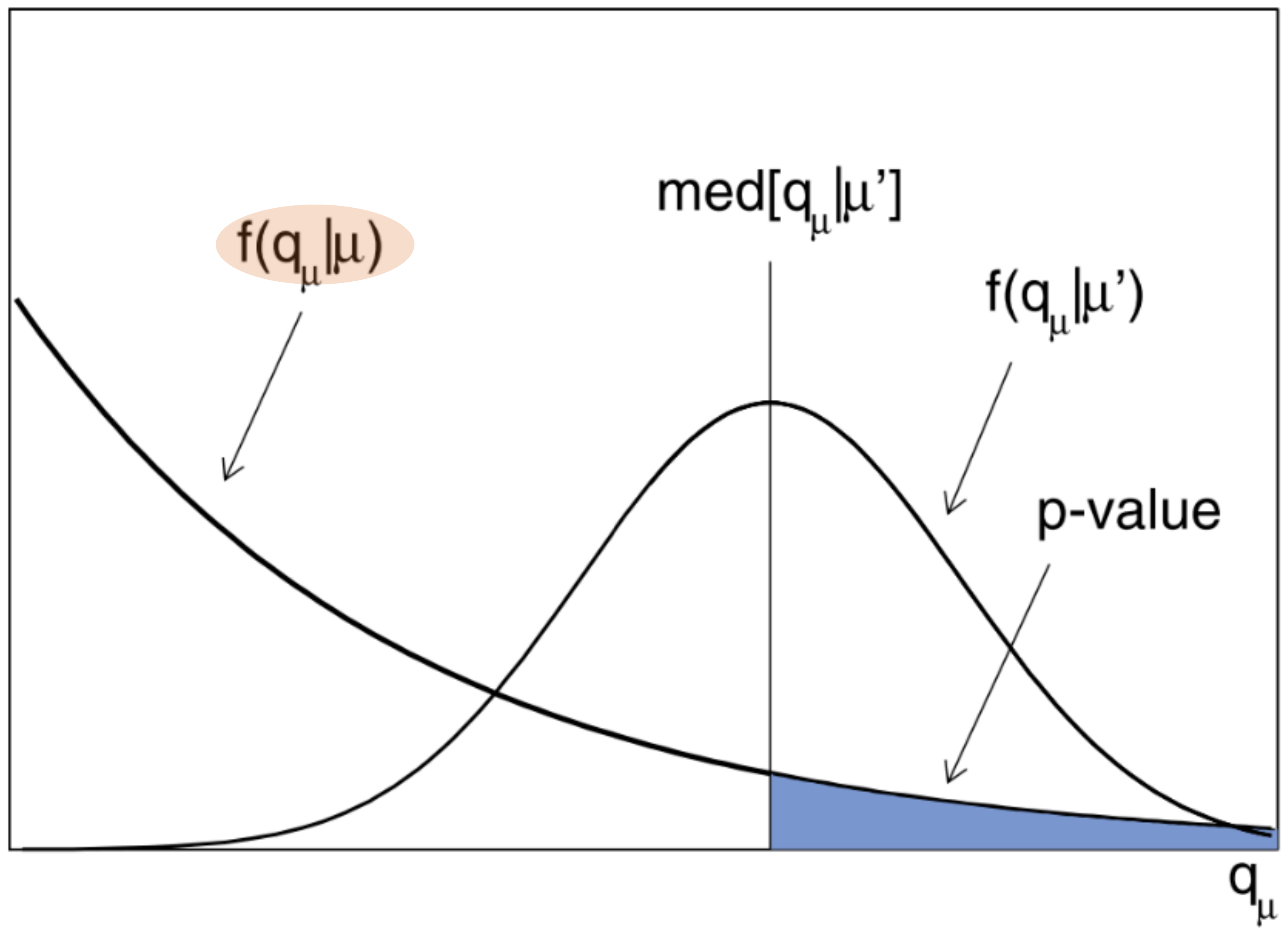
Sensitivity of an experiment

Asimov data sets

- characterised not only in the significance obtained from a single data set, but rather in the expected median significance to reject different values of μ;
- estimator is evaluated by using the so called "Asimov" data set, replacing the ensemble of real data with a distribution obtained by a suitable weighted addition of the MC samples, according to the chosen test statistic.

For exclusion limit setting
sensitivity characterised by the median significance, assuming data generated using the $\mu = 0$ hypothesis, with which one rejects a nonzero value of μ (usually $\mu = 1$ is of greatest interest)

For discovery
sensitivity characterised by the median, under the assumption of the nominal signal model ($\mu = 1$), with which one would reject the background-only ($\mu = 0$) hypothesis



pdf for q_μ assuming a strength parameter μ

Sensitivity of an experiment

Asimov data sets

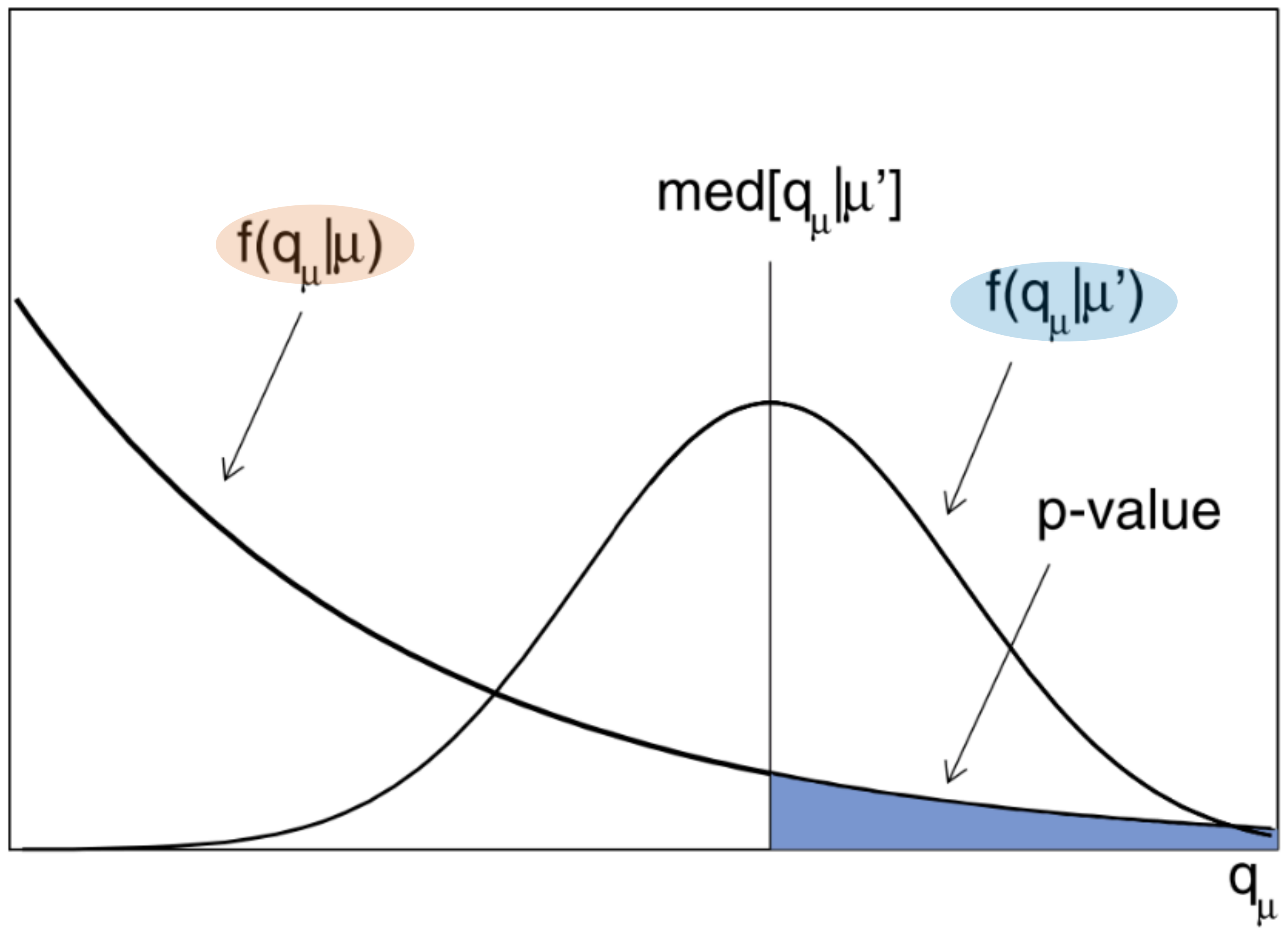
- characterised not only in the significance obtained from a single data set, but rather in the expected median significance to reject different values of μ;
- estimator is evaluated by using the so called "Asimov" data set, replacing the ensemble of real data with a distribution obtained by a suitable weighted addition of the MC samples, according to the chosen test statistic.

For exclusion limit setting

sensitivity characterised by the median significance, assuming data generated using the $\mu = 0$ hypothesis, with which one rejects a nonzero value of μ (usually $\mu = 1$ is of greatest interest)

For discovery

sensitivity characterised by the median, under the assumption of the nominal signal model ($\mu = 1$), with which one would reject the background-only ($\mu = 0$) hypothesis



pdf for q_μ assuming a strength parameter μ

pdf for q_μ assuming a different strength parameter value μ'
shifted to higher value of q_μ, corresponding on average to lower p-values

Sensitivity of an experiment

Asimov data sets

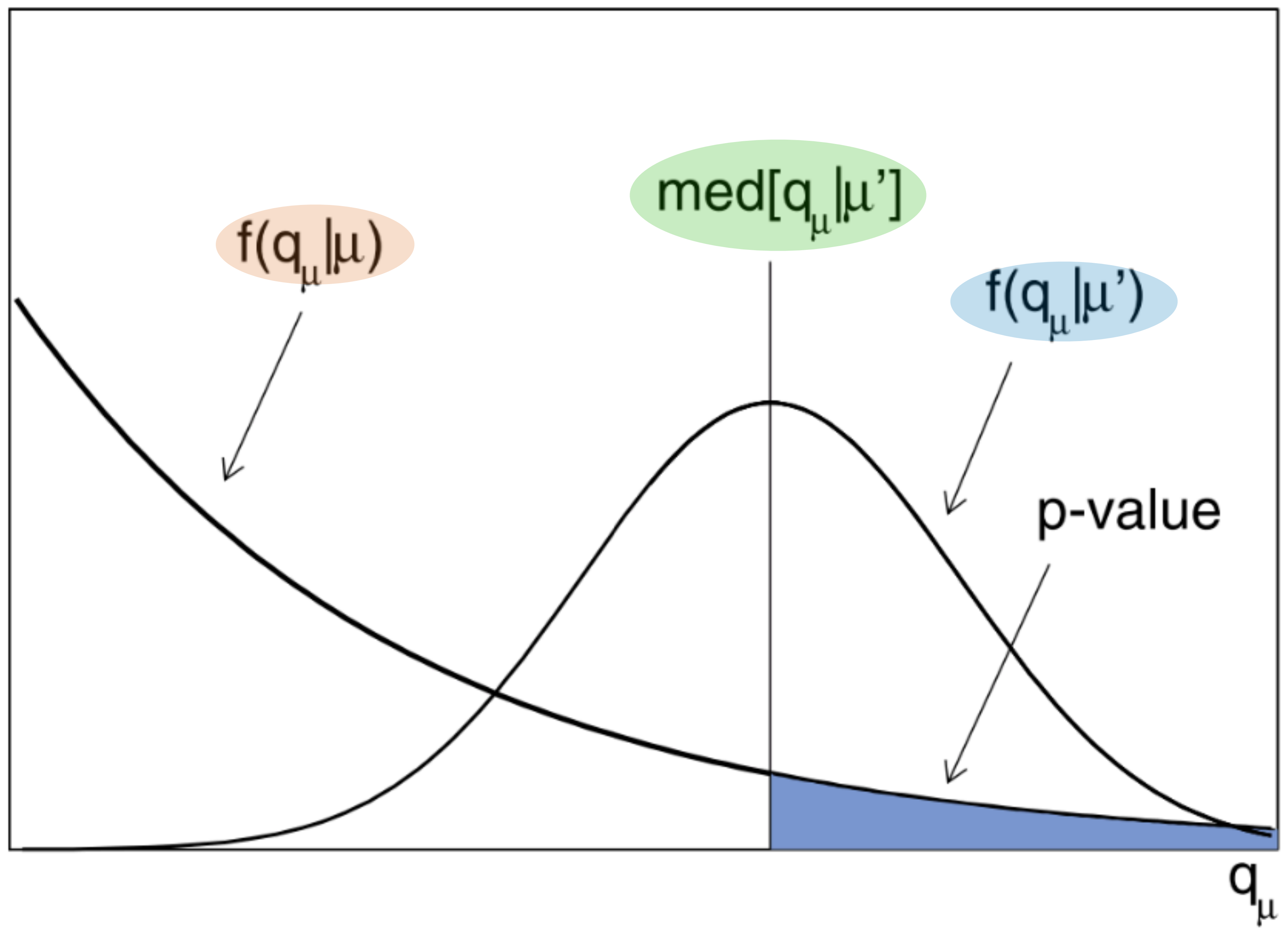
- characterised not only in the significance obtained from a single data set, but rather in the expected median significance to reject different values of μ;
- estimator is evaluated by using the so called "Asimov" data set, replacing the ensemble of real data with a distribution obtained by a suitable weighted addition of the MC samples, according to the chosen test statistic.

For exclusion limit setting

sensitivity characterised by the median significance, assuming data generated using the $\mu = 0$ hypothesis, with which one rejects a nonzero value of μ (usually $\mu = 1$ is of greatest interest)

For discovery

sensitivity characterised by the median, under the assumption of the nominal signal model ($\mu = 1$), with which one would reject the background-only ($\mu = 0$) hypothesis



pdf for q_μ assuming a strength parameter μ

pdf for q_μ assuming a different strength parameter value μ'
shifted to higher value of q_μ, corresponding on average to lower p-values

median q_μ assuming the alternative value μ'
as the p-value is a monotonic function of q_μ, this is equal to the median p-value assuming μ'

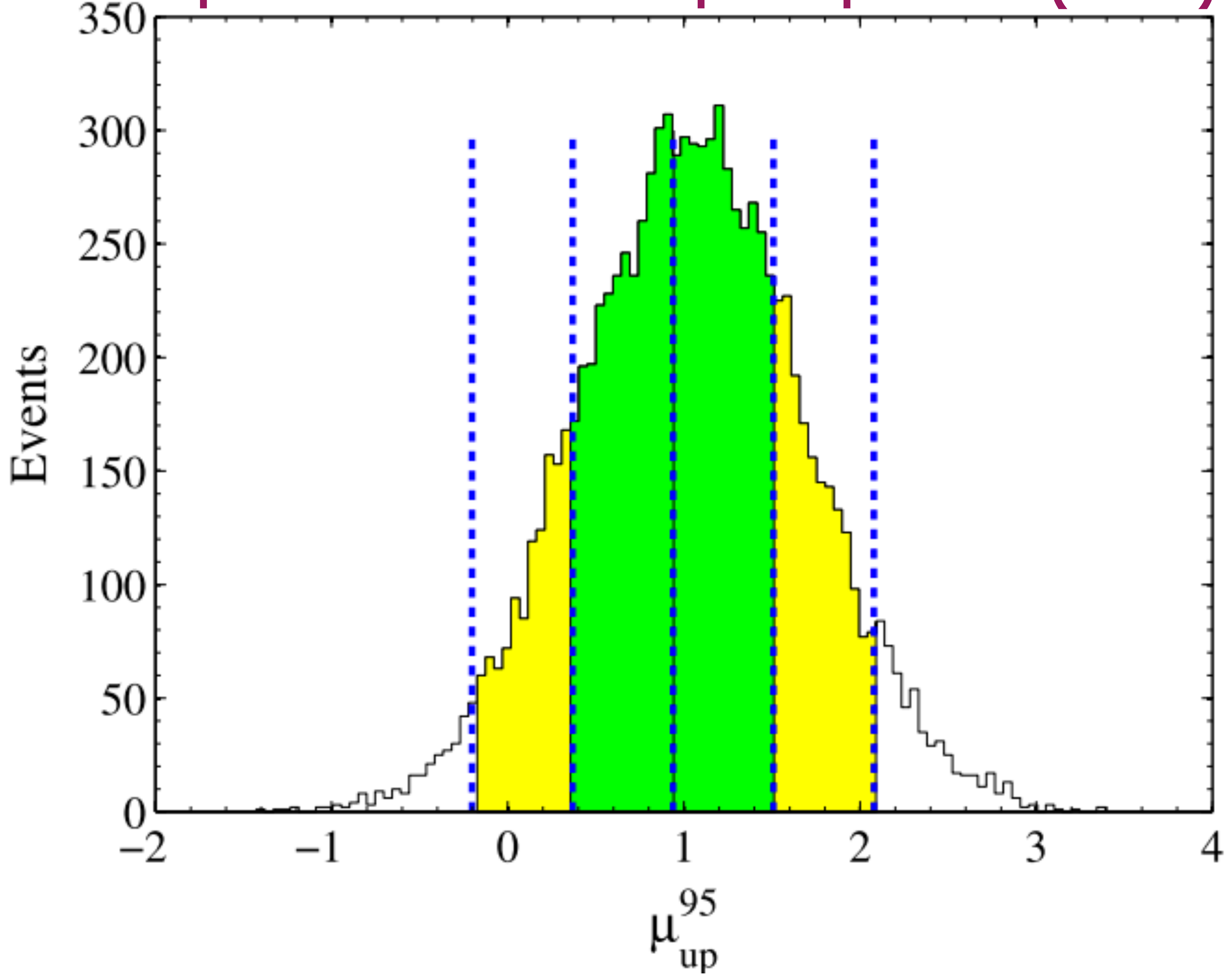
► **this characterises the sensitivity of an experiment!**

Sensitivity of an experiment

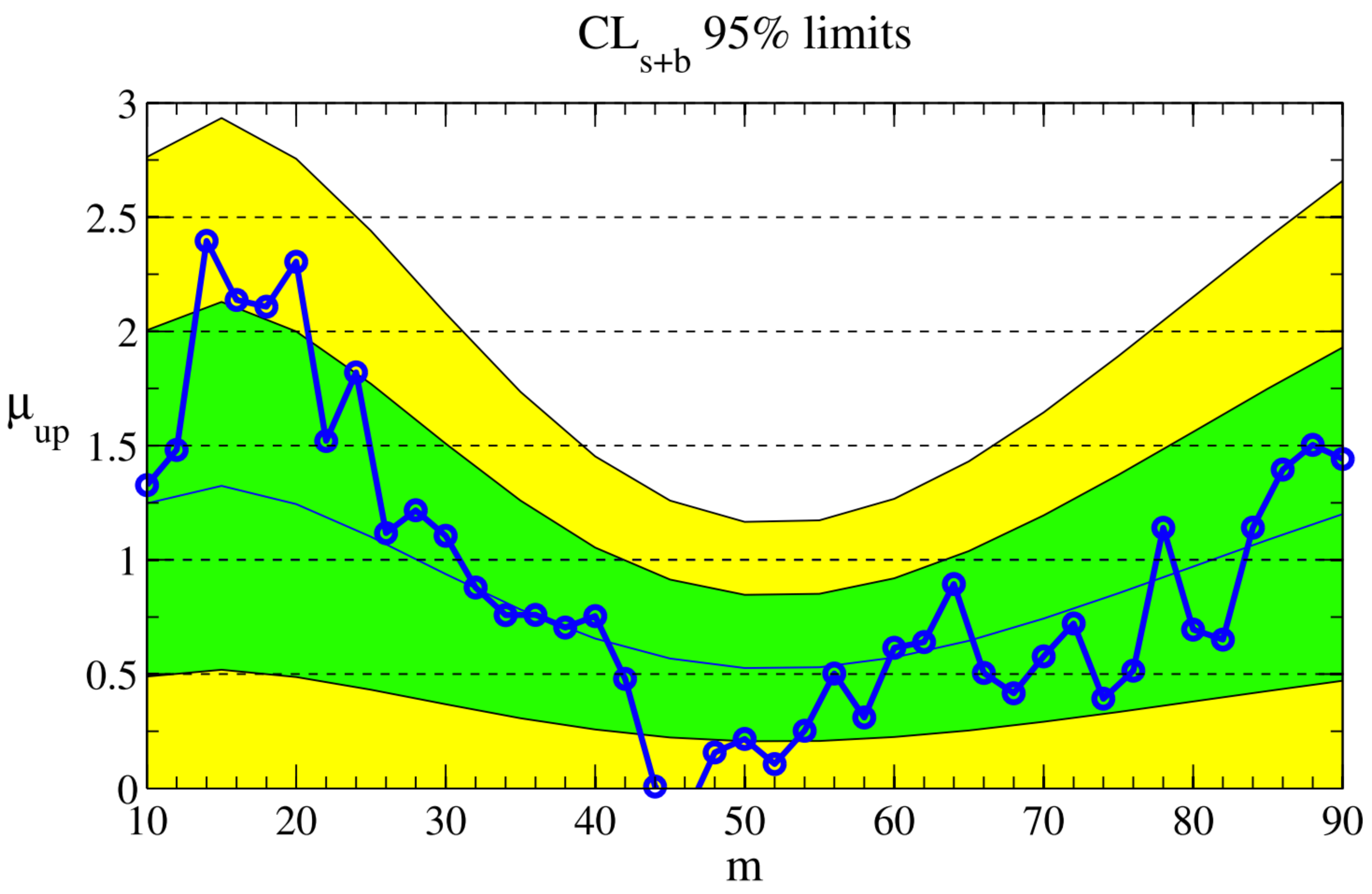
Asimov data sets

- By simulating the experiment many times with Monte Carlo, it is possible to obtain a histogram of the upper limits on μ at 95% CL;
- assuming data corresponding to the background-only hypothesis:

specific value of the peak position (mass)



search for a signal of unknown mass



The **median**, **±1σ** and **±2σ** error bands obtained from the MC pseudo-experiments.

The **vertical lines** (only left plot) are the error bands estimated directly without MC.

CL$_s$ Method

Modified Frequentist CL$_s$ method

- CL for excluding the possibility of signal on top of background (the s+b hypothesis), can be defined as:

$$\alpha_{s+b} = P_{s+b}(q_\mu \le q_{\mu,obs})$$

probability, assuming the presence of both signal and background at their hypothesised levels, that the test statistic would be less than or equal to that observed in the data

Problem

- To quote exclusion limit: confidence level $CL_{s+b} = 1 - \alpha_{s+b}$;
- if too few candidates are observed to account for the estimated background, then any signal, and even the background itself, may be excluded at a high confidence level!

CL$_s$ method is the solution!

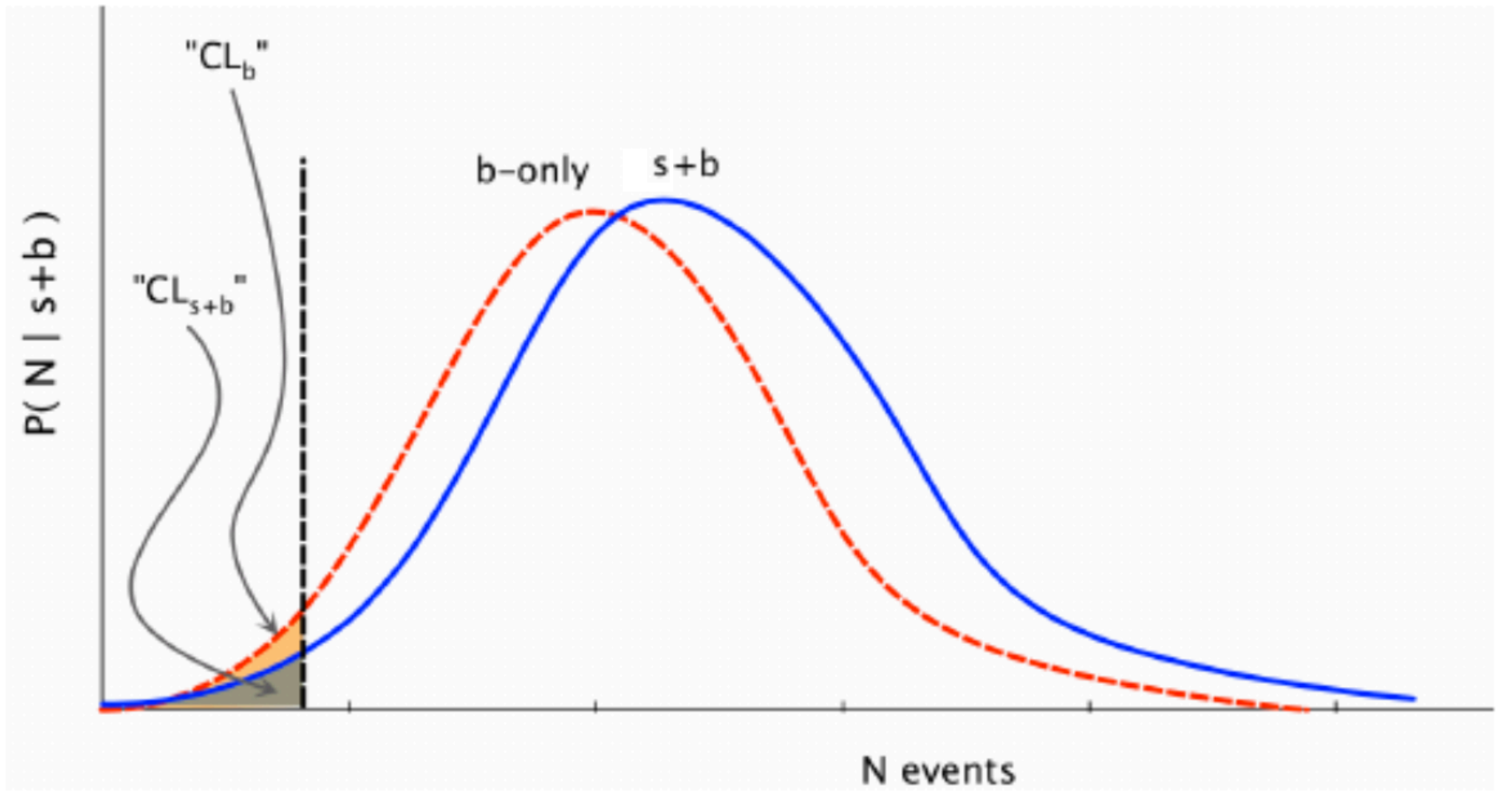
- computing the confidence level for the background alone;

$$\alpha_b = P_b(q_\mu \le q_{\mu,obs})$$

to quantify the confidence of a potential discovery, as it expresses the probability that background processes would give a number of events smaller than or equal to the number of observed candidates

- To quote exclusion limit: confidence level $CL_b = 1 - \alpha_b$;
- CLs computed as the ratio:

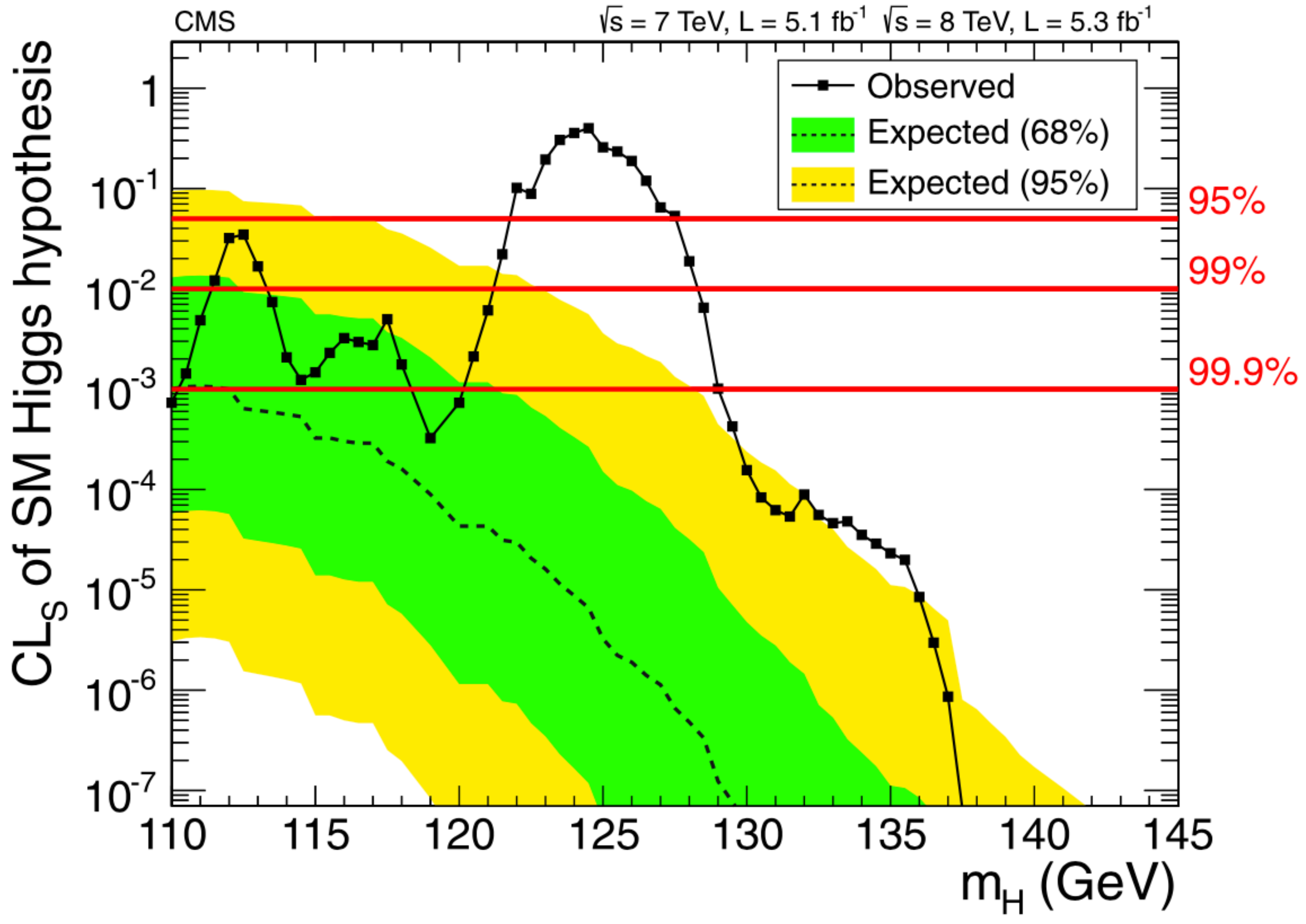
$$CL_s = CL_{s+b}/CL_b$$

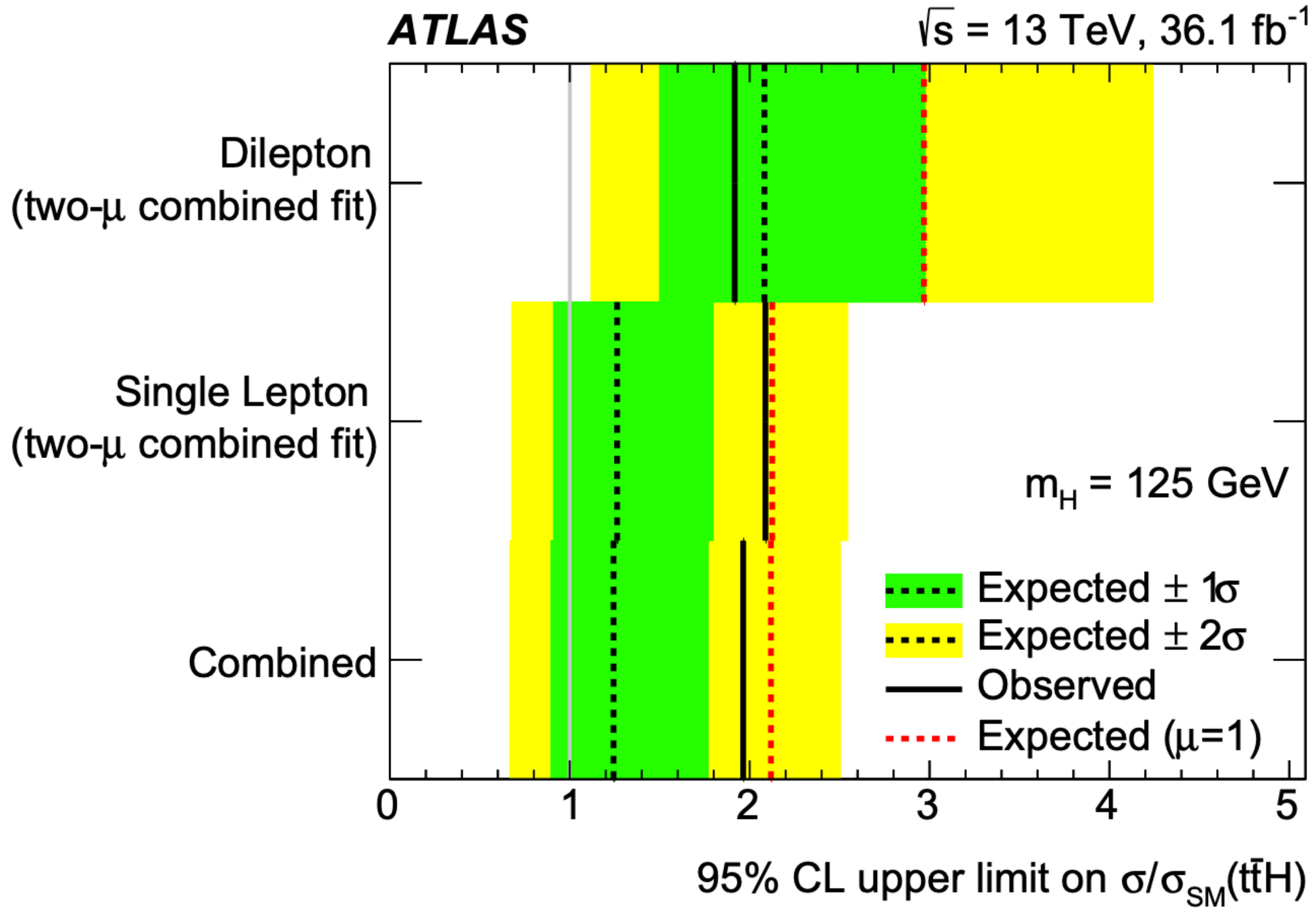


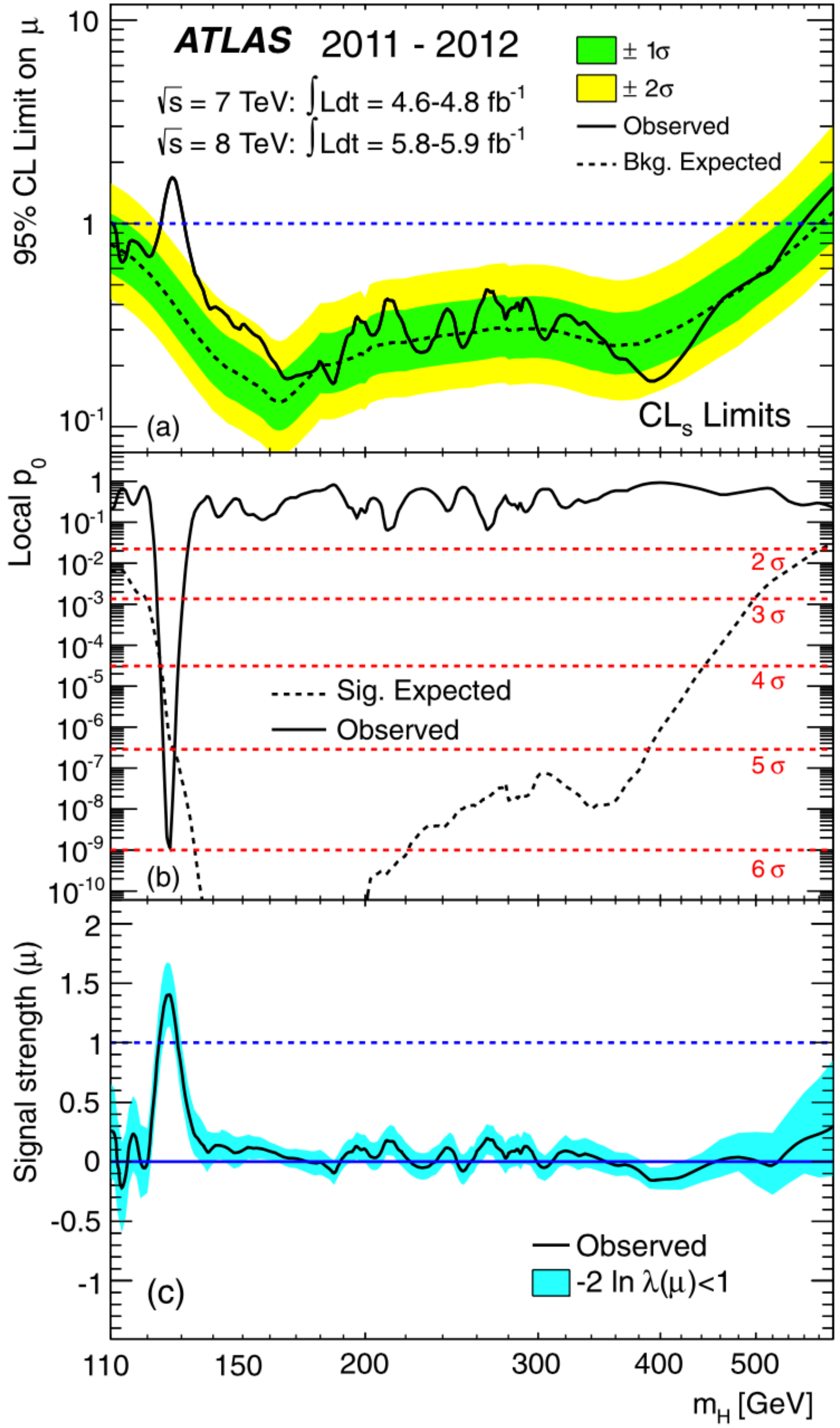
CLs Method - examples

examples

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







Supporting material