Limit determination: Part II

Christian Autermann

24.03.2015

Contents

Yesterday

- Confidence levels
- Confidence level belt construction
- Confidence intervals
- Coverage

Today

- Hypotheses, decisions, and tests
- Limit calculation
- The CLs method

Overview

- Decisions and Tests
  - Statistical hypotheses testing (Frequentist)
  - Significance
  - Signal and Background Hypotheses
  - Neyman-Pearson
- Limits
- CLs
- Tools
- Exercises

Statistical hypotheses testing (Frequentist)

Interpretation of Data

Often (not only in Physics) measured data has to be interpreted within a given theory.

Therefore:

- A hypothesis has to be defined (The Model)
- Perhaps, parameters of the model are determined
- Test the hypothesis with measurements

Goal: Quantify the agreement between theory model and the measured data

Methods: Statistical hypothesis testing, χ^2-Test, Student's *t*-Test, Kolmogorov-Smirnov-Test, ...

The Null-Hypothesis

Karl Popper: *„Empirical theories are characterized by falsifiability. Science should adopt a methodology based on falsification, because no number of experiments can ever prove a theory, but a single experiment can contradict one."*

The **null-hypothesis** typically proposes a general or default position, and can be tested against an **alternative hypothesis**. If the data rejects the null-hypothesis, then one can conclude that the opposite is true.

The null-hypothesis should be defined with great care, and before the experiment is started!

Binomial Example: Coin-toss

Toss a random coin 10 times, it comes down head 7 times.

- Null-hypothesis: The coin is **fair**.
- Alternative hyposesis: The coin is **biased**.

Binomial Example: Coin-toss

Toss a random coin 10 times, it comes down head 7 times.

- Null-hypothesis: The coin is **fair**.
- Alternative hyposesis: The coin is **biased**.

$$P_{binomial}(0.5, 10, 7) = 0.12, \qquad P_{binomial}(0.5, 10, 8) = 0.044,$$
$$P_{binomial}(0.5, 10, 9) = 9.8 \cdot 10^{-3}, \qquad P_{binomial}(0.5, 10, 10) = 9.8 \cdot 10^{-4}$$

The probability to get **head up or head down** seven or more times out of ten is $2 \times (12\% + 4\% + 1\% + 0.1\%) = 34.4\%$, that's not significant.

Binomial Example: Coin-toss

Toss a random coin 10 times, it comes down head 7 times.

- Null-hypothesis: The coin is **fair**.
- Alternative hyposesis: The coin is **biased towards heads**.

$$P_{binomial}(0.5, 10, 7) = 0.12, \qquad P_{binomial}(0.5, 10, 8) = 0.044,$$
$$P_{binomial}(0.5, 10, 9) = 9.8 \cdot 10^{-3}, \qquad P_{binomial}(0.5, 10, 10) = 9.8 \cdot 10^{-4}$$

The probability to get **head up or head down** seven or more times out of ten is $2 \times (12\% + 4\% + 1\% + 0.1\%) = 34.4\%$, that's not significant.

The probability to get **head up** seven or more times out of ten is $12\% + 4\% + 1\% + 0.1\% = 17.2\%$, that's hardly significant either!

Binomial Example: Coin-toss

Toss a random coin 10 times, it comes down head 7 times.

- Null-hypothesis: The coin is **fair**.

$$P_{binomial}(0.5, 10, 7) = 0.12, \qquad P_{binomial}(0.5, 10, 8) = 0.044,$$
$$P_{binomial}(0.5, 10, 9) = 9.8 \cdot 10^{-3}, \qquad P_{binomial}(0.5, 10, 10) = 9.8 \cdot 10^{-4}$$

The probability to get **head up or head down** seven or more times out of ten is $2 \times (12\% + 4\% + 1\% + 0.1\%) = 34.4\%$, that's not significant.

The probability to get **head up** seven or more times out of ten is $12\% + 4\% + 1\% + 0.1\% = 17.2\%$, that's hardly significant either!

To show signs of bias at the 95% confidence level, the coin would need to come down at least 9 out of 10 times head or tails!

Example: Coin toss

" ...

Memo to all teams playing Belgium in the World Cup this year: don't let them use their own coins for the toss. Mathematicians say the coins issued in the eurozone's administrative heartland are more likely to land heads up than down.

...

DIE WELT

„Wir müssen Amerika erneuern"

theguardian

Backlash over Blair's school revolution

UK link to terror snatches

Bigger isn't always better...

King Albert, who appears on Belgian coins, appears to be a bit of a lightweight, according to Polish mathematicians Tomasz Gliszczynski and Waclaw Zawadowski. The two professors and their students at the Podlaska Academy in Siedlce spun a Belgian one euro coin 250 times, and found it landed heads up 140 times. The cent coins proved even more likely to land heads up.

"The euro is struck asymmetrically," Prof Gliszczynski, who teaches statistics, told Germany's Die Welt newspaper. "

...

The Guardian, 4.1.2002

Example: Coin toss

Is that really significant?

Null-hypothesis: 125 times head out of 250, p=0.5.
Two ways to calculate the significance:

1. Statistical uncert.: $\sigma = \sqrt{N \cdot p(1-p)} = 7.9$
   Fluctuation: $s_\alpha = \frac{signal}{uncertainty} = \frac{140-125}{7.9} = 1.9$
2. By summing up the Binomial probabilites from $P_{binomial}(0.5, 250, 140)$ till $P_{binomial}(0.5, 250, 250)$.

The double (single) sided significance that the coin is *not* biased (towards heads) is $\alpha \approx 6.6\%$ ($\alpha \approx 3.3\%$).

Example: Coin toss

Is that really significant?

1. The original „Die Welt" article states further, that the Professor has examined **all** types of 1-Euro and 2-Euro coins. $\rightarrow$ there are 2×15 different common national coins (excluding all rare designs).
2. The **look-elsewhere** effect has to be considered:
3. If „the experiment", i.e. tossing a coin 250 times is repeated 30 times, than it is not credible to quote only the one result with the largest difference!
4. After 30 experiments, a minimal *p*-value of any of these experiments of $1/30 = 3.3\%$ is expected!
5. The observed significance of 3.3% (6.6%) is actually very similar (less significant) as expected

Hypothesis testing

Hypothesis testing

- The Null-hypothesis and the alternative hypothesis have to be clearly defined
- ideally before the experiment is carried out!

- Example 1: The Belgium 2-Euro coin is not biased.
- Example 2: No type of 1-Euro or 2-Euro coins is biased.

- Defining the hypotheses after looking at the outcome of the experiment is cheating (at least)!

Particle physics: Signal and Background Hypotheses

- In particle-physics the „Null-hypothesis" is usually the expectation, that the observed data will follow the predictions of the Standard Model.
- The Standard Model predicts the process cross-sections. With the integrated luminosity and the selection efficiency the background probability density function is determined.
- The **test-statistic** is usually the event yield in a given selection.

Particle physics: Signal and Background Hypotheses

- In particle-physics the „Null-hypothesis” is usually the expectation, that the observed data will follow the predictions of the Standard Model.
- The Standard Model predicts the process cross-sections. With the integrated luminosity and the selection efficiency the background probability density function is determined.
- The **test-statistic** is usually the event yield in a given selection.

Example:

background expectation:

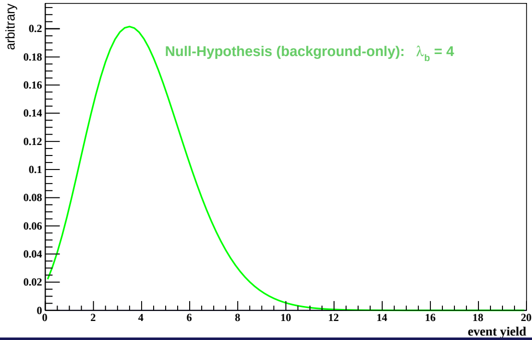
$$b = 4$$

signal expectation:

$$s = 11$$

observed data:

$$d = 7$$



Particle physics: Signal and Background Hypotheses

- In particle-physics the „Null-hypothesis" is usually the expectation, that the observed data will follow the predictions of the Standard Model.
- The Standard Model predicts the process cross-sections. With the integrated luminosity and the selection efficiency the background probability density function is determined.
- The **test-statistic** is usually the event yield in a given selection.

Example:

background expectation:

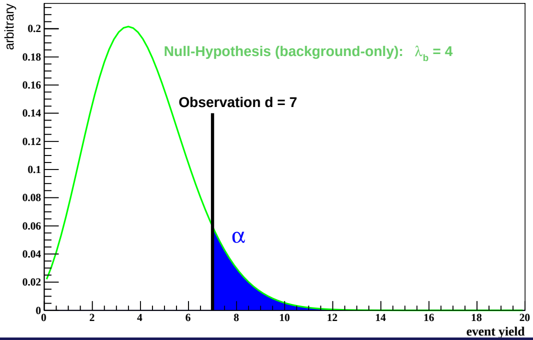
$$b = 4$$

signal expectation:

$$s = 11$$

observed data:

$$d = 7$$



Particle physics: Signal and Background Hypotheses

- In particle-physics the „Null-hypothesis" is usually the expectation, that the observed data will follow the predictions of the Standard Model.
- The Standard Model predicts the process cross-sections. With the integrated luminosity and the selection efficiency the background probability density function is determined.
- The **test-statistic** is usually the event yield in a given selection.

Example:

background expectation:

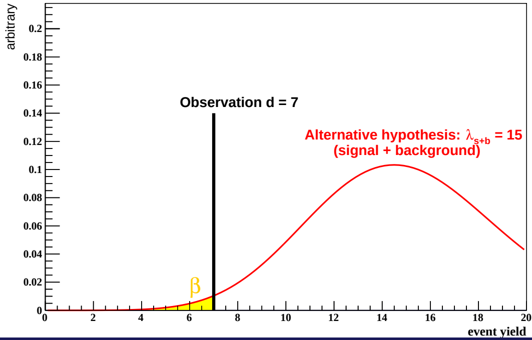
$$b = 4$$

signal expectation:

$$s = 11$$

observed data:

$$d = 7$$



Particle physics: Signal and Background Hypotheses

- In particle-physics the „Null-hypothesis" is usually the expectation, that the observed data will follow the predictions of the Standard Model.
- The Standard Model predicts the process cross-sections. With the integrated luminosity and the selection efficiency the background probability density function is determined.
- The **test-statistic** is usually the event yield in a given selection.

Example:

background expectation:

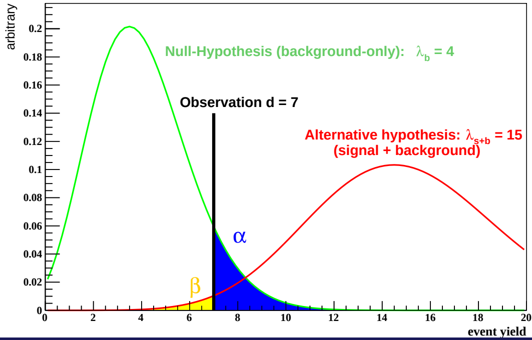
$$b = 4$$

signal expectation:

$$s = 11$$

observed data:

$$d = 7$$



What does it mean?

- The probability (**p-value**) to reject the null-hypothesis H_0, while H_0 is true, is $\int_d^\infty f(x|H_0)dx = \alpha < 1 - \mathrm{CL}_{\text{critical}}$
- Similarly, the probability to reject the alternative hypothesis H_1 if it's true is $\int_{-\infty}^d f(x|H_1)dx = \beta < 1 - \mathrm{CL}_{\text{critical}}$
- Usual choice: $\mathrm{CL}_{\text{critical}} = 95\%$.
- Here: $\alpha = 11\%$, $\beta = 1.8\%$

Example:

background expectation:

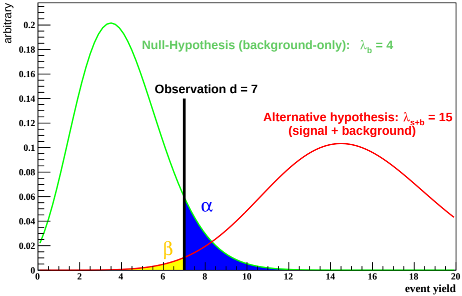
$$b = 4$$

signal expectation:

$$s = 11$$

observed data:

$$d = 7$$



Hypothesis testing

How to distinguish between hypotheses?

- In general, there is no „observed data" defining the value of the test statistics x_d and therefore the probability to accept or to reject a hypothesis
- In general we might be interested to find x_d such, that the null-hypothesis or the alternative hypothesis are accepted or rejected with certain efficiencies.

Type I and Type II Errors

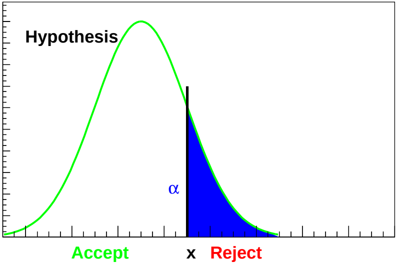
If we regard two mutually exclusive hypotheses that are either true or false

There are four possible outcomes:

- Accepting a true hypothesis
- Rejecting a wrong hypothesis
- Rejecting a true hypothesis (Type I error)
- Accepting a wrong hypothesis (Type II error)

If α **is the significance of the test**, then Type I errors are bound to occur less than or equal to α:

$$\int_X^\infty P_h(x)dx \leq \alpha$$



Type I and Type II Errors

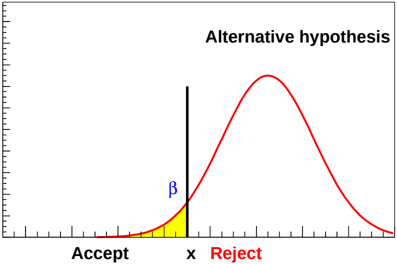
If we regard two mutually exclusive hypotheses that are either true or false

There are four possible outcomes:

- Accepting a true hypothesis
- Rejecting a wrong hypothesis
- Rejecting a true hypothesis (Type I error)
- Accepting a wrong hypothesis (Type II error)

The probability to mistakenly accept the hypothesis H_a is β, and $1-\beta$ **is the power of the test**:

$$\int_{-\infty}^{X} P_a(x)dx \leq \beta$$



Type I and Type II Errors

If we regard two mutually exclusive hypotheses that are either true or false

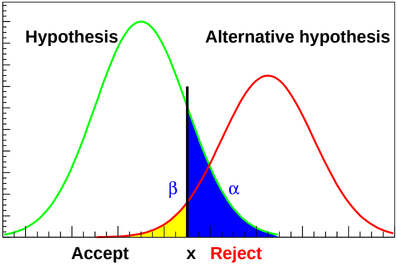
There are four possible outcomes:

- Accepting a true hypothesis
- Rejecting a wrong hypothesis
- Rejecting a true hypothesis (Type I error)
- Accepting a wrong hypothesis (Type II error)

Both, α and β should be as small as possible, but there is a tradeoff between minimizing α and β.

The relative importance of α or β depends on the problem!

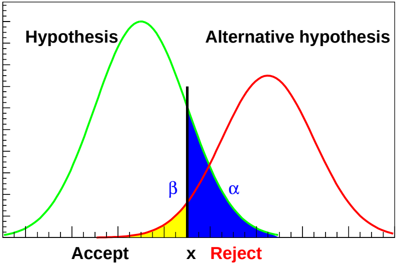
Example b-tagging: decide between *b*-jet and light-jet hypothesis.



Example: Cow-fever epidemic

Let's assume the really dangerous „cow-fever" disease results always in a fever of 39.7 C with a Gaussian spread of 0.2 C. Patients with normal flu only have temperature 39.2 ± 0.2 C and are 100 times more likely.

Where to cut (fever-threshold for treating a patient ambulant or stationary) if we want a test-power of $1 - \beta = 90\%$?

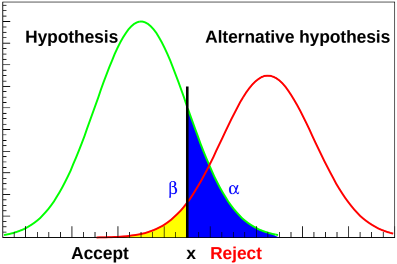


Example: Cow-fever epidemic

Let's assume the really dangerous „cow-fever" disease results always in a fever of 39.7 C with a Gaussian spread of 0.2 C. Patients with normal flu only have temperature 39.2 ± 0.2 C and are 100 times more likely.

Where to cut (fever-threshold for treating a patient ambulant or stationary) if we want a test-power of $1 - \beta = 90\%$?

Case 1: Accepting $\beta = 10\%$ of normal flu patients, leads to the rejection of $\sim 15\%$ cow-fever patients (Type I error) and more than 92% beds are occupied by normal flu patients.

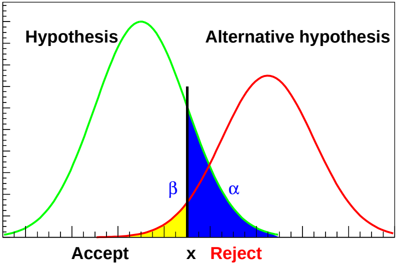


Example: Cow-fever epidemic

Let's assume the really dangerous „cow-fever" disease results always in a fever of 39.7 C with a Gaussian spread of 0.2 C. Patients with normal flu only have temperature 39.2 $\pm$ 0.2 C and are 100 times more likely.

Where to cut (fever-threshold for treating a patient ambulant or stationary) if we want a test-significance of $\alpha = 5\%$?

Case 2: Accepting 95% of cow-fever patients (significance $\alpha = 5\%$) by cutting at 39.37 C leads to a Type II error (accepted normal flu patients) of $\beta \approx 80\%$. Now, more than 98.8% beds are occupied by normal flu patients.

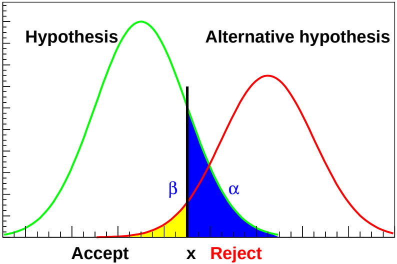


Example: Cow-fever epidemic

Let's assume the really dangerous „cow-fever" disease results always in a fever of 39.7 C with a Gaussian spread of 0.2 C. Patients with normal flu only have temperature 39.2 $\pm$ 0.2 C and are 100 times more likely.

Where to cut (fever-threshold for treating a patient ambulant or stationary) if we want a test-significance of $\alpha = 5\%$?

In reality the clinic will probably choose α and β according to other constraints, e.g. the number of available beds, etc.



The best test, for which both α and β are small as possible is called a **Neyman-Pearson test**:

The Neyman-Pearson lemma

When performing a hypothesis test between two hypotheses H_h and H_a, then the **likelihood-ratio test** which rejects H_a in favour of H_h when

$$Q = \frac{L_{H_h}(x)}{L_{H_a}(x)} \geq Q_0 \qquad \text{for a given significance } \alpha$$

is the **most powerful test-statistic** to minimize both α and β.

- Both hypothesis H_h and H_a have to be explicitely defined and have to be simple.
- The acceptance region giving the highest power $1 - \beta$ for a given significance α is the region comprised by the above (in)equation.
- In the one-dimensional case, a cut on x for a specific α (e.g. b-tag efficiency) determines β (and therefore the purity).

Overview

- Decisions and Tests
- Limits
  - Event yields
- CLs
- Tools
- Exercises

Limits

http://pix.echtlustig.com

Limit definition

- The **observed limit** on the signal event-yield at $\mathrm{CL}_{s+b} = 95\%$ is defined as the s, for which

$$\begin{aligned} \beta &= \int_{-\infty}^{d} Q(x|H_1)dx \\ &\leq 1 - \mathrm{CL}_{s+b}. \end{aligned}$$

- In this example $Q(x|H_1)$ is the Poisson p.d.f. with a mean $\lambda = s + b$.

Example:

background expectation:

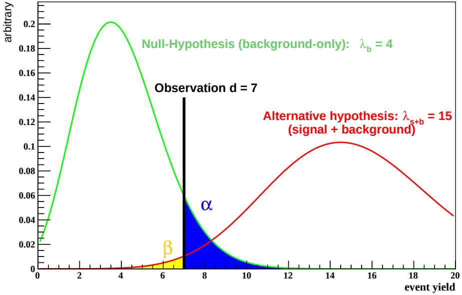
$$b = 4$$

observed data:

$$d = 7$$

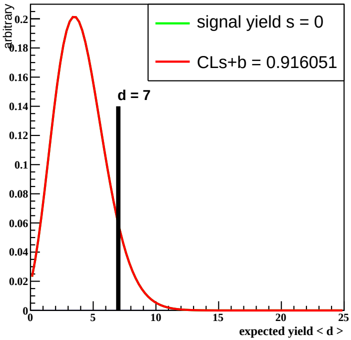
limit on s + b at 95% C.L.:

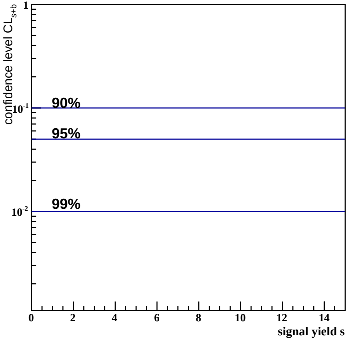
$$s + b = 12.5$$



Limit calculation: Example

background expectation: $b = 4$
observed data: $d = 7$
limit on s + b at 95% C.L.: $s + b = 12.5$





PDF reader with Java (e.g. Adobe Acrobat) necessary for animation

Remarks on limit calculation

- The test-statistic $Q(x|H_1)$ was in the example a Poisson pdf modelling the under the H_1 hypothesis expected statistical uncertainties of the measurement (the data).
- The test-statistics $Q(x|H_1)$ may incorporate also systematical uncertainties on the background σ_b and on the signal estimation σ_s, e.g.

$$Q(x|H_1) = \text{Poisson}(\lambda_{s+b}) \otimes \text{Gauss}(b, \sigma_b) \otimes \text{Gauss}(s, \sigma_s)$$

- In general, $Q(x|H_0)$ and $Q(x|H_1)$ may be defined by a likelihood that distinguishes both hypotheses
- Essentially a one-dimensional non-linear minimization problem, numerical solution quite time consuming
- In this case, the agreement of the measured data with the background-only expectation, i.e. the null-hypothesis H_0, is not directly considered.
- The limit on the signal event yield does not depend on the expected signal event yield!

Multi-channel limits

- Since likelihood functions are multiplicative, multiple statistically exclusive channels i.e. from different exclusive selections or histogram bins can be easily combined:

$$L(x) = \prod_{b=1}^{\text{bins}} L_b(x)$$

  where the $L_b(x)$ are the test-statistics of the individual single-bin counting experiments.

- Systematic uncertainties affecting the estimation of the background or signal prediction σ_i^b and σ_i^s may be correlated among different bins. This can be considered when drawing pseudo-data from the hypotheses test-statistics.

Expected limits and limit uncertainties

- Expected limits are usually defined as the 50% quantile, i.e. the median, of the distribution of observed limits for a number of pseudo-experiments; where the pseudo-observations are drawn according to the background-only null-hypothesis test-statistic H_0.
- The $\pm 1\sigma$ uncertainties on the expected limit are equivilantly the 16% and 84% quantiles.

Overview

- Decisions and Tests
- Limits
- CLs
  - Modified frequentist approach
  - LEP Higgs limit
  - LHC Higgs limits
  - LHC SUSY limits
- Tools
- Exercises

Limit definition

- An **observed limit** on the signal event-yield at $\mathrm{CL}_{s+b} = 95\%$ can be defined as the s, for which

$$\beta = \sum_{d=0}^{d_{\text{obs}}} \frac{e^{s+b}(s+b)^d}{d!} \leq 1 - \mathrm{CL}_{s+b}.$$

Example:

background expectation:

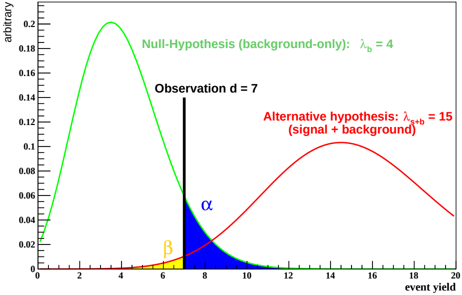
$$b = 4$$

observed data:

$$d = 7$$

limit on s + b at 95% C.L.:

$$s + b = 12.5$$



The modified Frequentist procedure (CLs)

CLs is a frequentist like statistical analysis which avoids excluding or discovering signals, that the analysis is not really sensitive to.

The null-hypothesis is that there is no signal and the alternate hypothesis that it exists.

The modified Frequentist procedure (CLs)

CLs is a frequentist like statistical analysis which avoids excluding or discovering signals, that the analysis is not really sensitive to.

The null-hypothesis is that there is no signal and the alternate hypothesis that it exists.

Example: Observing less than the mean expected background events could be accommodated best with a negative signal cross-section. The exclusion may be so strong that even zero signal is excluded at a certain confidence level.

The modified Frequentist procedure (CLs)

CLs is a frequentist like statistical analysis which avoids excluding or discovering signals, that the analysis is not really sensitive to.

The null-hypothesis is that there is no signal and the alternate hypothesis that it exists.

Example: Observing less than the mean expected background events could be accommodated best with a negative signal cross-section. The exclusion may be so strong that even zero signal is excluded at a certain confidence level.

This is a perfectly valid result in terms of statistics, but it says more about fluctuations of the (known) background, than about the hypothetical signal we are interested in!

The modified Frequentist procedure (CLs)

CLs is a frequentist like statistical analysis which avoids excluding or discovering signals, that the analysis is not really sensitive to.

The null-hypothesis is that there is no signal and the alternate hypothesis that it exists.

Example: Observing less than the mean expected background events could be accommodated best with a negative signal cross-section. The exclusion may be so strong that even zero signal is excluded at a certain confidence level.

This is a perfectly valid result in terms of statistics, but it says more about fluctuations of the (known) background, than about the hypothetical signal we are interested in!

$\longrightarrow$ Normalize the confidence level in the signal+background hypothesis CL_{s+b} to the confidence level for the background-only hypothesis CL_b.

Introducing CLs method

The modified frequentist re-normalization is simply:

$$CL_s = \frac{CL_{s+b}}{CL_b}$$

CL$_s$ gives an approximation to the confidence in the signal hypothesis one might have obtained if the experiment had been performed in the complete absence of background. CL$_s$ tries to reduce the dependency on the uncertainty due to the background.

Strictly, CL$_s$ is not a confidence, but a ratio of confidences.

- Consequentially, the false exclusion rate is generally less than the nominal rate CL,
- it increases the „coverage” of the analysis,
- it gives a consistent performance compared to CL$_{s+b}$ at small expected signal but different background rates.

CLs: Single counting experiment

For a counting experiment with a single channel, CL_s takes the following form:

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{\text{Poisson}(s+b, d_{obs})}{\text{Poisson}(b, d_{obs})}$$

where $s+b$ (or b) come from the Poisson distributions of number of events for the signal+background (background-only) hypotheses, and d_{obs} is the number of events observed.

The modified frequentist signal exclusion confidence becomes:

$$CL = 1 - \frac{\sum_{n=0}^{d_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{d_{obs}} \frac{e^{-b}b^n}{n!}}$$

Which is (accidentally!) similar to the result we obtained by computing the constrained Bayesian integral with a flat prior.

CLs: Likelihood-ratio test-statistic for a counting experiment

Using a likelihood-ratio as test-statistic to compute CL_{s+b} and CL_b:

$$X = \frac{\mathrm{Poisson}(s(m_H) + b, d_{obs})}{\mathrm{Poisson}(b, d_{obs})}$$

Where the expected signal s depends e.g. on a model parameter (e.g. the Higgs mass m_H). Likelihoods are multiplicative, different N channels can be combined:

$$X(m_h) = \prod_i^N X_i(m_h)$$

If d_i data events are observed, then this leads to a value X_{obs} of the test-statistics. CL_{s+b} is then given by:

$$\begin{aligned} CL_{s+b} &= P_{s+b}(X \leq X_{obs}) \\ &= \int_{-\infty}^{X_{obs}} \frac{dX_{s+b}}{dx} dx \end{aligned}$$

where dX_{s+b}/dx is the p.d.f. distribution of the test-statistics X for signal+background experiments.

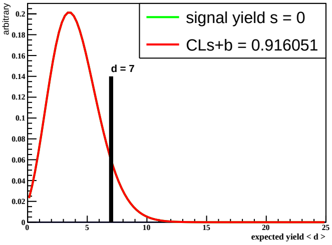
Frequentists confidence levels

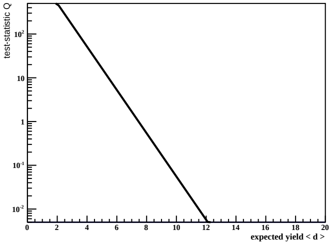
The Confidence Level CL_{s+b} is then in the case of a N channel couting experiment calculated as:

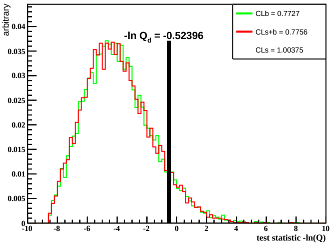
$$CL_{s+b} = \sum_{X(d_i') \leq X(d_i)} \prod_{i=1}^{N} \frac{e^{s_i+b_i}(s_i+b_i)^{d_i'}}{d_i'!}$$

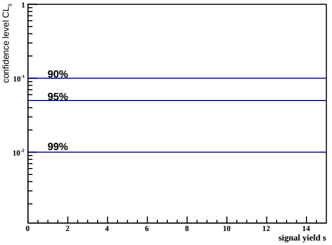
Small values of CL_{s+b} indicate poor compability with the $s+b$ hypothesis and favour the background-only hypothesis.

CL_b is calculated likewise.



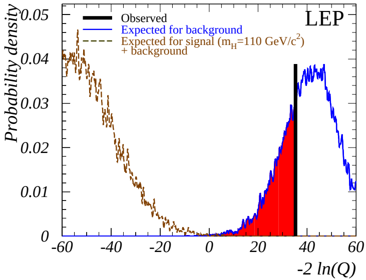


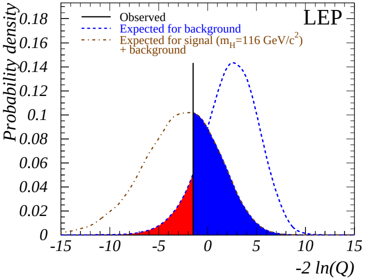


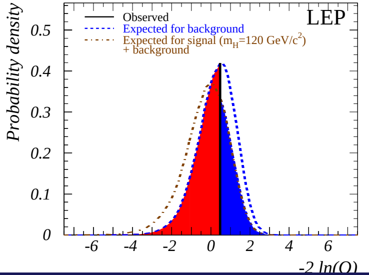


PDF reader with Java (e.g. Adobe Acrobat) necessary for animation

LEP combined SM Higgs limits





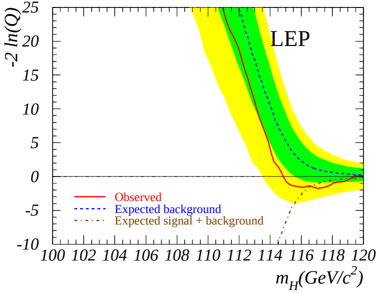


Probability density functions corresponding to fixed test-masses for the b and $s+b$ hypotheses. The observed test-statistic $-2\ln Q = Q_{obs}$ is indicated by the vertical line.

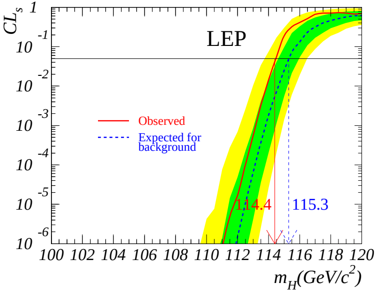
The shaded areas refer to $1 - CL_b$ and CL_{s+b}.

LEP combined SM Higgs limits

LHWG Note/2002-01 „Search for the Standard Model Higgs Boson at LEP”

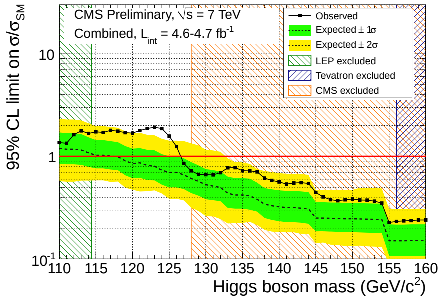


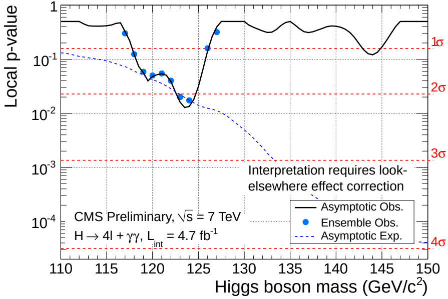
Observed and expected behaviour of the test-statistic $-2 \ln Q$ as a function of the test mass m_H.



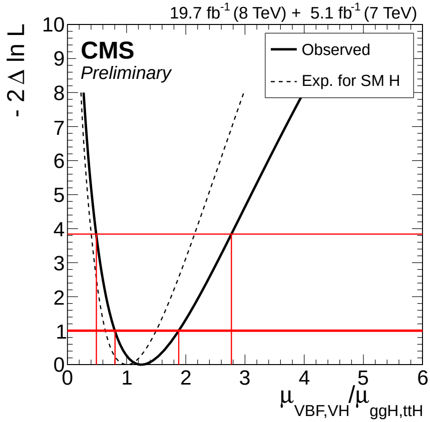
Confidence level CL_s for the signal+background hypothesis. The intersection of the observed limit with the horizontal $CL_s = 0.05$ line determines the 95% CL lower mass limit.

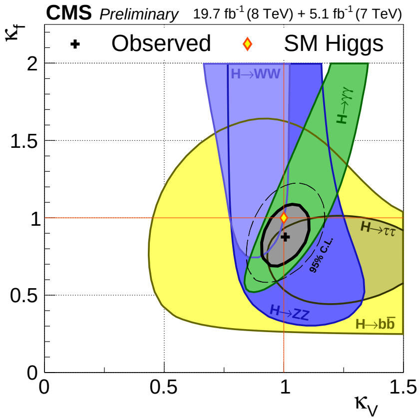
LHC Higgs limits (2011, 7 TeV, 5 fb^{-1})



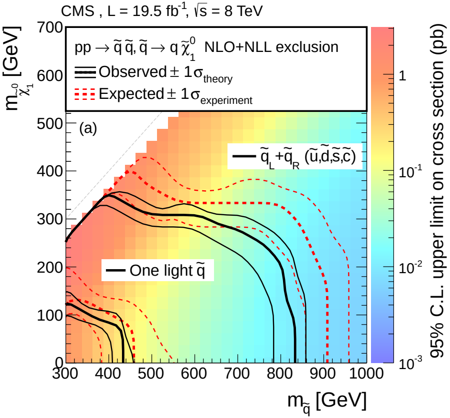


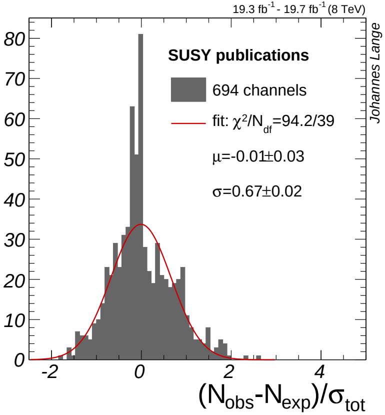
LHC Higgs: measurement





SUSY 2D limits





Overview

- Decisions and Tests
- Limits
- CLs
- Tools
  - TLimit
  - RooStat
- Exercises

CLs method

Alex Read, „Presentation of search results: the CLs technique", Journal of Physics G: Nucl. Part. Phys. **28** p. 2693-2704 (2002).

Tom Junk, „Confidence level computation for combining searches with small statistics", NIM **A434**, p. 435-443, (1999).

ROOT::TLimit

```
TFile* infile= new TFile("file.root","READ"); infile->cd();
TH1* sh=(TH1*)infile->Get("signal");
TH1* bh=(TH1*)infile->Get("background");
TH1* dh=(TH1*)infile->Get("data");
TLimitDataSource * mydata = new TLimitDataSource(sh,bh,dh);
TConfidenceLevel * myconf =
                  TLimit::ComputeLimit(mydatasource,50000);
cout << " CLs  :"<< myconfidence->CLs()  << "\n"
     << " CLsb :"<< myconfidence->CLsb() << "\n"
     << " CLb  :"<< myconfidence->CLb()  << "\n"
     << "<CLs >:"<< myconfidence->GetExpectedCLs_b()  <<"\n"
     << "<CLsb>:"<< myconfidence->GetExpectedCLsb_b() <<"\n"
     << "<CLb> :"<< myconfidence->GetExpectedCLb_b()<< endl;
```

CLs method

Alex Read, „Presentation of search results: the CLs technique”, Journal of Physics G: Nucl. Part. Phys. **28** p. 2693-2704 (2002).

Tom Junk, „Confidence level computation for combining searches with small statistics”, NIM **A434**, p. 435-443, (1999).

ROOT::TLimit

```
TFile* infile= new TFile("file.root","READ"); infile->cd();
TH1* sh=(TH1*)infile->Get("signal");
TH1* bh=(TH1*)infile->Get("background");
TH1* dh=(TH1*)infile->Get("data");
TLimitDataSource* mydata = new TLimitDataSource(sh,bh,dh);
TConfidenc

cout << " CLs  :"<< myconfidence->CLs()  << "\n"
     << " CLsb :"<< myconfidence->CLsb() << "\n"
     << " CLb  :"<< myconfidence->CLb()  << "\n"
     << "<CLs >:"<< myconfidence->GetExpectedCLs_b()  <<"\n"
     << "<CLsb>:"<< myconfidence->GetExpectedCLsb_b() <<"\n"
     << "<CLb> :"<< myconfidence->GetExpectedCLb_b()<< endl;
```

can handle also systematic uncertainties

Root Statistic Tools

RooStat

RooStats is a project to create statistical tools for ROOT built on top of RooFit and distributed in ROOT. It is a joint project between the LHC experiments and the ROOT team. Included since ROOT v5.22.

Detailed information with many nice tutorials and examples:
`https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome`
`https://twiki.cern.ch/twiki/bin/view/RooStats/TutorialsOctober2009`

Core developers: K. Cranmer (ATLAS), Gregory Schott (CMS), Wouter Verkerke (RooFit), Lorenzo Moneta (ROOT).

Overview

- Decisions and Tests
- Limits
- CLs
- Tools
- Exercises
  - Task 3: Modified frequentist upper limit
  - Task 4: Upper limit on measured negative signal yield: Frequentist vs Bayesian

Task 3: Modified frequentist upper limit

Modified frequentist (CLs) limits

Revisit the last item of task 2:

- $\mu_{bkg} = 3$, $N_{obs} = 0$

This time set a 90% upper limit using the modified frequentist approach:

$$CL_s = CL(S+B)/CL(B) = 0.1$$

Note: the definitions are:

$CL(B) = p(\mu_{bkg}, N_{obs}) = \sum_{i \leq N_{obs}} e^{-\mu_{bkg}} \frac{\mu_{bkg}^i}{i!}$ and

$CL(S+B)p(\mu_{bkg} + \mu_{sig}, N_{obs}) = \sum_{i \leq N_{obs}} e^{-(\mu_{bkg}+\mu_{sig})} \frac{(\mu_{bkg}+\mu_{sig})^i}{i!}$

Task 4: Frequentist vs Bayesian

After background subtraction, an experiment "observes" a yield of -2 ± 1 particles. The uncertainty is assumed to be Gaussian. Determine an 90% upper limit μ_{lim} for the expectation value of the number of events using the

- Frequentist approach: taking the result at face value
  Instruction: determine the 90% upper limit from

$$CL = \int_{\mu_{lim}}^{\infty} dx' \, \frac{1}{2\pi} e^{\frac{-(x'+2)^2}{2}} = 10\%.$$

  Hint: The solution can be read off from the CL curves for a Gaussian.

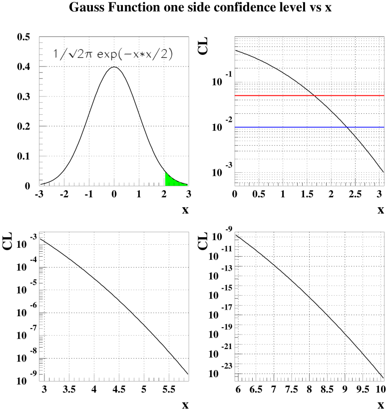
- Bayesian approach: the result has to be positive
  Instruction: determine the 90% upper limit from

$$CL = \frac{\int_{\mu_{lim}}^{\infty} dx' \, \frac{1}{2\pi} e^{\frac{-(x'+2)^2}{2}} \theta(x')}{\int_{0}^{\infty} dx' \, \frac{1}{2\pi} e^{\frac{-(x'+2)^2}{2}}} = 10\%.$$

  Hint: The $\theta(x')$ can be ignored since only positive values of μ_{lim} will solve the equations. The solutions to both integrals can be read off from the CL curves for a Gaussian.

Appendix 1 - One sided Gaussian confidence levels

$$\mathrm{CL}(x) = \int_x^\infty dx' \frac{1}{\sqrt{2\pi}} e^{-x'^2/2}$$



Appendix 2 - Two sided Gaussian confidence levels

$$\mathrm{CL}(x) = 2 \int_x^\infty dx' \, \frac{1}{\sqrt{2\pi}} e^{-x'^2/2}$$