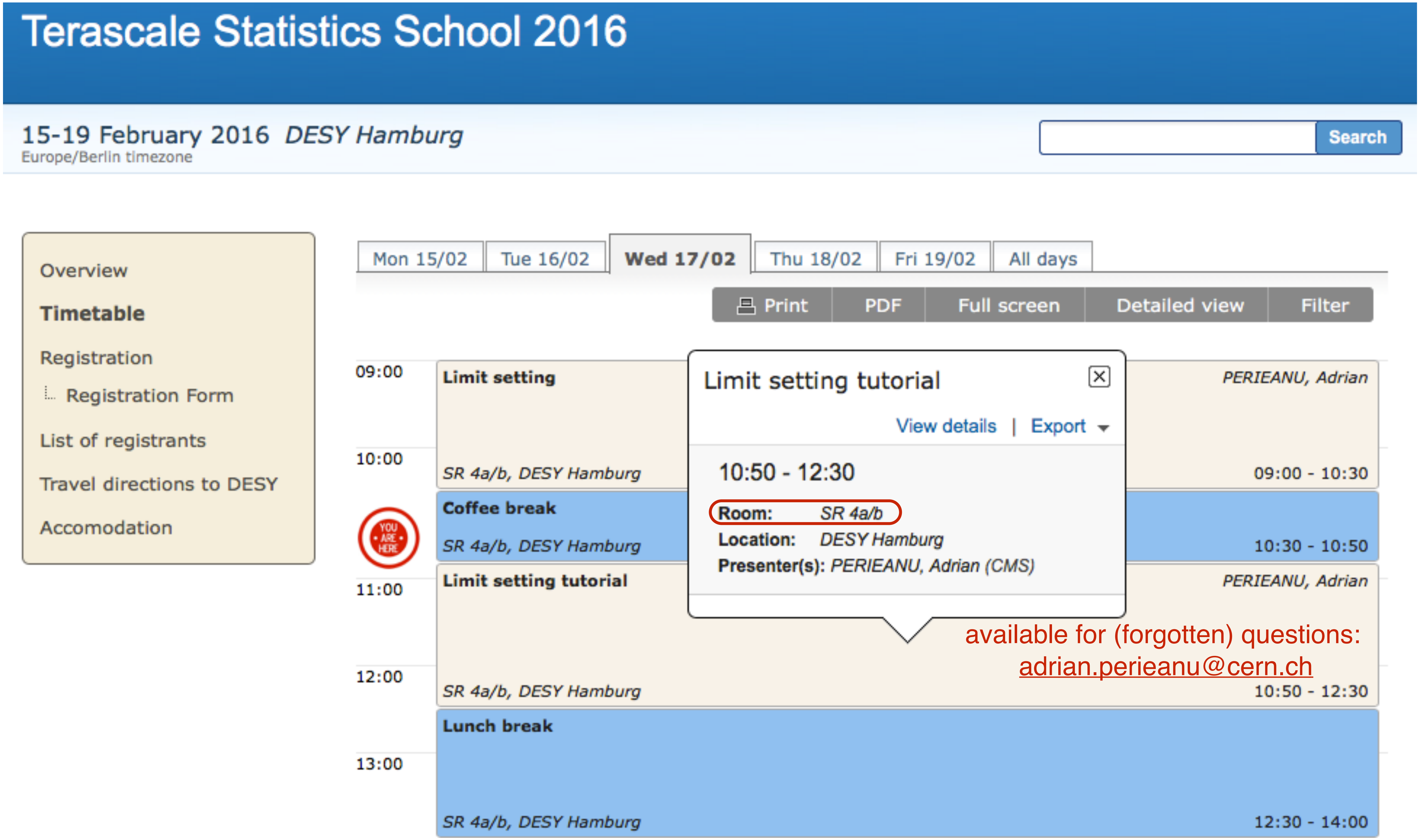
Calculation of Exclusion Limits

17th February 2016

Adrian Perieanu

where are you? still here?



available for (forgotten) questions:
adrian.perieanu@cern.ch

reminder
— part I —

* now you are sure that you can **read a limit plot**

* you know how a **confidence interval** is defined (single- and double-sided)

* you learned about the **confidence belt**

* and you know the major difference between the **frequentist (give for each theory the probability to be observed)** and **bayesian (assigns probabilities to theories)** approaches

*** next to come in part II:**
— formulate null-hypothesis
— how to calculate the observed (expected) limit with and w/o systematic uncertainties
— estimate and interpret the p-value
— explain limits to “outside” wild world

hypothesis

*** goal: quantify the agreement between theory model and measured data**

*** what we need before we start quantifying:**
— to define a hypothesis (the model)
— setup a list of parameters of the model to be determined
— measure the wanted parameters
— test hypothesis with measurements

*** methods:**
— statistical hypothesis test
— χ^2 test
— student's t-test
— Kolmogorov-Smirnov test
— Mann–Whitney U test

null-hypothesis

* proposes a general/default position
* can be tested against of an alternative hypothesis
* should be defined with care
* definitely should be defined before making the experiment/analysis

* if data rejects the null-hypothesis than the opposite is true

binomial null-hypothesis test

* take a 2 EURO Belgian coin and toss it three times
* let's say it came head down each time

* null-hypothesis: the coin is fair
one in a million coins is phony

* alternative hypothesis: the coin is phony
one in twenty coins is phony

* estimate the probabilities using bayesian statistics approach

* how often does the coin have to came head down to be phony
with a 95% CL?

binomial null-hypothesis test

* take a 2 EURO Belgian coin and toss it three times
* let's say it came head down each time

* null-hypothesis: the coin is fair
one in a million coins is phony: 10^{-6}

* alternative hypothesis: the coin is phony
one in twenty coins is phony: 5%

use Bayesian theorem

* conditional probability $P(A|B)$: probability of A to occur under the condition that B has occurred

likelihood prior probability

posterior probability $$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

evidence

binomial null-hypothesis test

* take a 2 EURO Belgian coin and toss it three times
* let's say it came head down each time

$$P(\text{phony}|\text{3 heads}) \;=\; \frac{P(\text{3 heads}|\text{phony})}{P(\text{3 heads})} \cdot P(\text{phony})$$

$$P(\text{3 heads}) = P(\text{3 heads}|\text{fair}) \cdot P(\text{fair}) + P(\text{3 heads}|\text{phony}) \cdot P(\text{phony})$$

null-hypothesis: *P (phony/3 heads) = $(1{*}10^{-6})/[(1/2)^{3}{*}(1-10^{-6}) + 1{*}10^{-6}] = 8{*}10^{-6}$*

✗ ***alternative-hypothesis:*** *P (phony/3 heads) = $(1{*}0.05)/[(1/2)^{3}{*}(1-0.05) + 1{*}0.05] = 0.30$*

✗ ***alternative-hypothesis:*** *P (phony/4 heads) = $(1{*}0.05)/[(1/2)^{4}{*}(1-0.05) + 1{*}0.05] = 0.45$*

✗ *. . .*

✓ ***alternative-hypothesis:*** *P (phony/9 heads) = $(1{*}0.05)/[(1/2)^{9}{*}(1-0.05) + 1{*}0.05] = 0.96$*

* nine times in a row!!!

hypothesis: in particle physics

* **null-hypothesis:** we expect that the observed data follow the predictions of the Standard Model (SM)

* **from SM:** process cross-section

* **from experiment:** integrated luminosity and selection efficiency

* **background probability density function:**
from SM and experiment

* **test hypothesis:** with the test statistics - the event yield of your super sensitive selection

hypothesis: signal, bkg. & data

* toy example:
— let's assume we observe after our analysis 7 events in **data**: **d = 7**
— SM predicts a **background** of 4 events: **b = 4**
— the theory of our best friend **predicts** a **signal** yield of 11 events: **s = 11**

* **null-hypothesis: background only** $\lambda_b = b = 4$

* **alternative hypothesis: signal & background** $\lambda_{s+b} = s + b = 15$

hypothesis: signal, bkg. & data

* **null-hypothesis: background only** $\lambda_b = b = 4$

* **alternative hypothesis: signal & background** $\lambda_{s+b} = s + b = 15$

* probability (**p-value**) to reject **null-hypothesis H_0 while H_0 is true:**

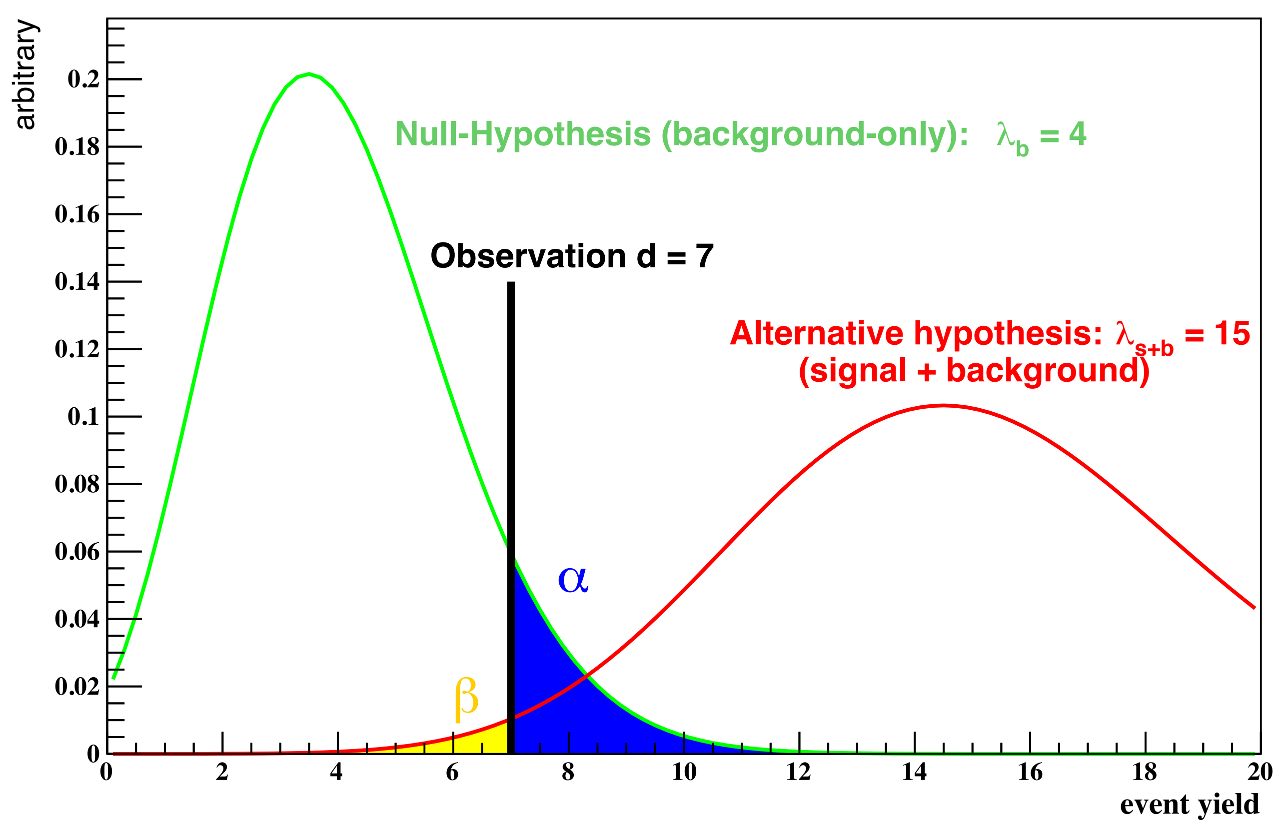
$$\int_d^{\infty} f(x|H_0)dx = \alpha < 1 - CL_{critical}$$

* **probability to reject the alternative hypothesis H_a while H_a is true:**

$$\int_{-\infty}^{d} f(x|H_a)dx = \beta < 1 - CL_{critical}$$

* **as usual choice: $CL_{critical} = 95\%$**

* **estimate for our example:** α = ?, β = ?%



hypothesis: signal, bkg. & data

* **null-hypothesis: background only** $\lambda_b = b = 4$

* **alternative hypothesis: signal & background** $\lambda_{s+b} = s + b = 15$

* probability (**p-value**) to reject **null-hypothesis H_0 while H_0 is true:**

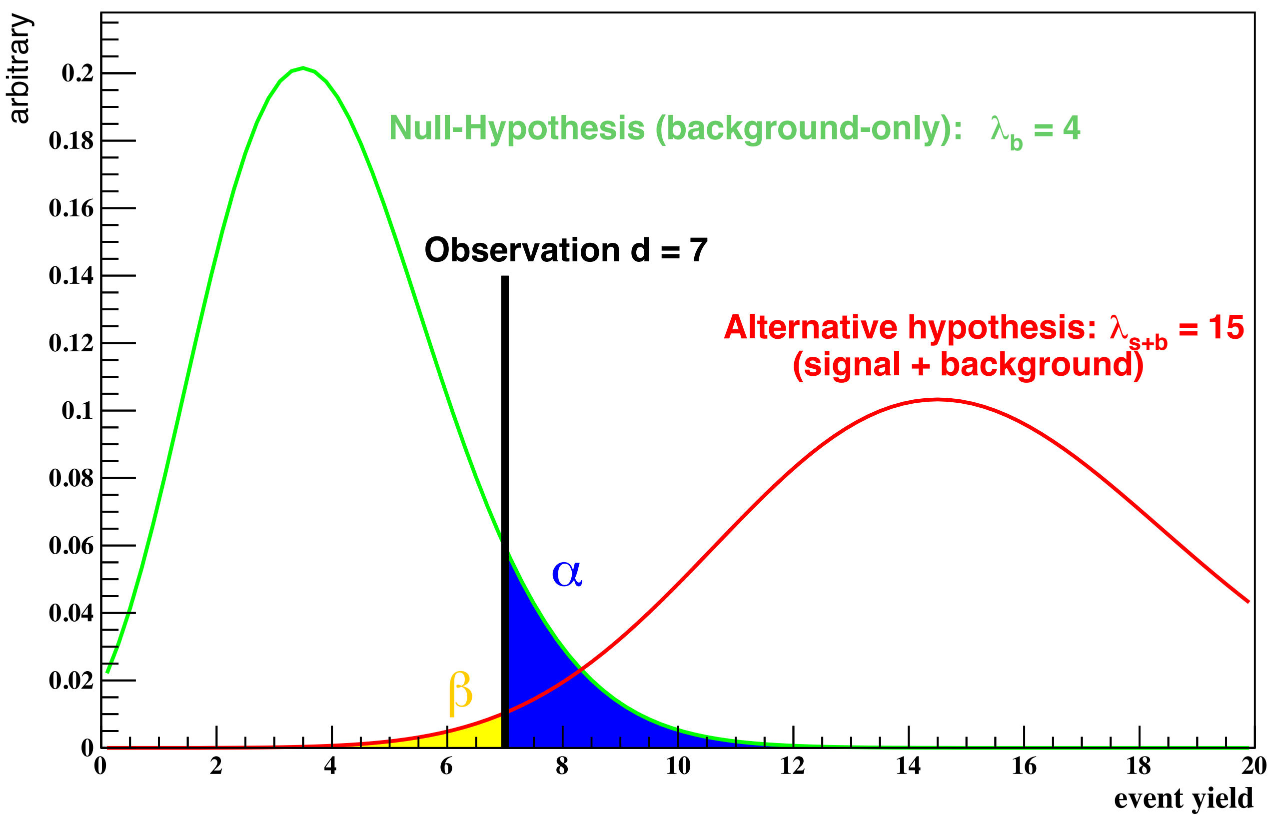
$$\int_d^{\infty} f(x|H_0)dx = \alpha < 1 - CL_{critical}$$

* **probability to reject the alternative hypothesis H_a while H_a is true:**

$$\int_{-\infty}^{d} f(x|H_a)dx = \beta < 1 - CL_{critical}$$

* **as usual choice: $CL_{critical} = 95\%$**

* **in our example: $\alpha = 11\%$, $\beta = 1.8\%$**



hypothesis: type I and II erros

* let's regard two mutually exclusive hypotheses that are either true or false

* possible outcomes:

— **Accepting a true hypothesis**
— **Rejecting a wrong hypothesis**

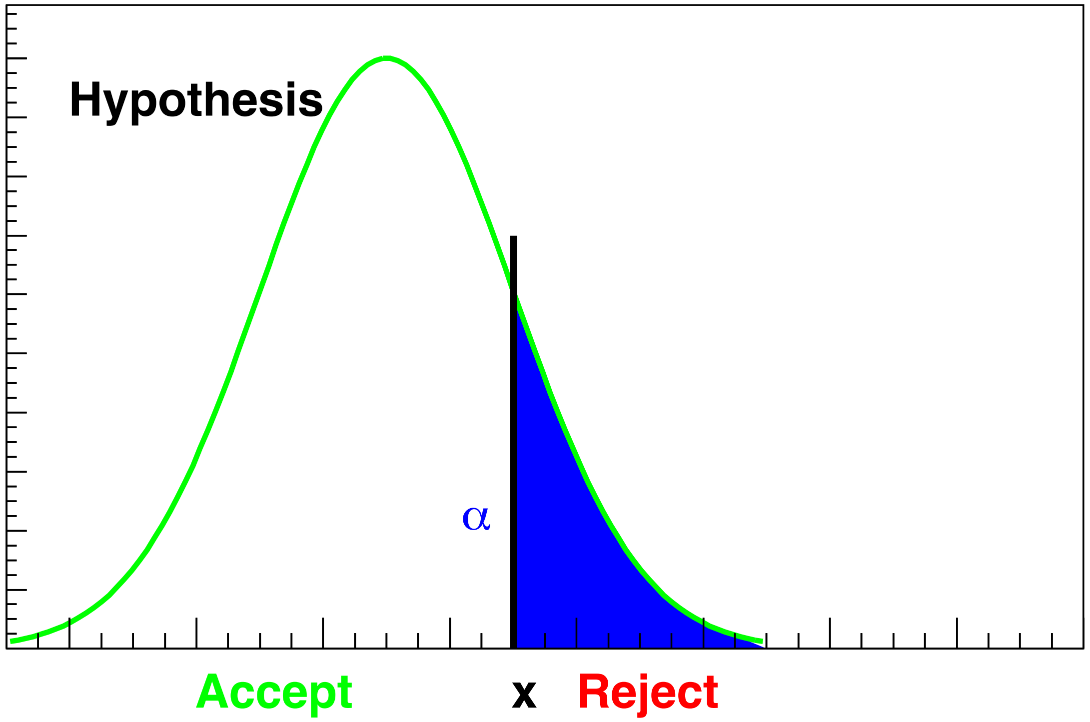
— **Rejecting a true hypothesis (type I error)**
— **Accepting a wrong hypothesis (type II error)**

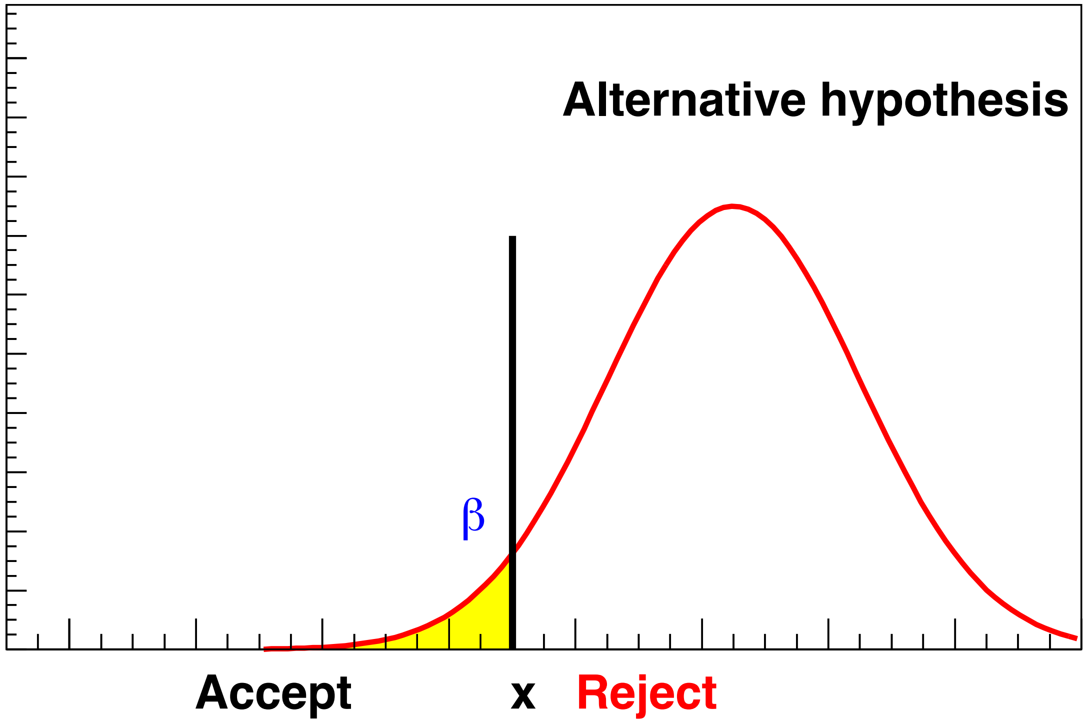
* for α, test significance, type I errors occur for:

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

* probability to accept the wrong hypothesis H_a is β, type II errors occur for:

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





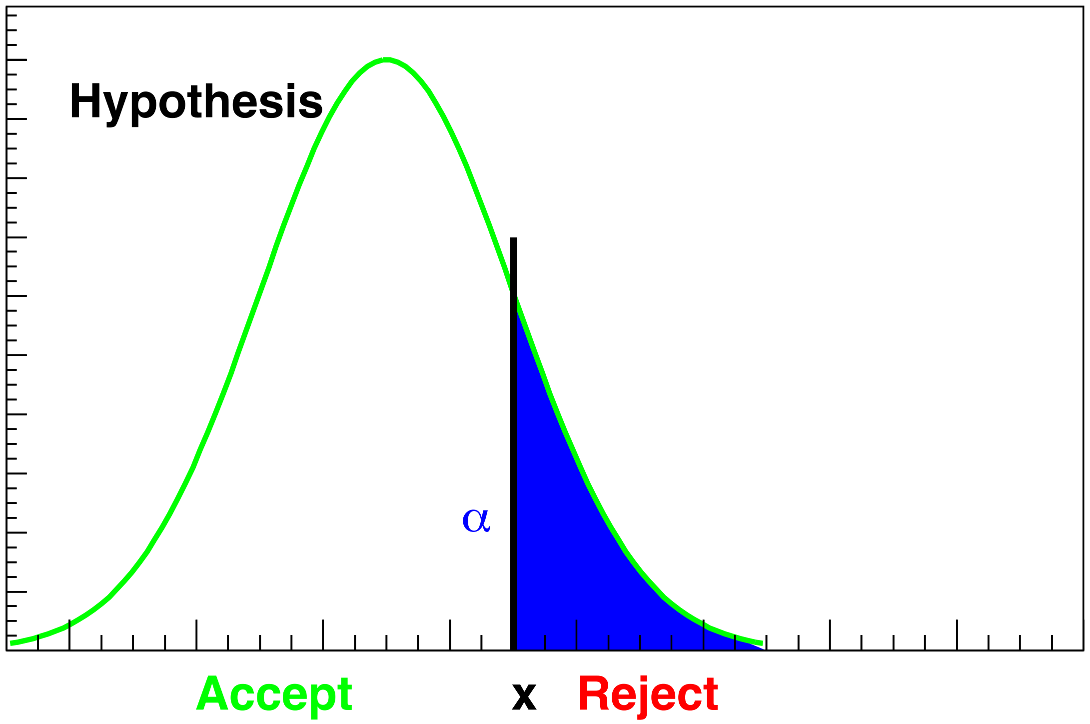
hypothesis: type I and II erros

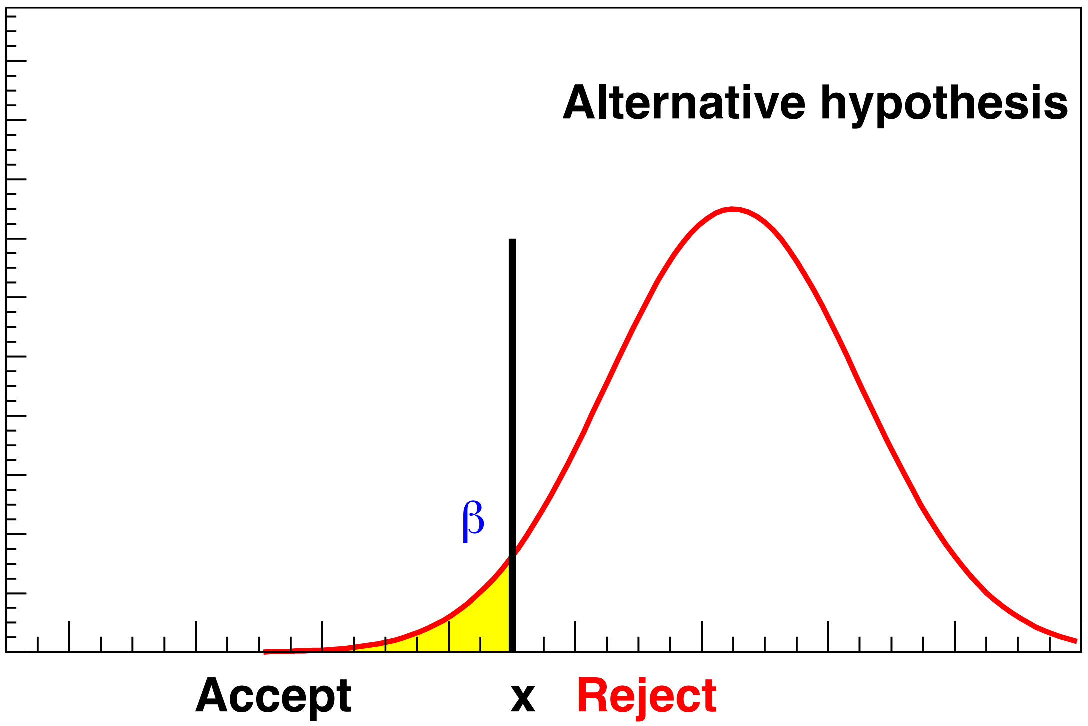
* let's regard two mutually exclusive hypotheses that are either true or false

* possible outcomes:

— Accepting a true hypothesis
— Rejecting a wrong hypothesis

— Rejecting a true hypothesis (type I error)
— Accepting a wrong hypothesis (type II error)





* α and β should be as small as possible
* a tradeoff between minimising α and β
* importance of α or β depends on topic

in practice:
— b-tagging: b-jet and light-jet hypothesis
— tau reconstruction: real and fake candidates

hypothesis: life decisions

* let's assume that a "bird flu" symptom is always fever with 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 in Europe (so far)

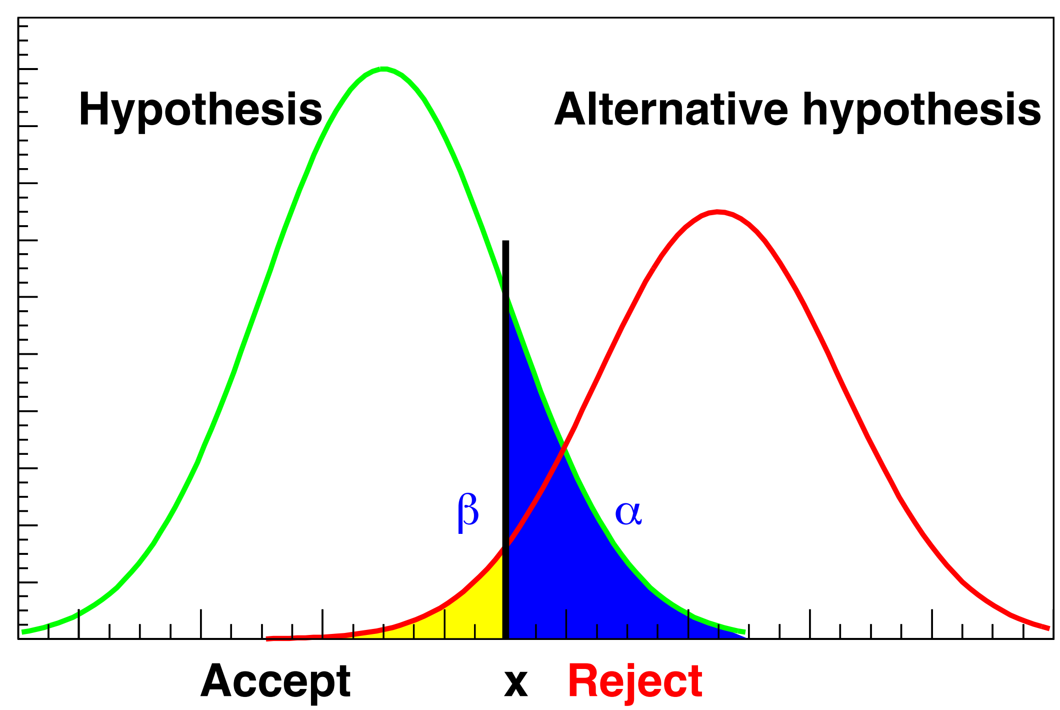
— where to decide (fever-threshold for treating a patient ambulant or stationary) if we want a **test-power of $1 - \beta = 90\%$**?

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

— how many patients with "bird flu" are rejected in the 2 cases?

— estimate how many of the patients will have normal flu

usually when I have flu I do not go to hospital… higher chances to get something more serious, but I am curios what your tests are saying



hypothesis: life decisions

* let's assume that a "bird flu" symptom is always fever with 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 in Europe (so far)

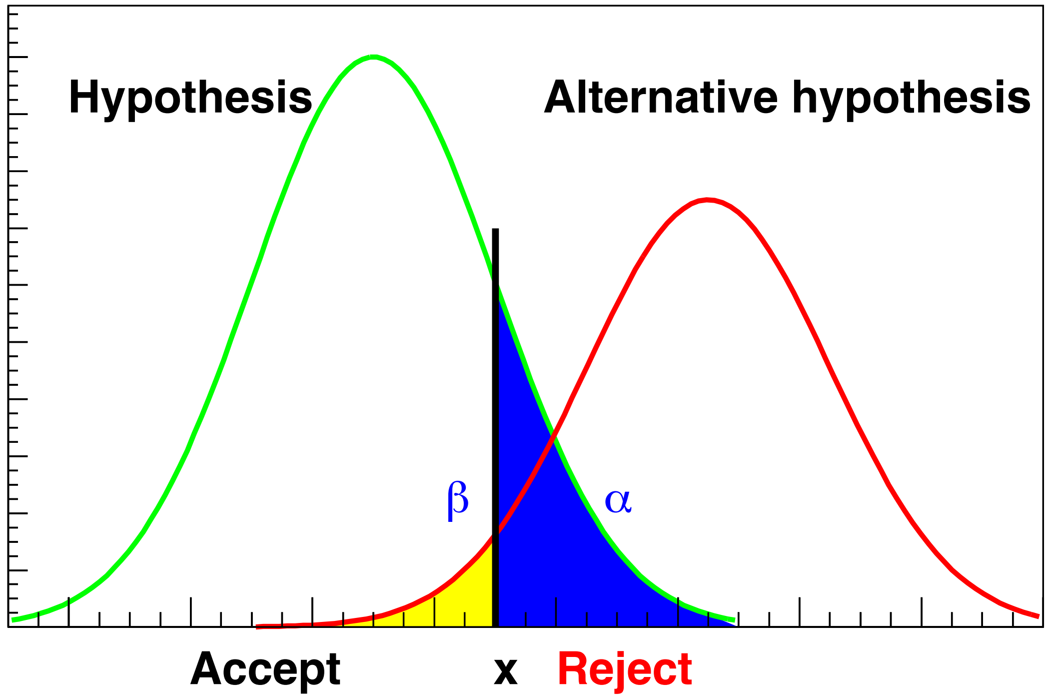
— where to decide (fever-threshold for treating a patient ambulant or stationary) if we want a **test-power of 1 – β = 90%**?

— where to decide (fever-threshold for treating a patient ambulant or stationary) if we want a **test-significance of α = 5%**?

— accepting β = 10% of normal flu patients, leads to the rejection of ~15% bird flu fever patients (Type I error)

> 92% beds are occupied by normal flu patients

usually when I have flu I do not go to hospital… higher chances to get something more serious, but I am curios what your tests are saying



hypothesis: life decisions

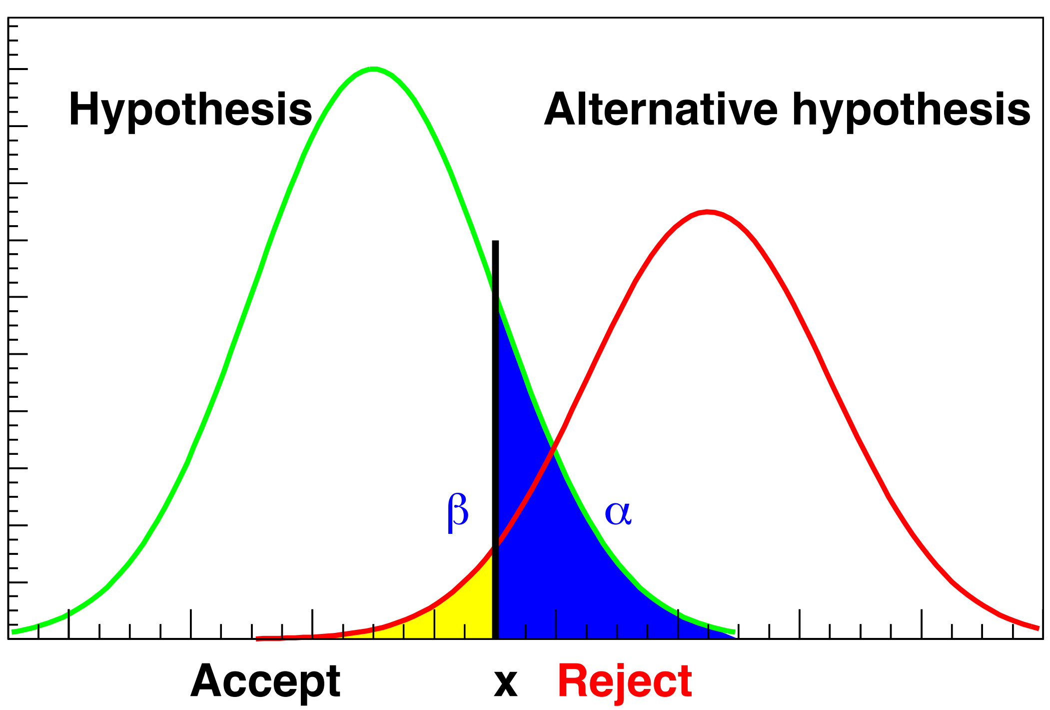
* let's assume that a "bird flu" symptom is always fever with 39.7 °C with a Gaussian spread of 0.2 °C
* patients with normal flu only have temperature 39.2 ± 0.2 °C and are 100 times more likely in Europe (so far)

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

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

— accepting 95% of bird flu fever patients (significance $\alpha = 5\%$) by cutting at 39.37 C leads to a Type II error (accepted normal flu patients) of $\beta \approx 80\%$
> 98.8% beds are occupied by normal flu patients

usually when I have flu I do not go to hospital… higher chances to get something more serious, but I am curios what your tests are saying



hypothesis: “best” test

*** Neyman-Pearson lemma:**
— when performing a hypothesis test between two hypotheses H_0 and H_a, then the likelihood-ratio test which rejects H_a in favour of H_0 when

$$\frac{L_{H0}(x)}{L_{Ha}(x)} \geq Q_0$$

for a given significance α

is the most powerful test-statistic to minimise both α and β

* H_0 and H_a *hypothesis* have to be:
— explicitly defined
— simple

* acceptance region giving the highest power 1 – β for a given significance α is the region comprised by the above equation

* in the one-dimensional case, a cut on *x* for a specific α, e.g., b-tag efficiency, determines β, therefore the purity

what is a likelihood?

* for a binomial distribution:

$$P(x|n,p) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

* probability question:
"if an event has probability p = 0.6, and we have n = 10 trials, what is the probability of the event occurring x = 3 times"?

* statistical question:
"in n = 10 trials I observed the event occur x = 3 times, so what is a good estimator of the success probability p?"

— same function, different point of view: now the data (x) are fixed, and we view the expression in terms of the parameter (p), and use it to obtain an estimator of the parameter (success probability p)

* likelihood function for binomial data as:

$$L(p|n,x) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

— same expression now viewed as a function of p instead of x

is time for limits

limits
— finally —

remember

*** a point beyond which it is not possible to go**

*** an amount or number that is the highest or lowest allowed**

* **observed limit** on the signal event yield at CL_{s+b} = 95% is defined as the *s*, for which:

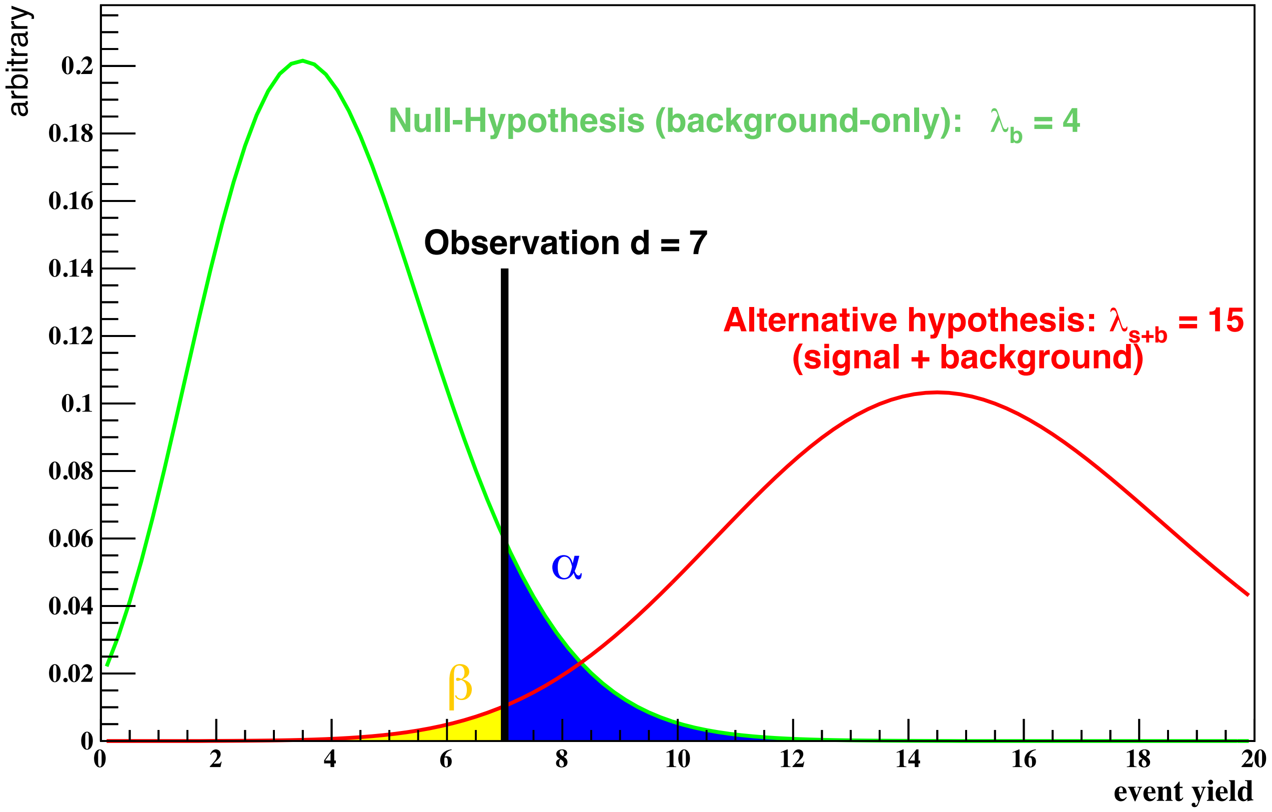
$$\int_{-\infty}^{d} Q(x|H_a)dx = \beta \leq 1 - CL_{s+b}$$

limits
— hands on —

* **observed limit** on the signal event yield at CL_{s+b} = 95% is defined as the *s*, for which:

$$\int_{-\infty}^{d} Q(x\,|H_a)\,dx = \beta \leq 1 - CL_{s+b}$$

* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**



* ***Q(x|Ha)*: Poisson probability density function with a mean λ = *s* + *b***

* **calculate the limit on s + b at 95% CL**

limits
— hands on —

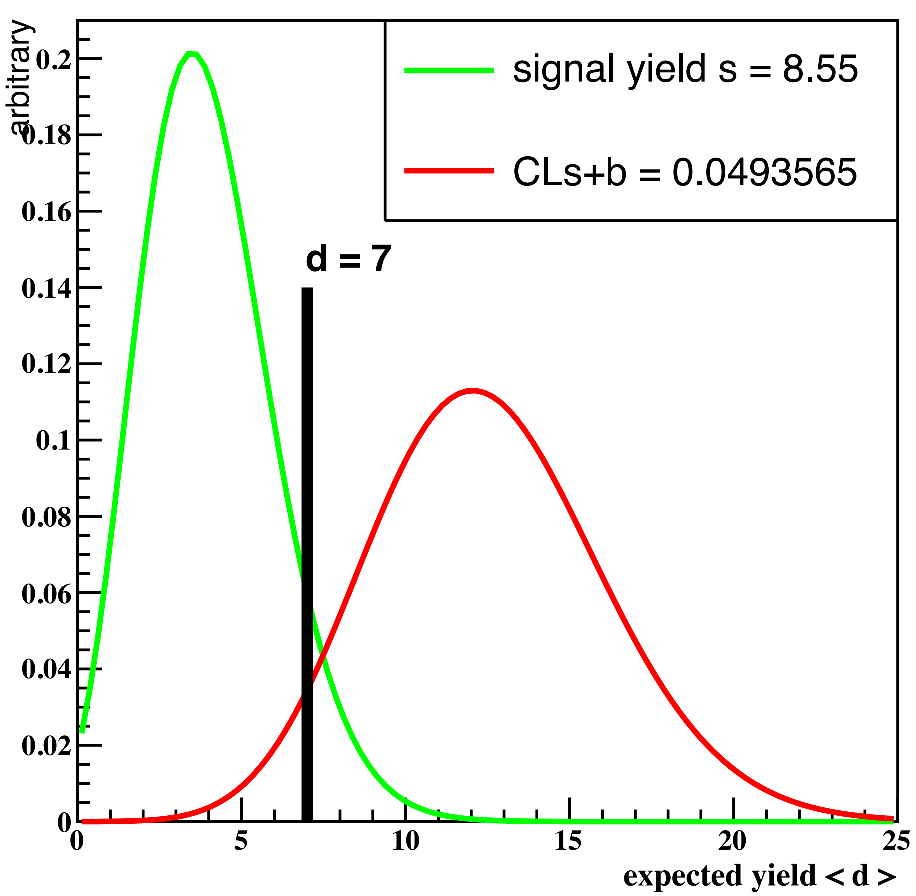
* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**
— **limit on s + b at 95% CL**: $s + b = ?$

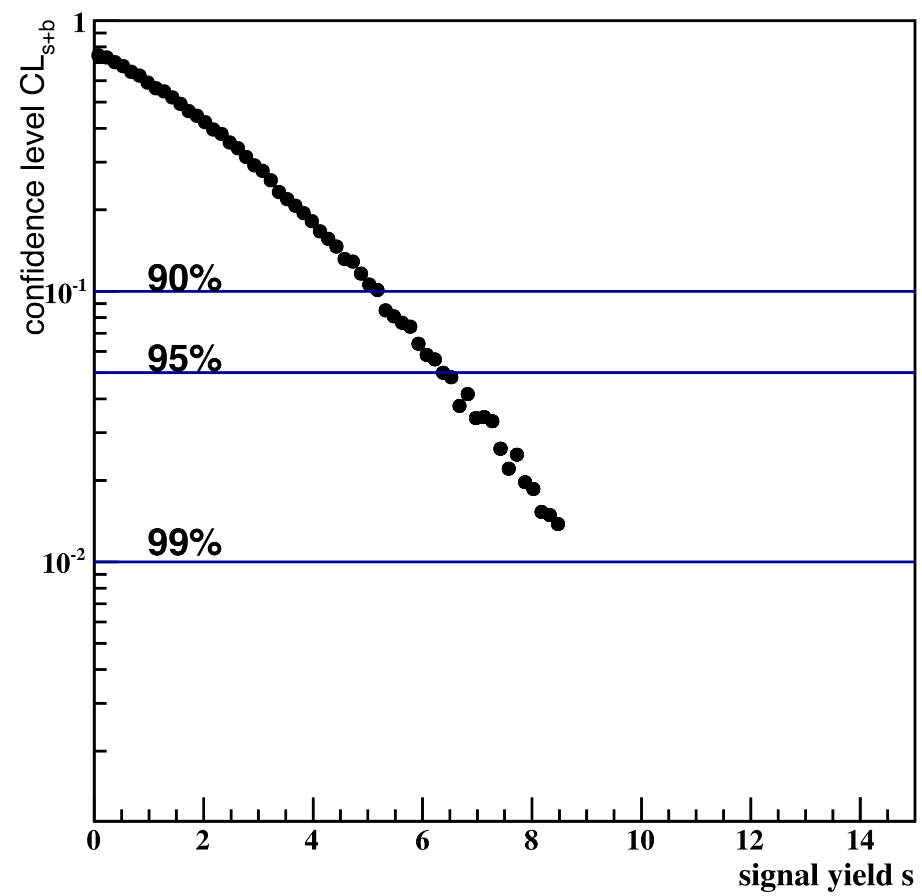
$$\beta = \sum_{d=0}^{d_{\text{obs}}} \frac{e^{-(s+b)}(s+b)^d}{d!}$$

limits
— hands on —

* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**
— **limit on s + b at 95% CL**: $s + b = 12.5$

$$\beta = \sum_{d=0}^{d_{obs}} \frac{e^{-(s+b)}(s+b)^d}{d!}$$





limits

— few extra points —

* test-statistic $Q(x|H_a)$ was in the example a Poisson probability density function modelling the under H_a hypothesis expected statistical uncertainties of the measurement (the data)

* test-statistics $Q(x|H_a)$ may incorporate also systematical uncertainties on the background σ_b and on the signal estimation σ_s, e.g. ,

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

* $Q(x|H_0)$ and $Q(x|H_a)$ may be defined by a likelihood that distinguishes both hypotheses

* agreement of the measured data with the background-only expectation, i.e. the null-hypothesis H_0, is not directly considered

* limit on the observed signal event yield does not depend on the expected signal event yield

more than one channel

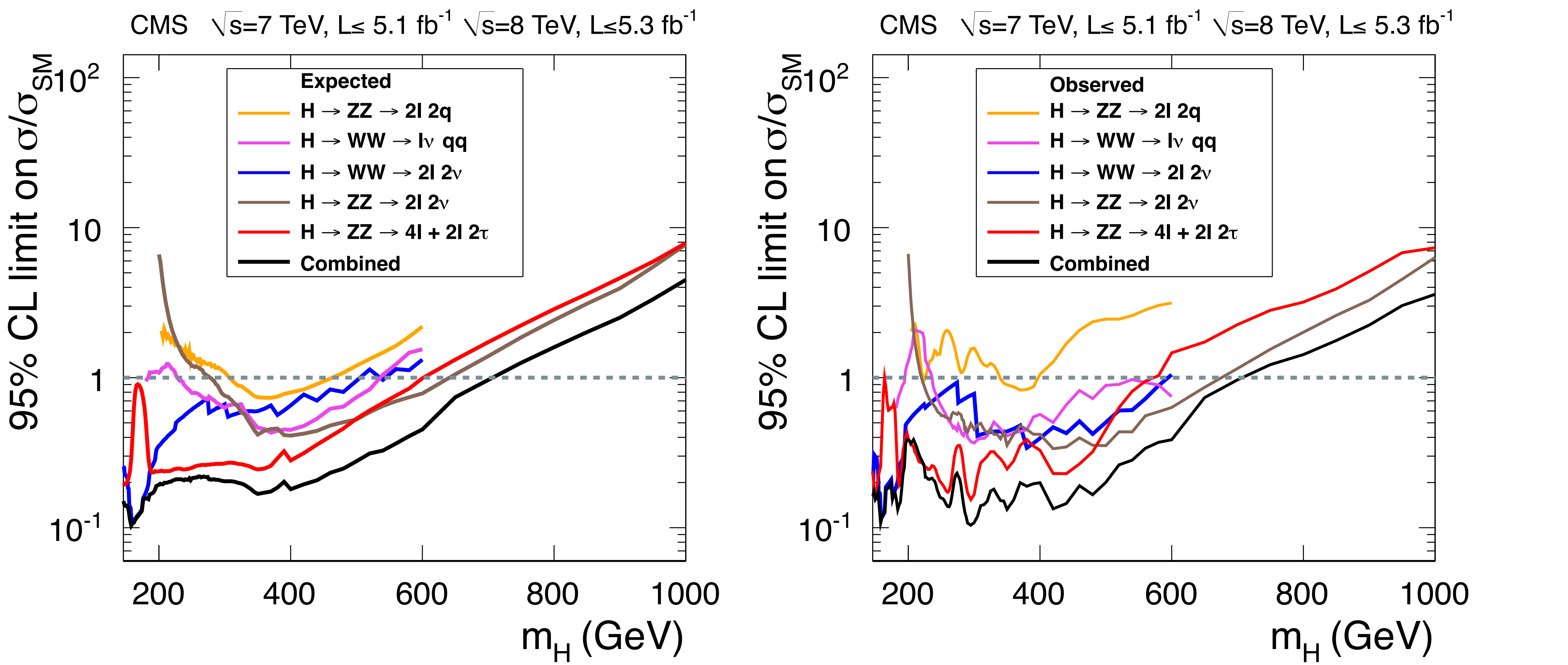
* because likelihood functions are multiplicative, multiple statistically exclusive channels, i.e. from different exclusive selections or histogram bins, can be combined:

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

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 needs to be considered

more than one channel



Eur. Phys. J. C 73 (2013) 2469
arXiv:1304.0213

* one has to make sure that:
— different exclusive selections are used for the considered channels
— correlations of the systematic uncertainties need to be taken into account

expected limits and uncertainties

* expected limits are defined as the 50% quantile, i.e. the median of the distribution of observed limits for a number of pseudo-experiments

* pseudo-observations are drawn according to the background-only null-hypothesis test-statistic H_0.

* the ±1σ uncertainties on the expected limit are equivalent to 16% and 84% quantiles

0 quartile = 0 quantile = 0 percentile
1 quartile = 0.25 quantile = 25 percentile
2 quartile = 0.5 quantile = 50 percentile (median)
3 quartile = 0.75 quantile = 75 percentile
4 quartile = 1 quantile = 100 percentile

CL_s: modified frequentist procedure

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

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

* null-hypothesis is that there is no signal

* alternate hypothesis that signal exists

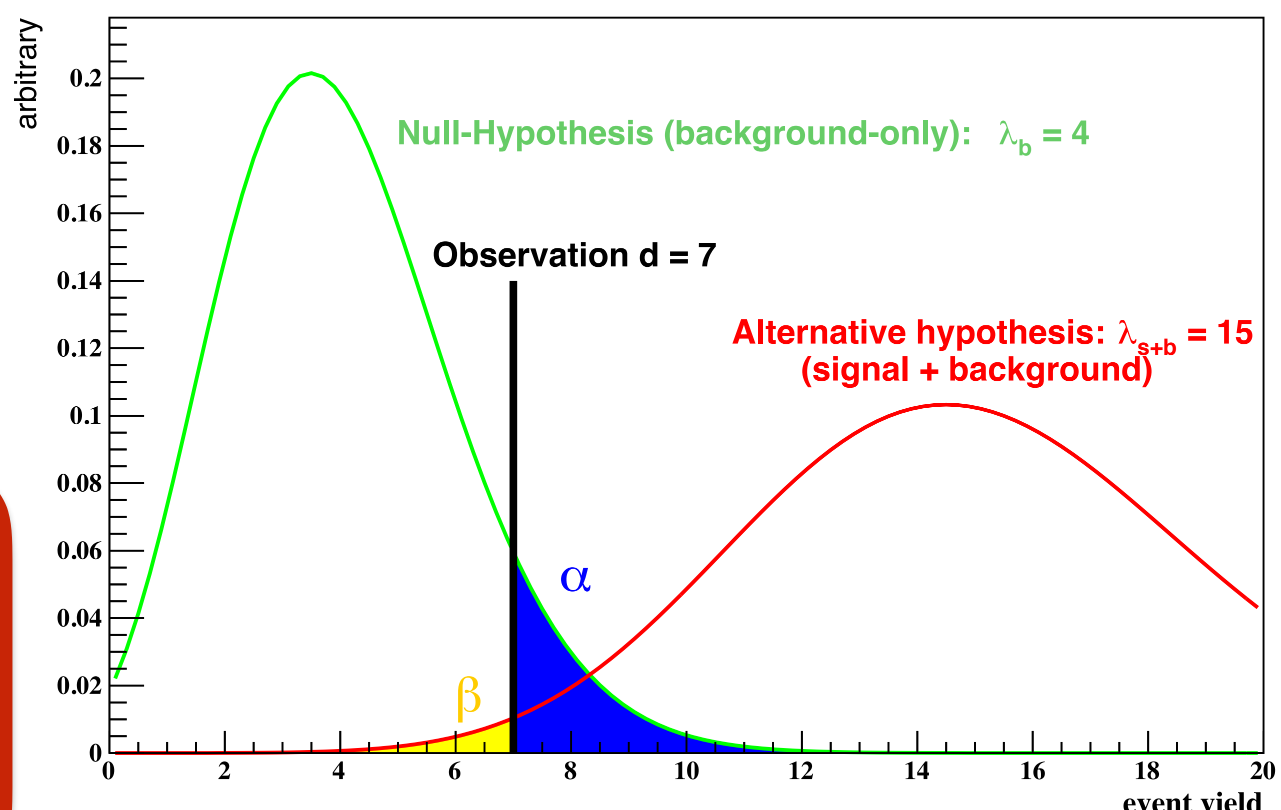
* CL_s gives an approximation to the confidence in the signal hypothesis one might have obtained if the experiment had been performed in the complete absence of background

* CL_s tries to reduce the dependency on the uncertainty due to the background

CL_s: single counting channel/experiment
— hands on —

* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**

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



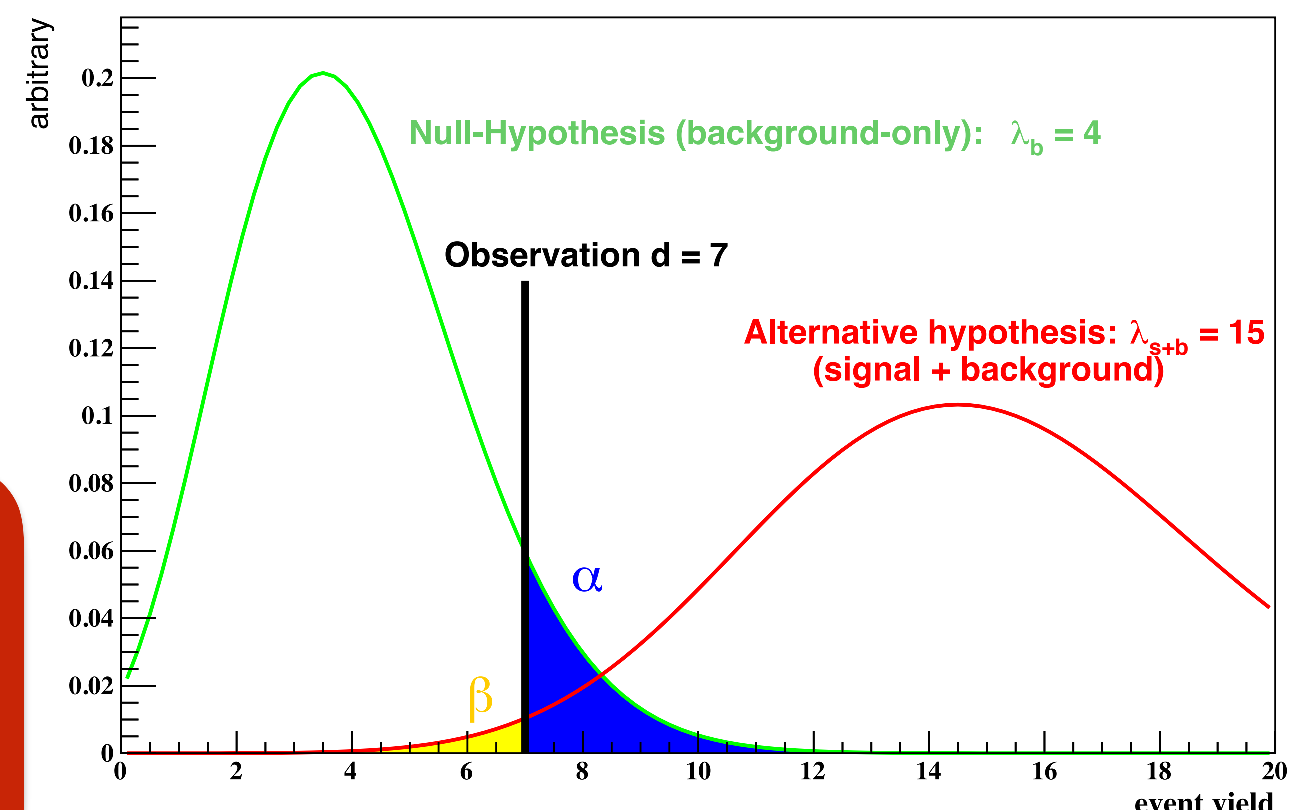
*** calculate the limit on s + b at 95% CL with CL_s**

CL_s: single counting channel/experiment
— hands on —

* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**

$$CL_{s+b} = \sum_{d=0}^{d_{obs}} \frac{e^{-(s+b)}(s+b)^d}{d!}$$

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



*** calculate the limit on s + b at 95% CL with CL_s**

CL_s: single counting channel/experiment
— hands on —

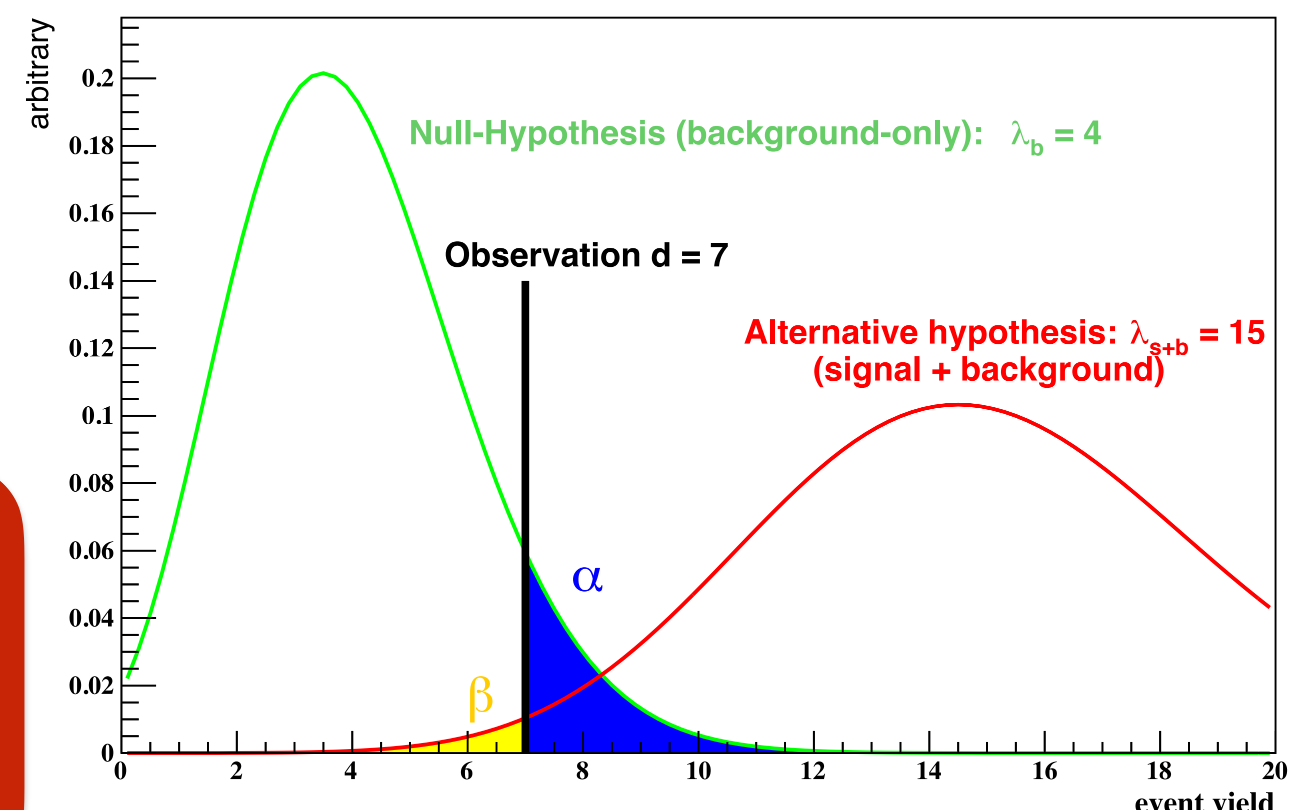
* for our toy example:
— events in **data**: **d = 7**
— **background** events: **b = 4**

$$CL_{s+b} = \sum_{d=0}^{d_{obs}} \frac{e^{-(s+b)}(s+b)^d}{d!}$$

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

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

*** limit on s + b at 95% CL: *s + b = 12.5***



CL_s: more than one channel

* with a likelihood-ratio as test-statistics compute CL_{s+b} and CL_b:

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

* expected signal *s* depends on a model parameter, e.g., the Higgs mass m_H

* likelihoods are multiplicative, different exclusive *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

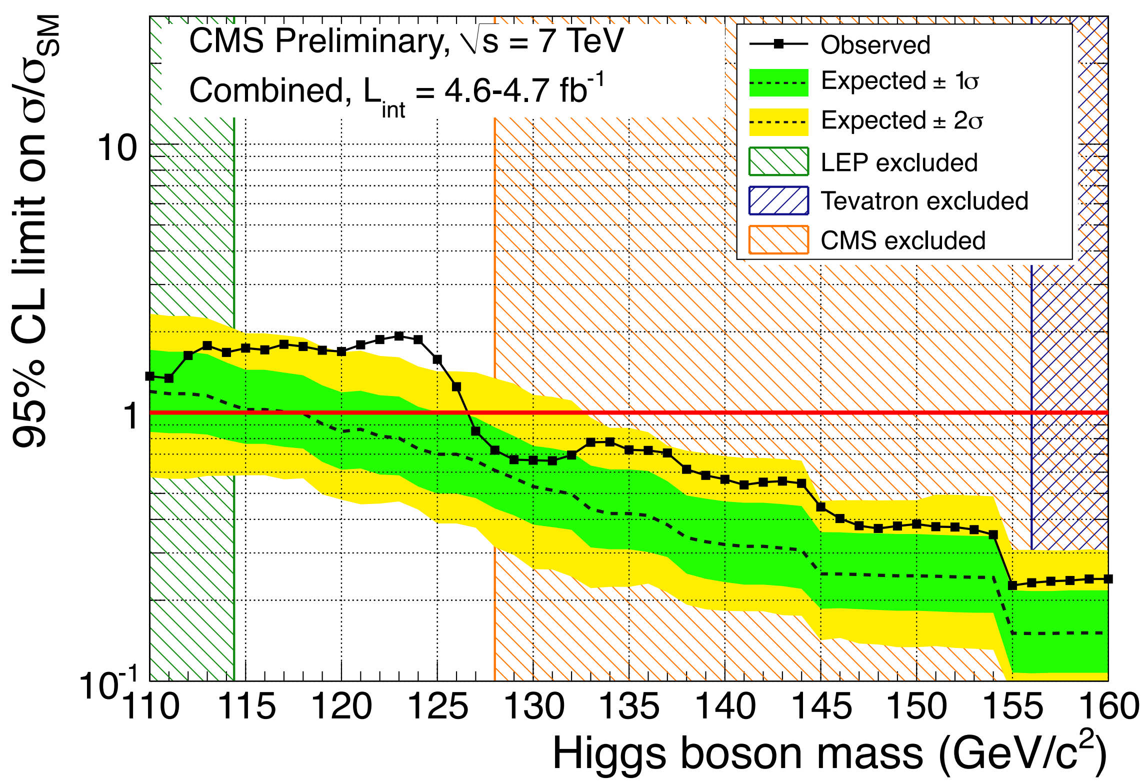
* CL_{s+b} is then given by:

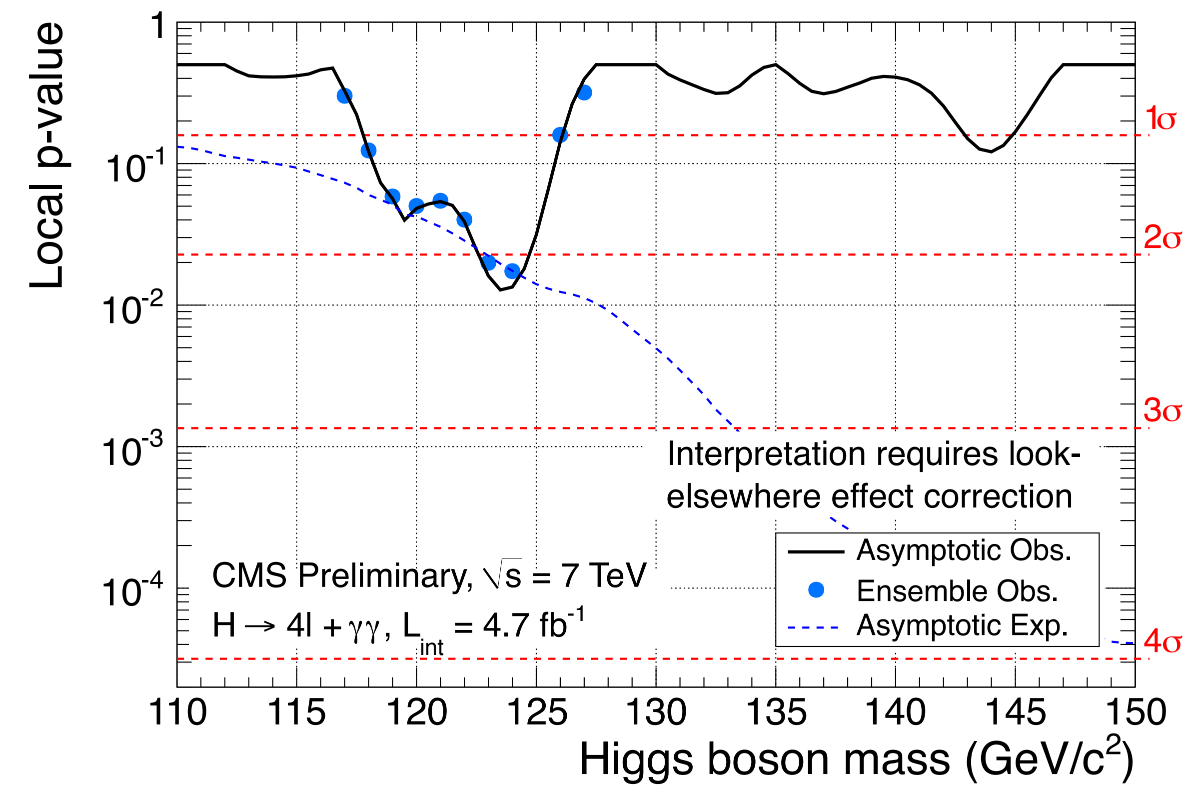
$$\begin{aligned} CL_{s+b} &= P_{s+b}(X \leq X_{obs}) \\ &= \int_{-\infty}^{X_{obs}} \frac{dX_{s+b}}{dx} dx \end{aligned}$$

where dX_{s+b}/dx is the probability density function distribution of the test-statistics X for signal+background experiments.

p-Value

from “former times”





*** once in a while it happens that you see something coming out of the two sigmas band: what do you do?**

— calculate p-Value

— p-Value helps you determine the significance of your results

p-Value in normal life

* p-Value ≤ 0.05: strong evidence against the null-hypothesis, so you reject the null hypothesis

* p-Value > 0.05: weak evidence against the null hypothesis, so you fail to reject the null hypothesis

— always report the p-value

* **classical example: pizza delivery!!!**

— to attract clients a pizzeria promote: **delivery times in less than 30'**

* you, the hungry particle physicist:

— **null-hypothesis H_0: mean delivery time is 30 minutes max**

— **alternative-hypothesis H_a: mean delivery time is greater than 30'**

* you randomly sample some delivery times and, after you eat, run the data through the hypothesis test: **$p = 0.001$**

— **how shall we interpret this?**

p-Value in normal life

* **classical example: pizza delivery!!!**

— to attract clients a pizzeria promote: **delivery times in less than 30'**

* you, the hungry particle physicist:

— null-hypothesis H_0: mean delivery time is 30 minutes max

— alternative-hypothesis H_a: mean delivery time is greater than 30'

* you randomly sample some delivery times and, after you eat, run the data through the hypothesis test**: p = 0.001**

— how shall we interpret this?

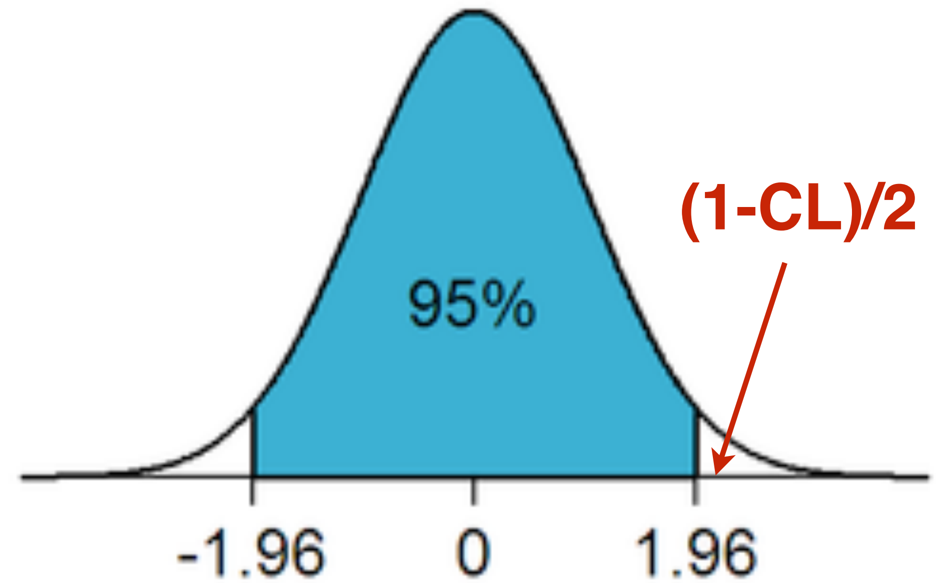
* there is a probability of 0.001 (of 0.1%) that you will wrongly reject the pizza place's claim that their delivery time is 30' max

ahhhaaa… so they are wrong, but did the right thing… during your sampling you became their client… good to know

p-Value ≤ 0.05
— how many sigmas? —

* estimate how many sigmas correspond to for p-Value ≤ 0.05

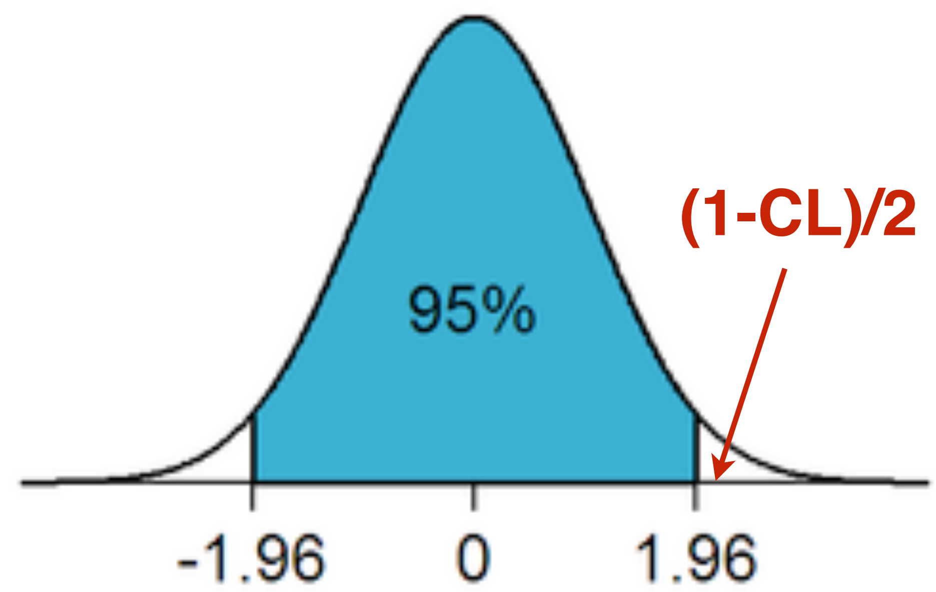
* write a ROOT macro and estimate the p-Value for 1σ, 2σ, 3σ, 4σ and 5σ



p-Value ≤ 0.05
— how many sigmas? —

* **estimate how many sigmas correspond to for p-Value ≤ 0.05**

* **write a ROOT macro and estimate the p-Value for 1σ,** 2σ, 3σ, 4σ and 5σ



Sigma	**1σ**	1.28	1.64	1.96	**2σ**	2.58	**3σ**
CI %	68.3%	**80%**	**90%**	**95%**	95.45%	**99%**	99.73%
P-value	0.317	0.20	0.10	0.05	0.0455	0.01	0.0027

α = (1-CL)

in normal life what is above 1.96 **σ** *is "significant"*

* **in particular physics we claim a discovery when we see "a signal with five standard deviations"**
— for 5 σ the corresponding **p-Value = 5.7 x 10^{-7}**

http://pdg.lbl.gov/2013/reviews/rpp2013-rev-statistics.pdf

questions

*
*
*

to take home with you
— instead of summary —

what about a little homework?

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

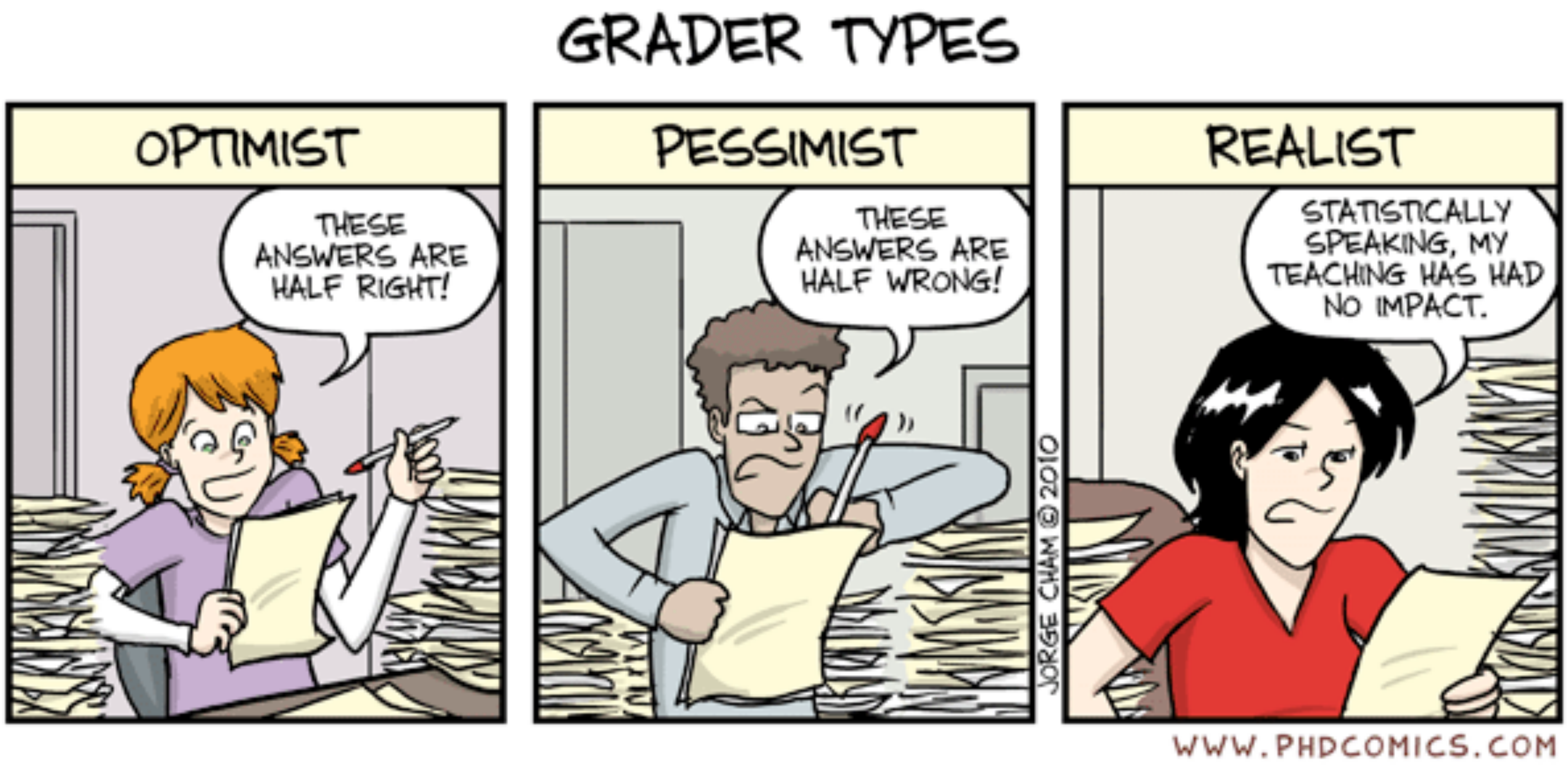
$$CL = \frac{\int_{\mu_{lim}}^{\infty} dx' \frac{1}{2\pi} e^{\frac{-(x'+2)^2}{2}} \theta(x')}{\int_{0}^{\infty} dx' \frac{1}{2\pi} e^{\frac{-(x'+2)^2}{2}}} = 10\%.$$

  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.

and a useful link in case you have more time

www.desy.de/~sschmitt/LimitStatSchool2013

before the end…



backup