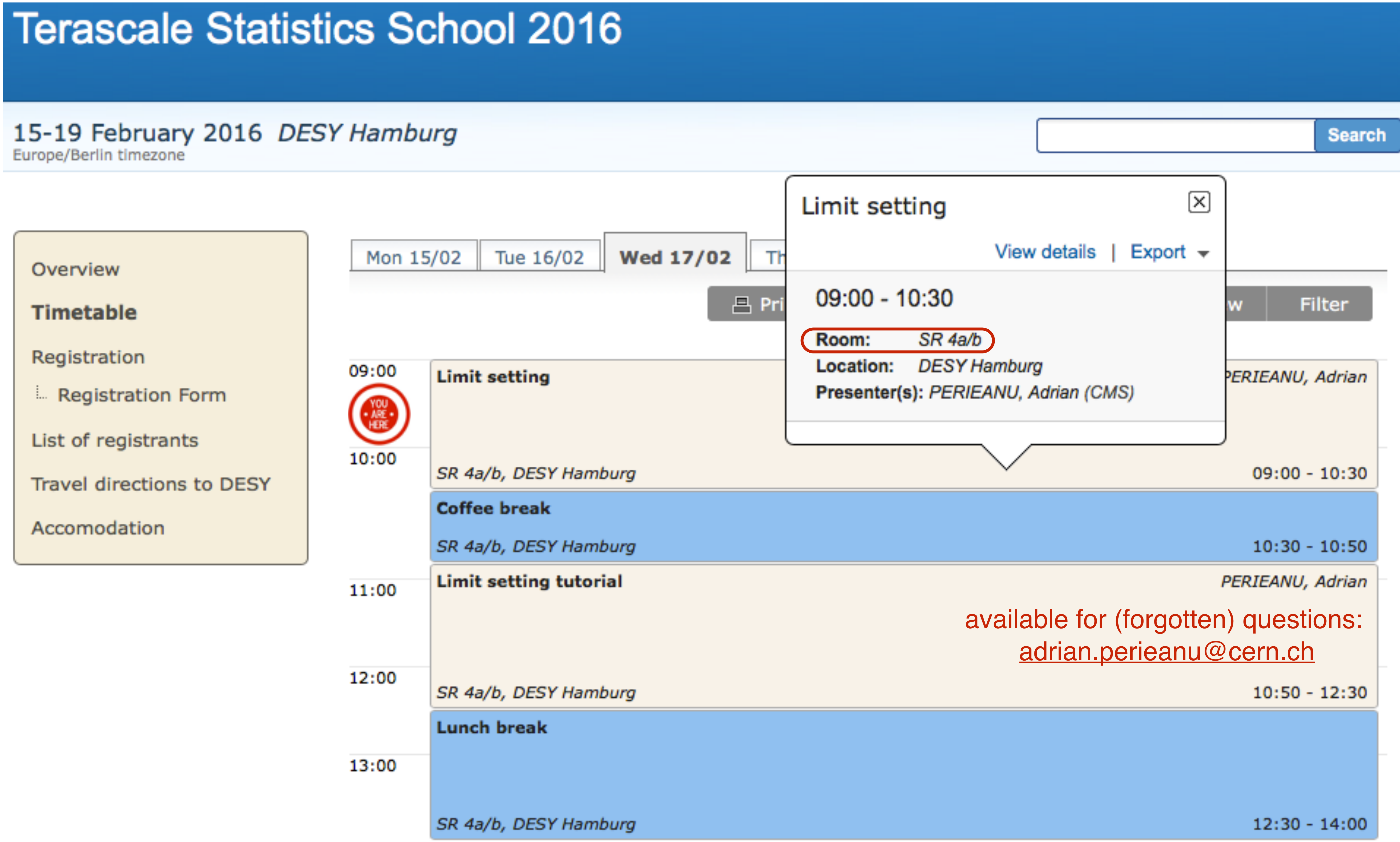
Calculation of Exclusion Limits

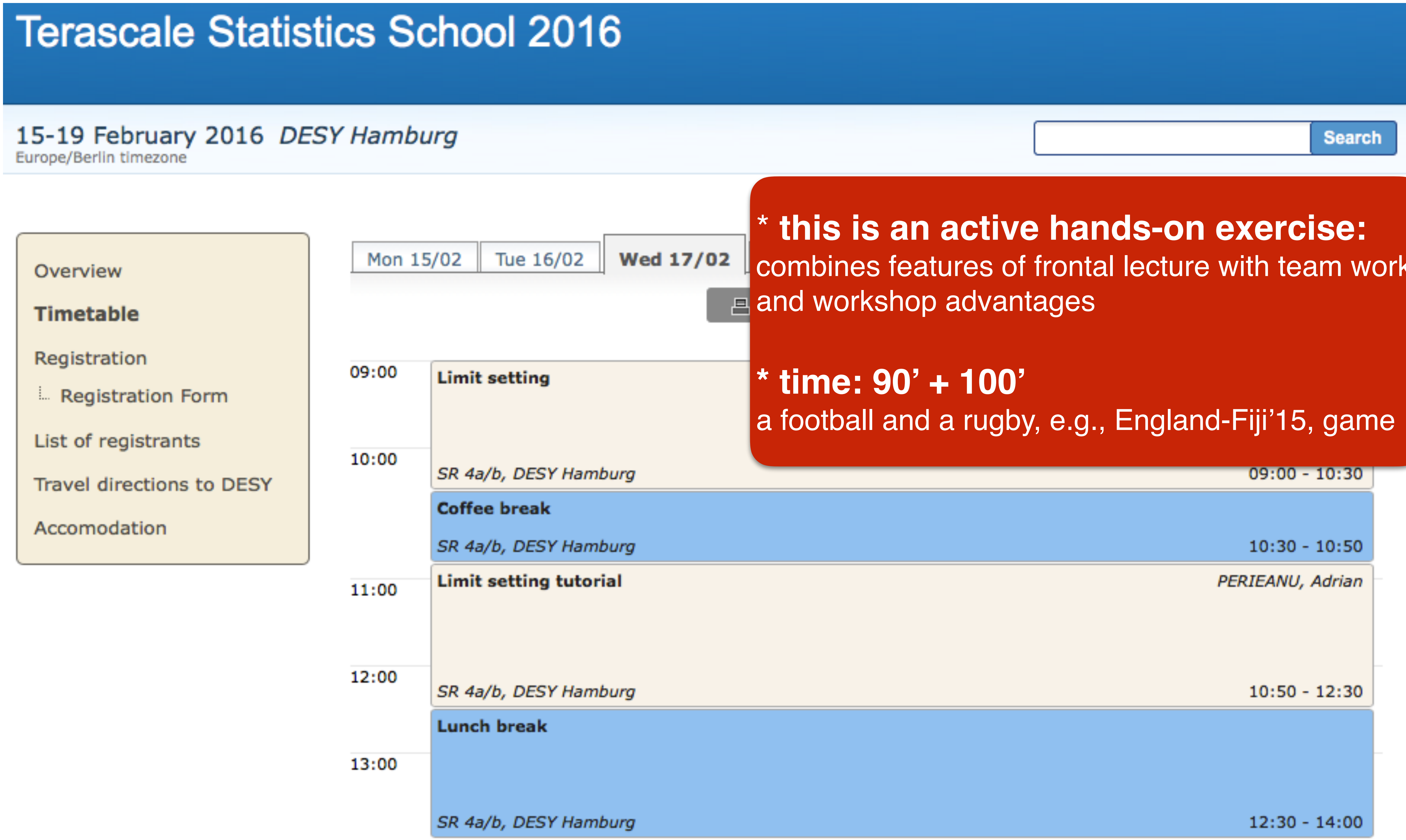
17th February 2016

Adrian Perieanu

where are you?



exercise structure



* **this is an active hands-on exercise:**
combines features of frontal lecture with team work and workshop advantages

* **time: 90’ + 100’**
a football and a rugby, e.g., England-Fiji’15, game

before we start

this lecture and tutorial are based on the material accumulated in the previous years from the lectures given by Stefan Schmitt and Christian Autermann

Calculation of Exclusion Limits
— goals —

* read a limit plot

* formulate null- and alternative- hypotheses

* understand the mechanism behind the limit calculation

* estimate and interpret the p-value

* explain limits to “outside” wild world

Calculation of Exclusion Limits
— rules —

well, there are no rules, just kind of guide lines…

* communication is (almost) everything - language? doesn't matter as long as we can understand each other

* if something is not clear, don't hesitate - **go ahead and ask**

* if you are hanging in a C++, root or python issue, don't waste more than 10 min searching for a solution - **go ahead and ask**

* if you are fed up and need a break - **go ahead and say it**, we all need this once in a while :)

"go ahead and ask" is kind of solving most of the problems

overview

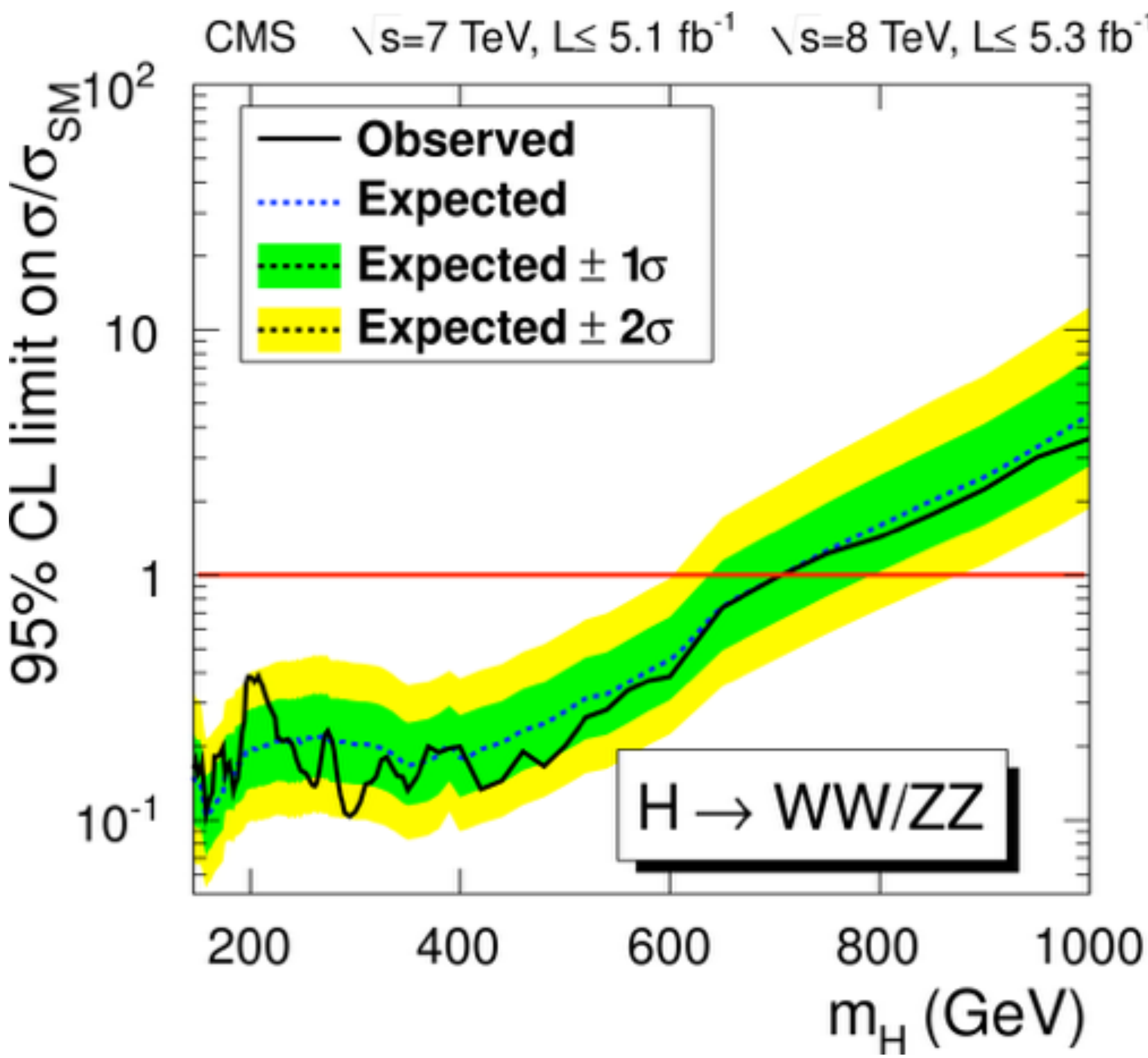
* a limit plot: what does it say? how is it done?

* confidence level
— confidence intervals
— confidence belt

* setting up limits: in general and in high energy physics

* p-value: what is it? how is it calculated?

a limit plot…



Eur. Phys. J. C 73 (2013) 2469

arXiv:1304.0213

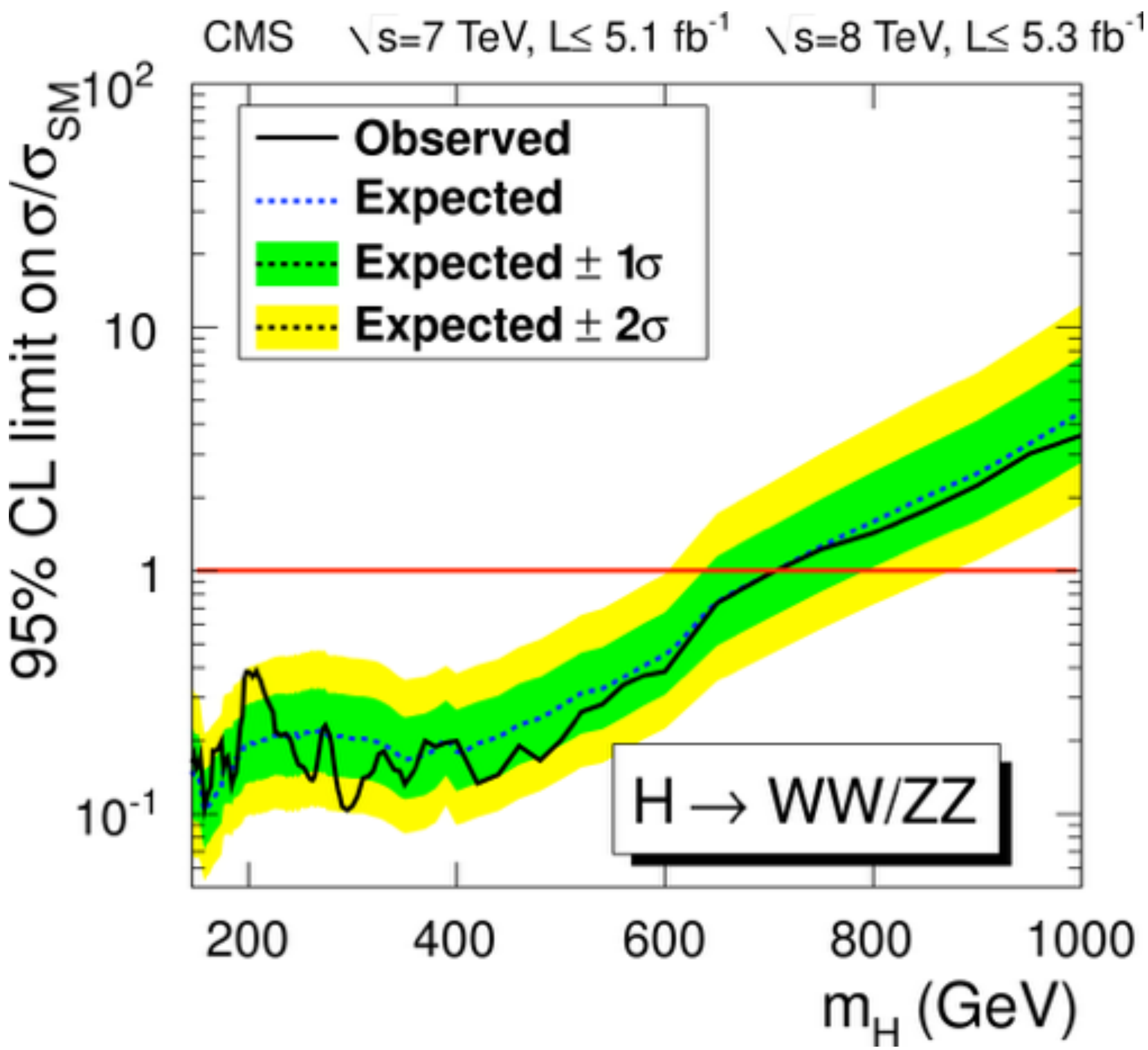
* what shall we understand from the y- and x-axes description ?

* what does the color code mean?

* where can we exclude a heavy standard-model-like Higgs?

for the ones that wake up late,
this is the first exercise

a limit plot…



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

* y-axis description:

* x-axis:

* colour code:

— **red:**

— **black:**

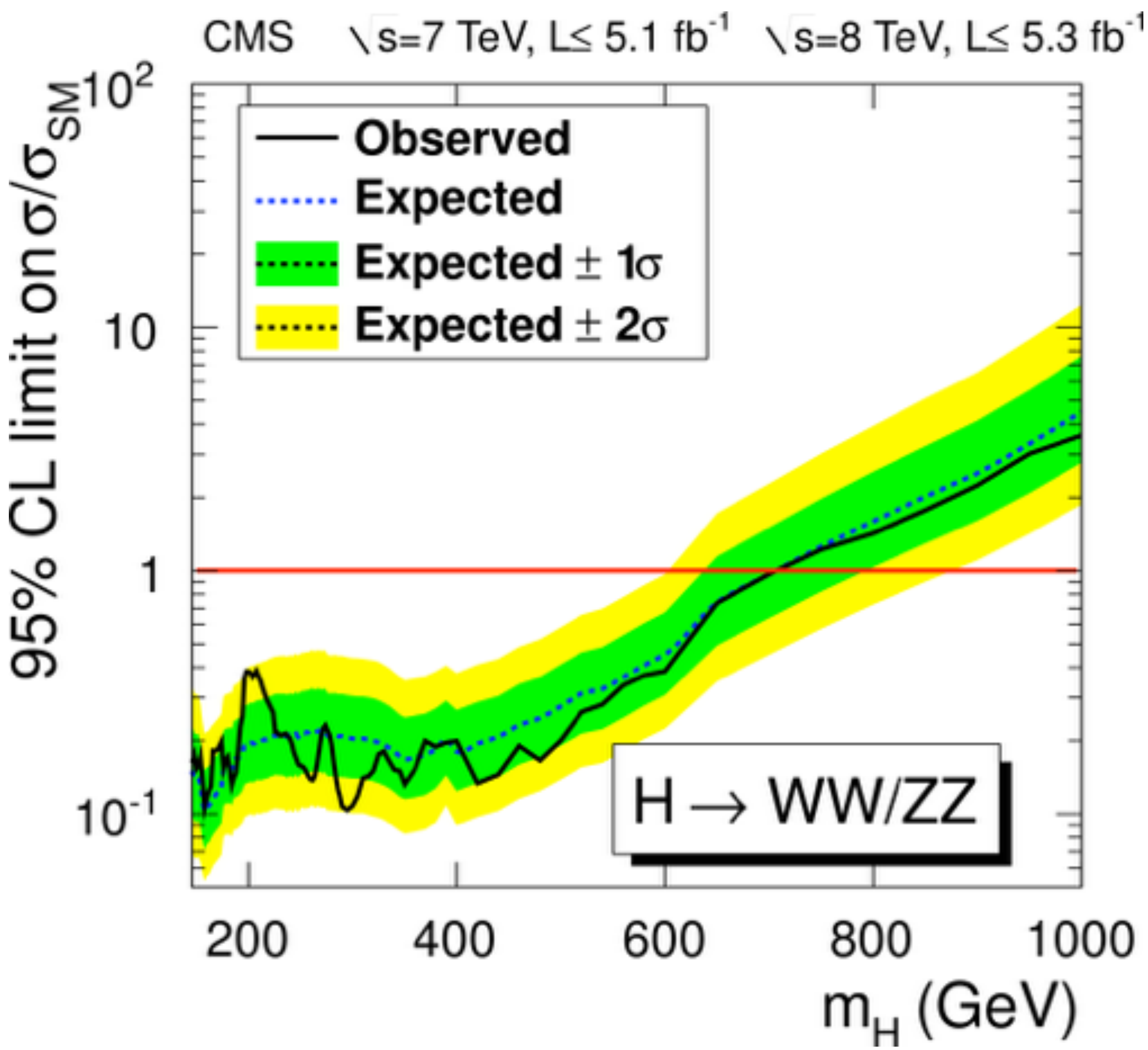
— **blue:**

— **green:**

— **yellow:**

Legend: black dashed line on yellow and on green are wrong, they should have been blue

a limit plot…



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

* y-axis description: **95% CL limit on σ/σ_{SM}**

95 % confidence level (CL) limit on the ratio of the production cross section (**σ**) to the standard model (SM) expectation (**σ_{SM}**)

* x-axis: **m_H (GeV)**

* colour code:

— **red: σ/σ_{SM}** = 1

— **black:** observed upper limit

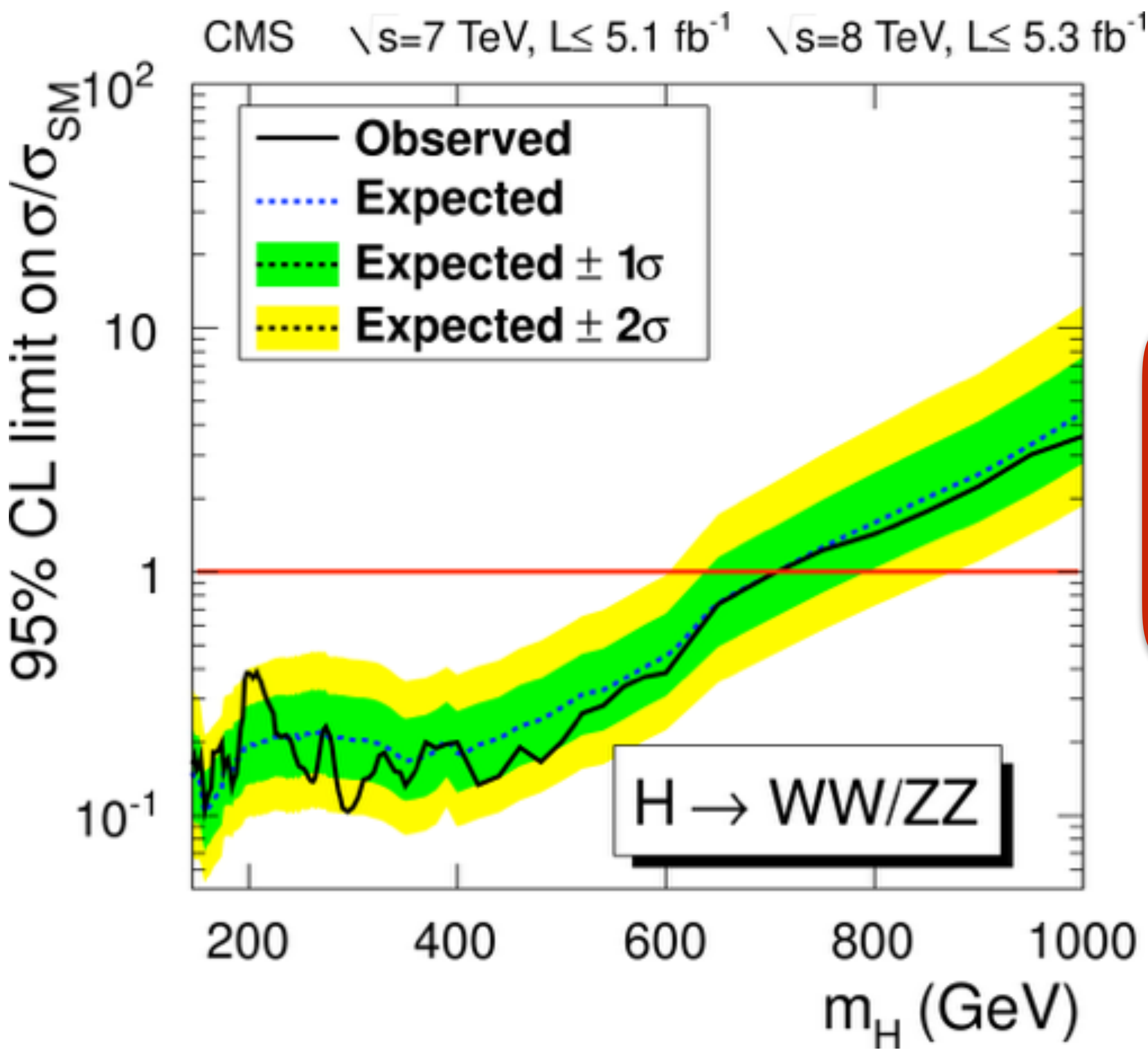
— **blue: dashed line** expected limit

— **green: 68%** (1σ) CL ranges of expectation

— **yellow: 95%** (2σ) CL ranges of expectation

Legend: black dashed line on yellow and on green are wrong, they should have been blue

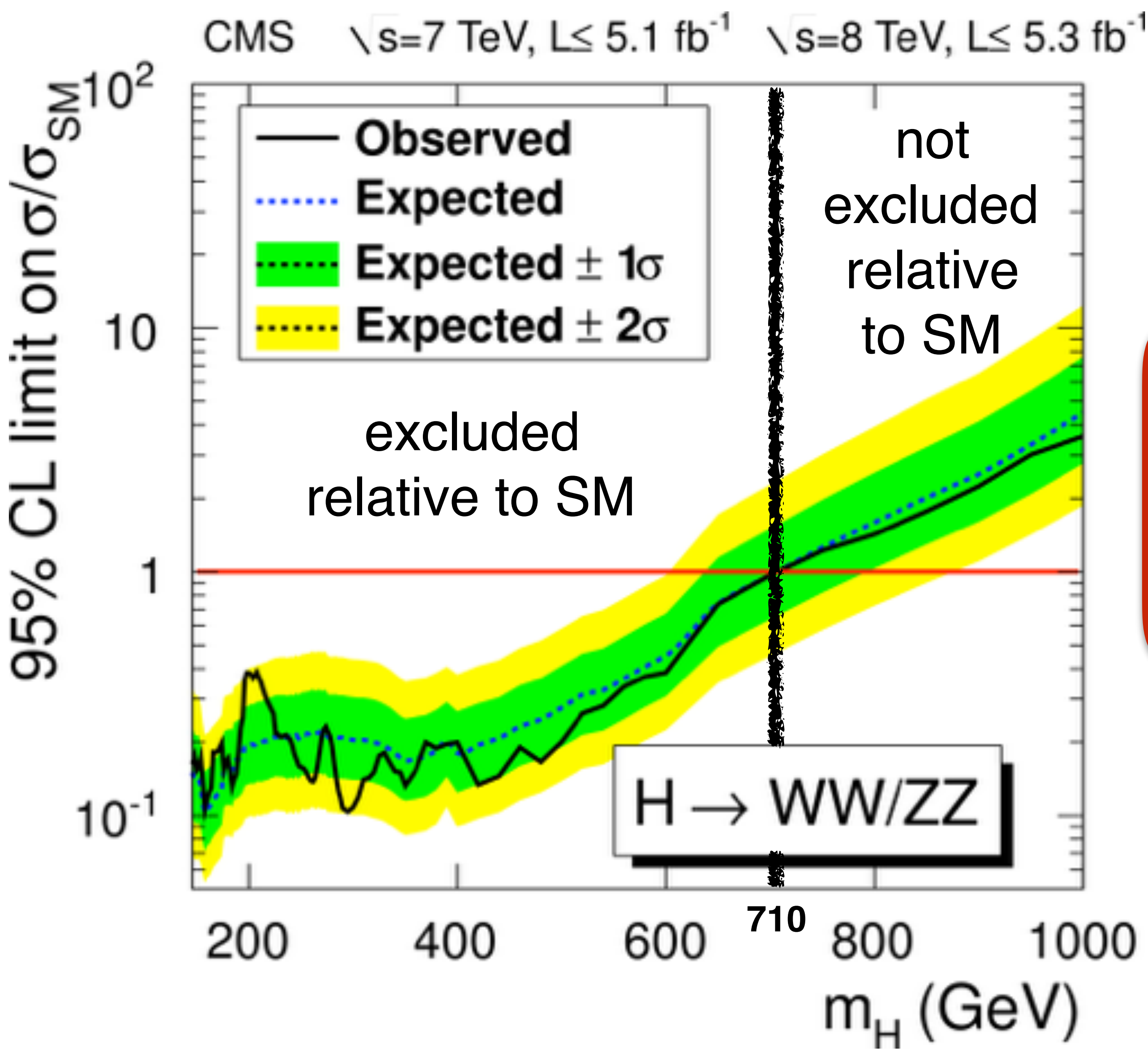
a limit plot…



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

* where can we exclude a heavy standard-model-like Higgs?

a limit plot...

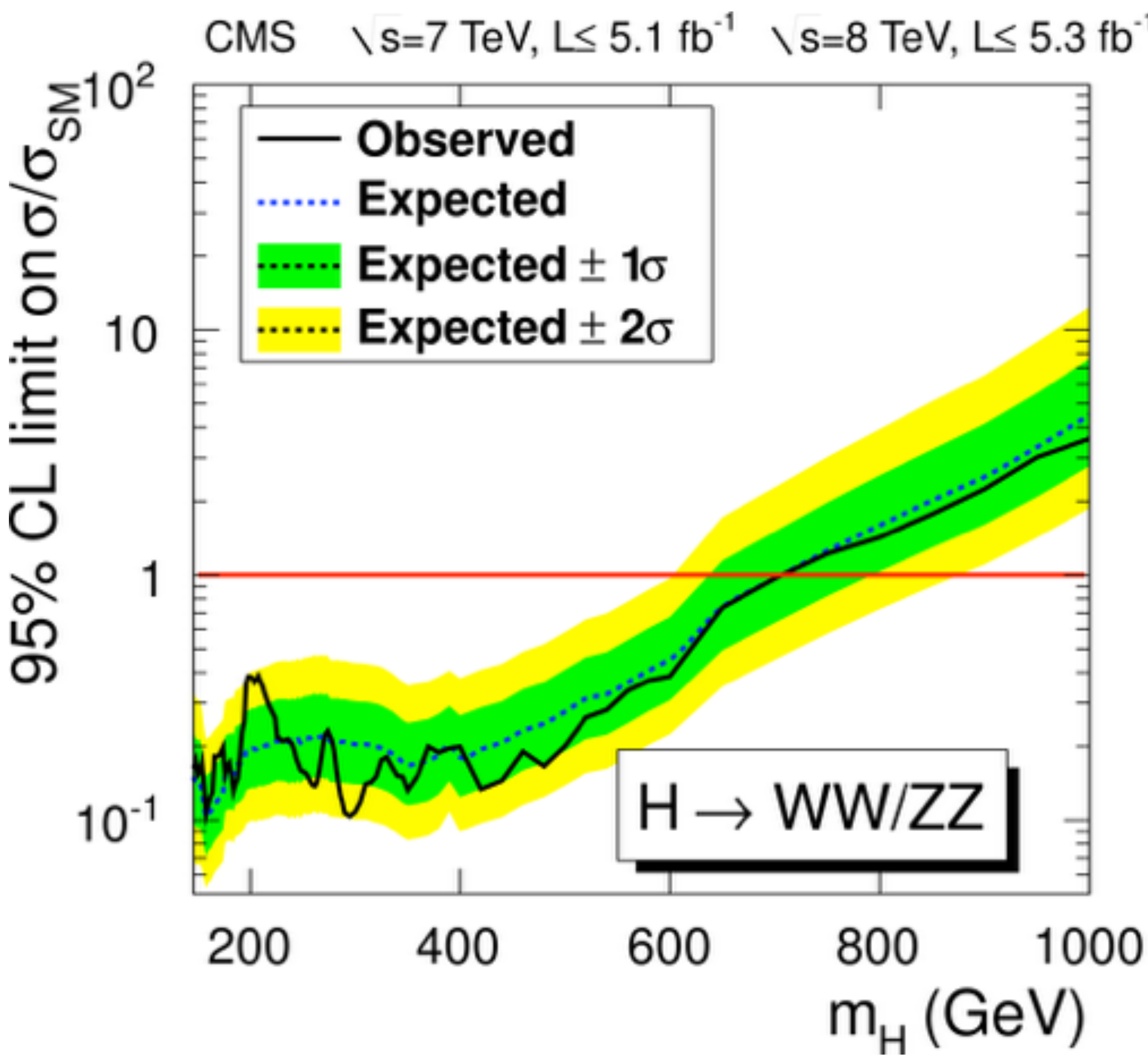


Eur. Phys. J. C 73 (2013) 2469
arXiv:1304.0213

*** where can we exclude a heavy standard-model-like Higgs?**

in the range $145 < m_H < 710$ GeV

a limit plot…
— let's compare with what authors wrote —



Eur. Phys. J. C 73 (2013) 2469

arXiv:1304.0213

from the abstract

The combined upper limits at 95% confidence level on products of the cross section and branching fractions exclude a standard-model-like Higgs boson in the range 145 < m_H < 710 GeV, thus extending the mass region excluded by CMS from 127–600 GeV up to 710 GeV.

in section: 4 Data Analysis

Figure 2: … The 68% (1σ) and 95% (2σ) CL ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The horizontal solid line at unity indicates the SM expectation.

in section: 6 Summary

Figure 11: Observed (solid line) and expected (dashed line) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson with all WW and ZZ channels combined.

few buzzwords:

* expected limit
* observed upper limit
* **95 % confidence level (CL)**
* **68%** (1σ) CL ranges of expectation
* **95%** (2σ) CL ranges of expectation

limit:

* a point beyond which it is not possible to go

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

— there is a **theory**: SM

— this theory (SM): predicts **signal-like events**

— there is also a **detector**: CMS (in this example)

— this detector: has a certain **accuracy** and can “measure” signal-like events

— there is also an **analysis** group: collaboration

— this analysis group: exploits or not most sensitive **methods**

limit:

* a point beyond which it is not possible to go
* an amount or number that is the highest or lowest allowed

* expected limit:

is the upper 95% CL limit that COULD BE achieved given the predicted signal by a theory, the detector accuracy and analysis methods sensitivity

* observed upper limit:

is the upper 95% CL limit that WAS MEASURED given the predicted signal by a theory, the detector accuracy and analysis methods sensitivity

confidence level

* statistical measure for a test results that can be expected to be within a specified range:

95 % CL: a result will probably meet the expectations 95% of the time

* in order to understand the meaning of a confidence level we need two ingredients:

— confidence interval

— confidence belt

confidence interval

* an interval which reflects the statistical uncertainty of the measured parameter

* it should:

— communicate the result in an objective mode

— give probability of containing the true parameter

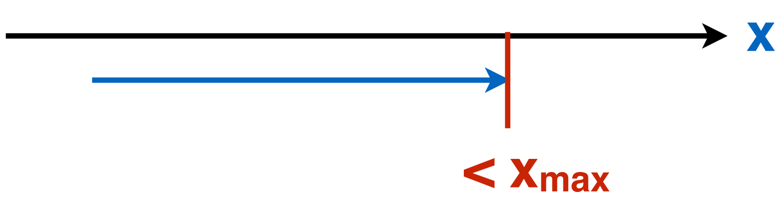
— if needed, provide information to draw conclusions about measured parameter (prior “beliefs”)

* it can be:

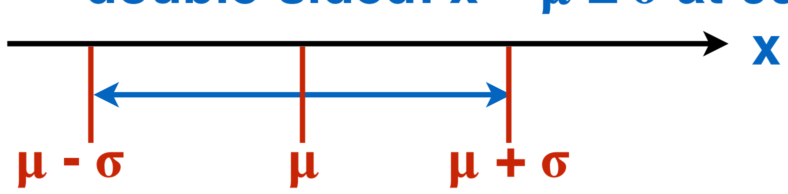
— single-sided

— double-sided

single-sided: $x < x_{max}$ at 95% CL

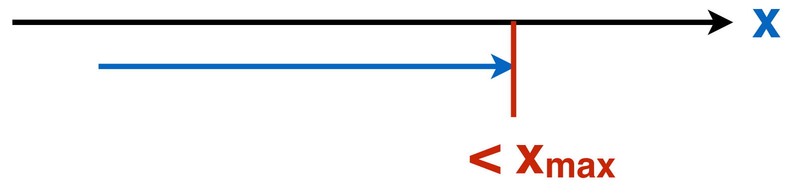


double-sided: $x = \mu \pm \sigma$ at 60% CL



confidence interval: single-sided

single-sided: $x < x_{max}$ at 95% CL

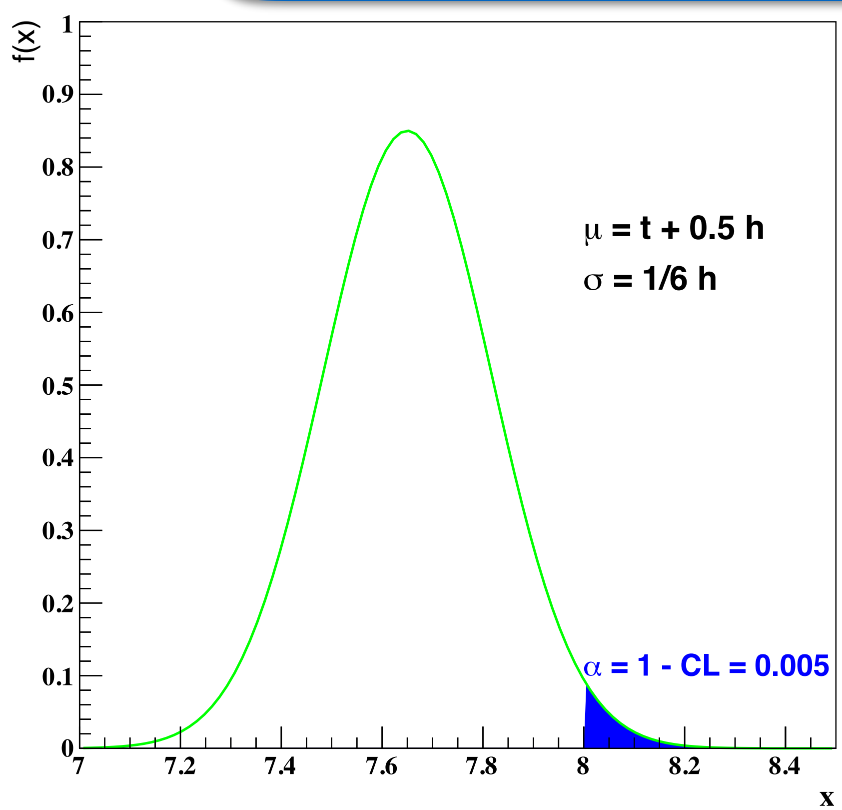


when do we need it? when we are lazy and want to sleep more, still not being fired for coming late at work

An employee needs to be at work at 8:00 o'clock sharp. The journey takes 30 minutes on average, with a Gaussian uncertainty of σ = 10 minutes due to traffic.

When must he leave home to be late only once a year (~0.5%)? Or to be in time in 99.5% of the cases?

in 2016 in Hamburg there are 254 working days, 30 days vacation: 1/224 ~0.0045



$$\int_{x_{arrival}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - CL$$

mean $\mu = x_{start} + 30'$

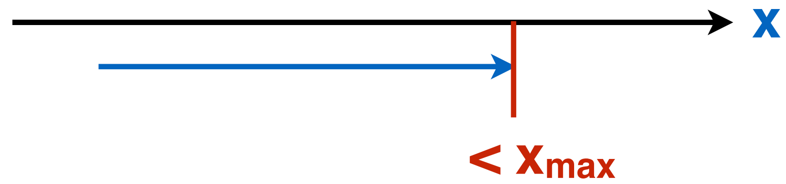
width $\sigma = 10'$

limit $x_{arrival} = 08{:}00$

x_{start} **(0.5%) = ?**

confidence interval: single-sided

single-sided: $x < x_{max}$ at 95% CL

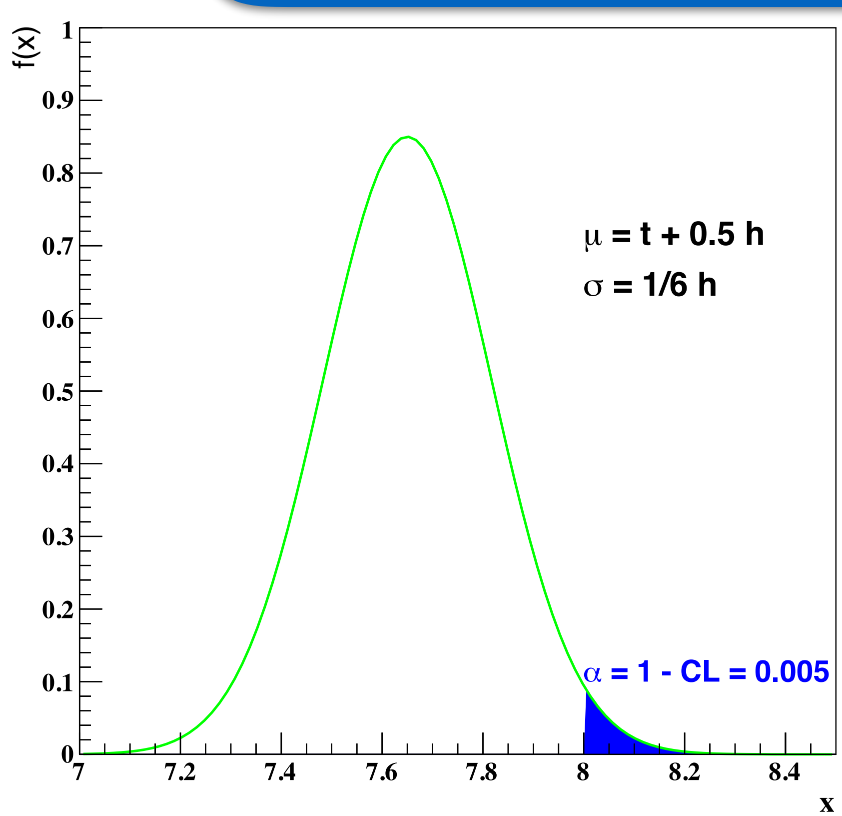


when do we need it? when we are lazy and want to sleep more, still not being fired for coming late at work

An employee needs to be at work at 8:00 o'clock sharp. The journey takes 30 minutes on average, with a Gaussian uncertainty of σ = 10 minutes due to traffic.

When must he leave home to be late only once a year (~0.5%)? Or to be in time in 99.5% of the cases?

in 2016 in Hamburg there are 254 working days, 30 days vacation: 1/224 ~0.0045



*** write a ROOT macro to estimate the correspondence between the number of sigmas and the single-sided interval for a Gauss distribution**

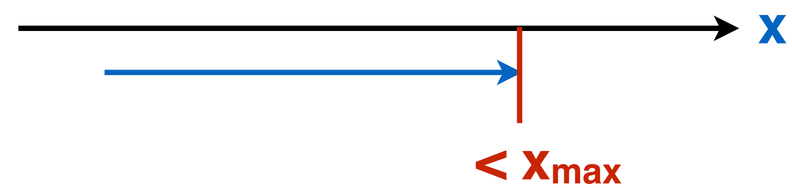
*** estimate number of sigmas for a single-sided interval of 1 - 0.005**

*** estimate starting time**

x_{start} (0.5%) = ?

confidence interval: single-sided

single-sided: $x < x_{max}$ at 95% CL

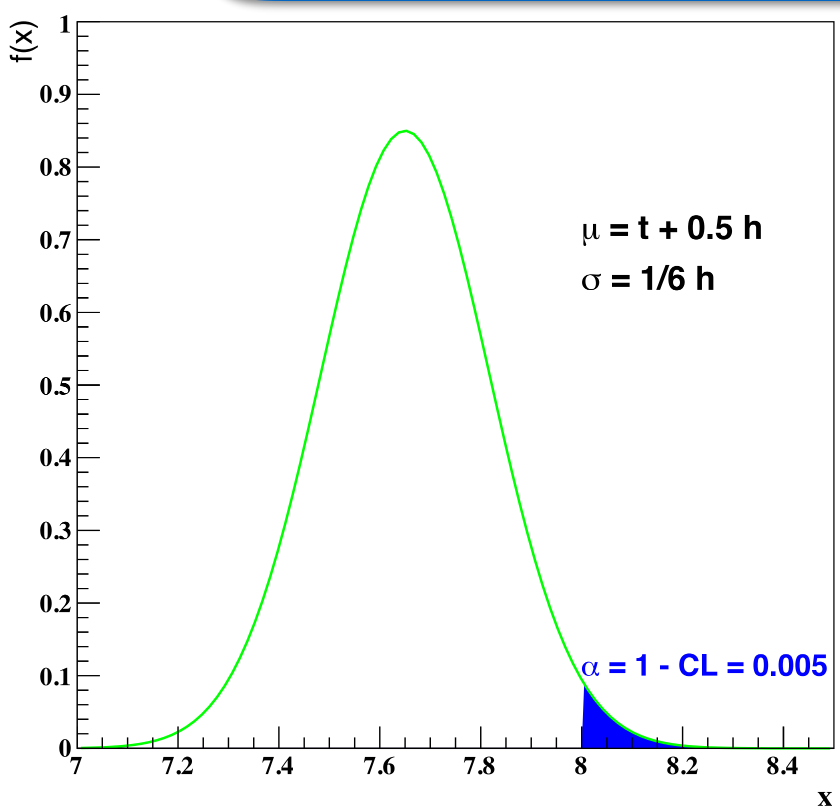


when do we need it? when we are lazy and want to sleep more, still not being fired for coming late at work

An employee needs to be at work at 8:00 o'clock sharp. The journey takes 30 minutes on average, with a Gaussian uncertainty of σ = 10 minutes due to traffic.

When must he leave home to be late only once a year (~0.5%)? Or to be in time in 99.5% of the cases?

in 2016 in Hamburg there are 254 working days, 30 days vacation: 1/224 ~0.0045



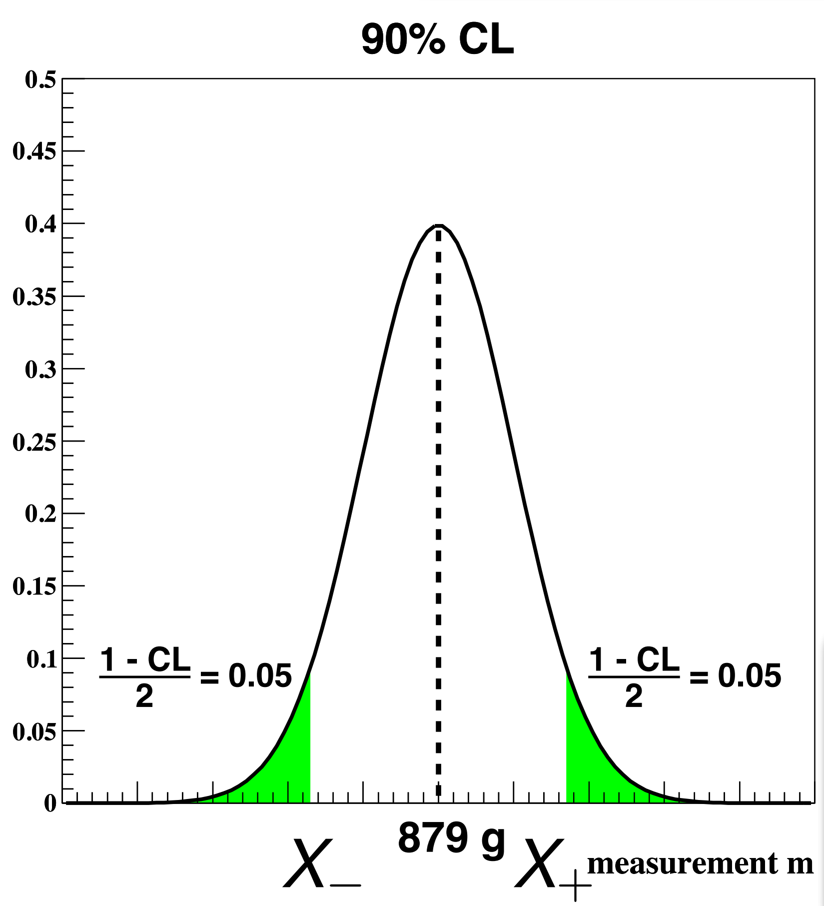
1 σ		84.13 %
2 σ		97.72 %
2.3 σ		98.92 %
2.6 σ	correspond to	99.53 %
2.7 σ		99.65 %
3 σ		99.86 %

x_{start} (0.5%) = 08:00 - 00:30 - 00:26 = 07:04

confidence interval: double-sided

* given a precisely known true value μ of a certain property, e.g., the weight of cereal packets 879g, one can ask:

— what is the weight-range into which a certain amount, e.g., 90%, of measurements x will fall?



central confidence intervals for a Gaussian distributions:

$$P(X_- \geq x \geq X_+) = \int_{X_-}^{X_+} P(x)dx = CL$$

x: measurement,
$X_\pm$: limits of the confidence interval.

* value of the measurement *m* lies in the interval *X-* . . . *X+* in „CL" % of the time

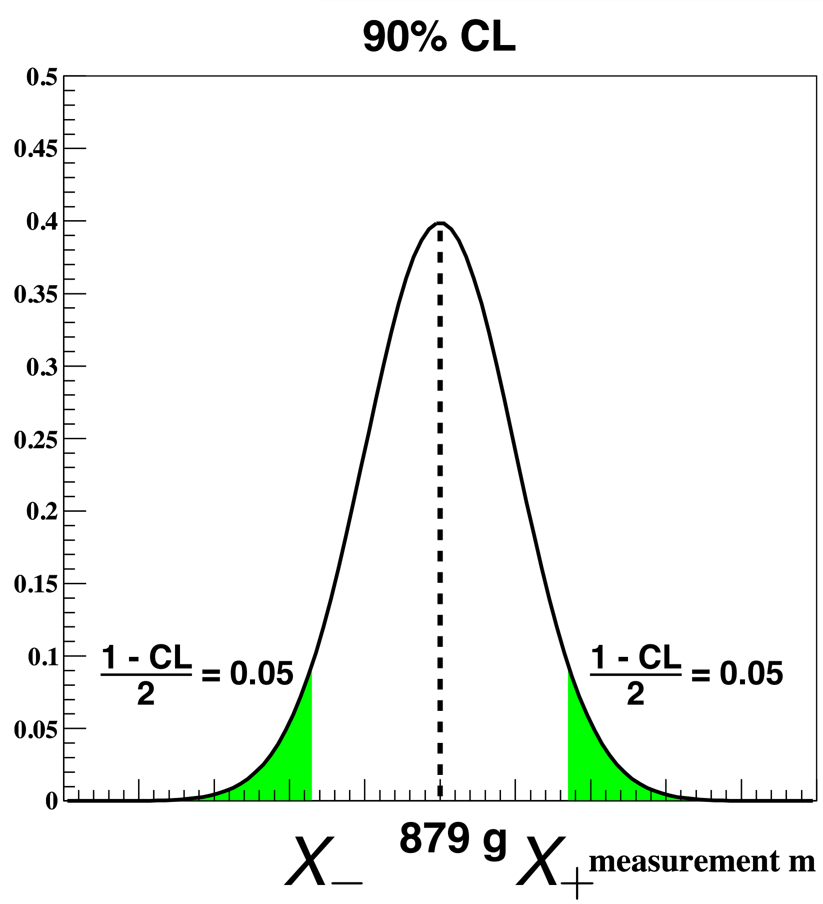
or

* saying „*m* will lie in the interval *X-* . . . *X+*" has "CL" % confidence

confidence interval: double-sided

* given a precisely known true value μ of a certain property, e.g., the weight of cereal packets 879g, one can ask:

— what is the weight-range into which a certain amount, e.g., 90%, of measurements x will fall?



* write a ROOT macro to estimate the correspondence between the number of sigmas and the double-sided interval for a Gauss distribution

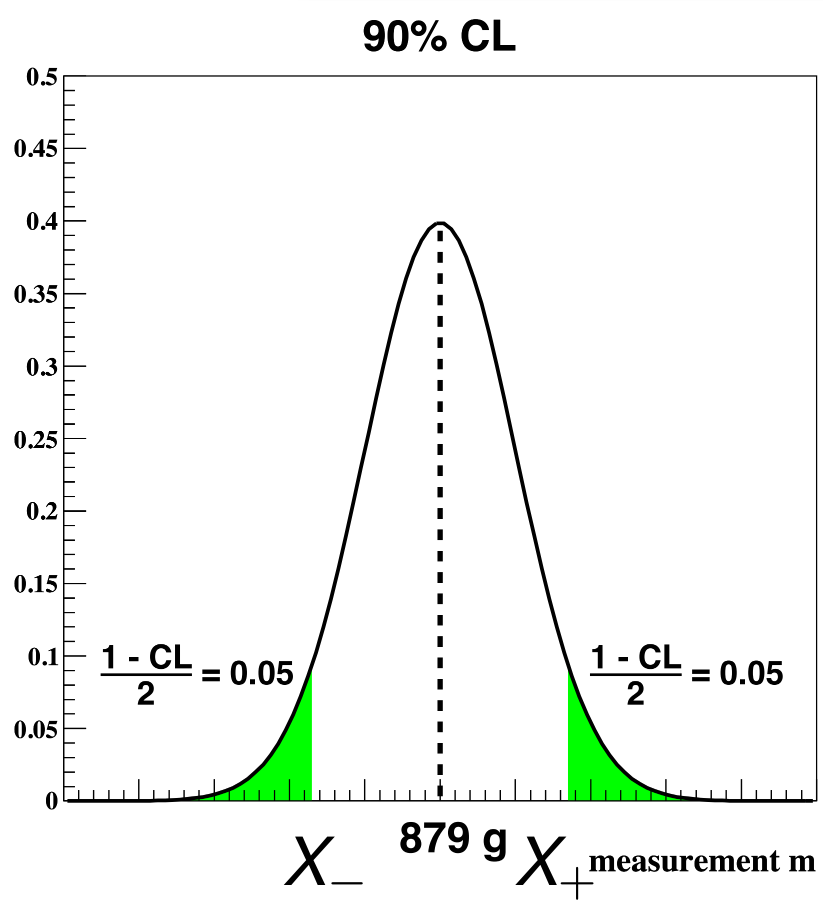
* estimate number of sigmas for a double-sided interval of 1 - 2*0.05

* estimate the weight-range

confidence interval: double-sided

* given a precisely known true value μ of a certain property, e.g., the weight of cereal packets 879g, one can ask:

— what is the weight-range into which a certain amount, e.g., 90%, of measurements x will fall?



1 σ		68.27 %
1.6449 σ		90.00 %
1.96 σ		95.00 %
2 σ	correspond to	95.45 %
3 σ		99.73 %

weight - range (90%) = μ ± 1.6449 σ at 90% CL

confidence: interval & belt

* so far we figured it out that 90% of the measurements of the cereal packet weight will be within $\mu \pm 1.6449\ \sigma$

* but, usually we buy only a packet, not the entire truck… so, what we could say about μ given that we have only one measurement x_0?

almost none of us like this one

some like this kind of belt to be confident

* we need a confidence belt to translate our measurement into a confidence interval

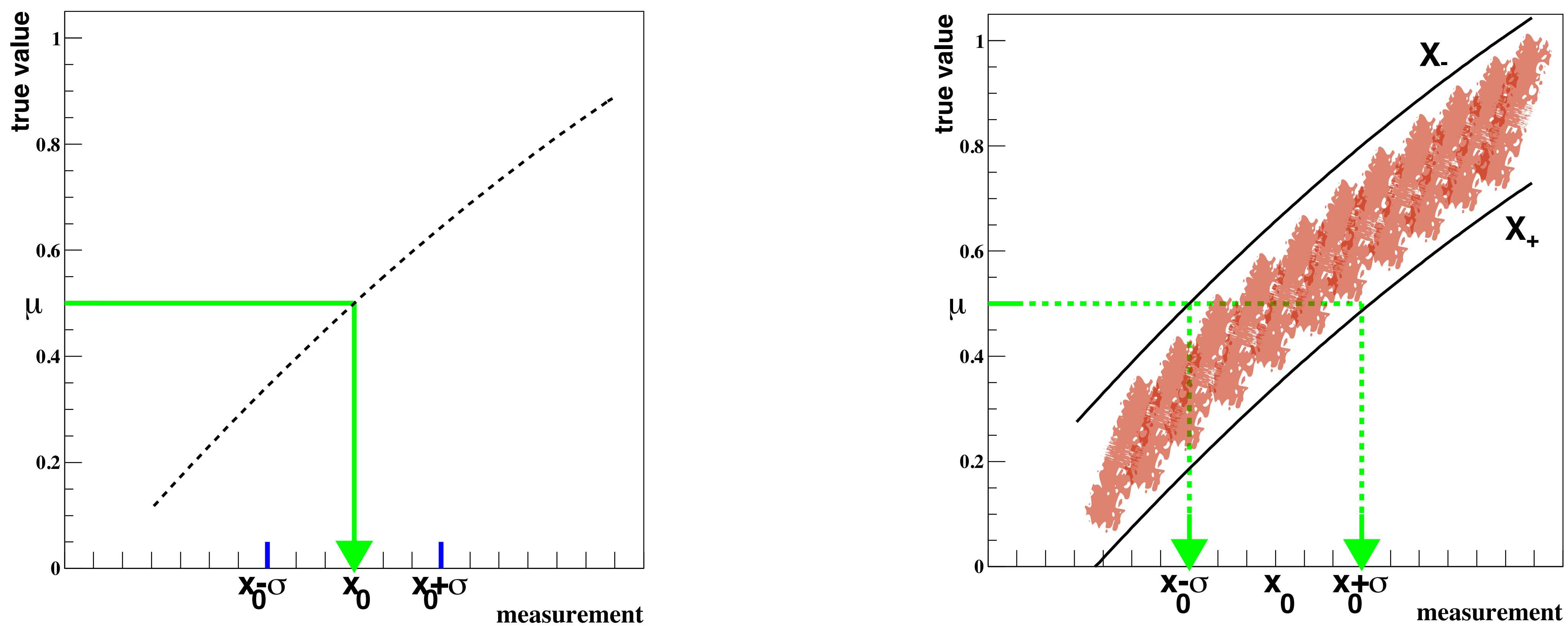
some others prefer this one

almost none of us would say no to wear this confidence belt

confidence belt

* for a particular true value μ, the probability density function $P(\mu,\sigma)$ defines the most probable measurement x_0, and the interval $x_0 - \sigma \ldots x_0 + \sigma$ into which the measurements will fall with a given CL

the confidence belt defined by $X_-(\mu)$ and $X_+(\mu)$



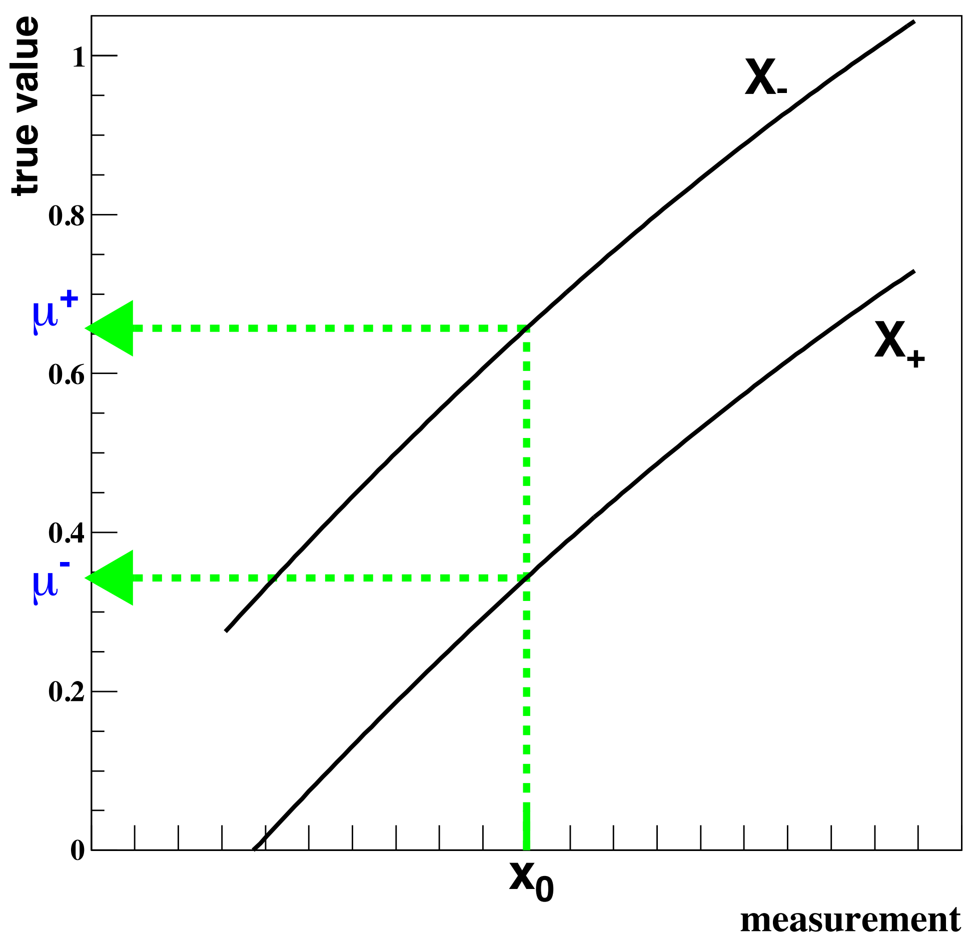
* different μ implies different measurements x_0 and limits $x_0 \pm \sigma$

* measurement limits $x_0 - \sigma$ and $x_0 + \sigma$ are functions, X_- and X_+, of true value μ

how to use a confidence belt
— "inside" statistics —

* back to our problem:

— usually we buy only a packet, not the entire truck… so, what we could say about μ given that we have only one measurement x_0?



* for a measurement x_0, a confidence interval for the true value: $\mu^- \ldots \mu^+$, can be constructed from the confidence belt

* confidence belt is *constructed horizontally* using the known probability density for all possible true values μ

* with a measurement x_0, confidence belt is *read vertically*

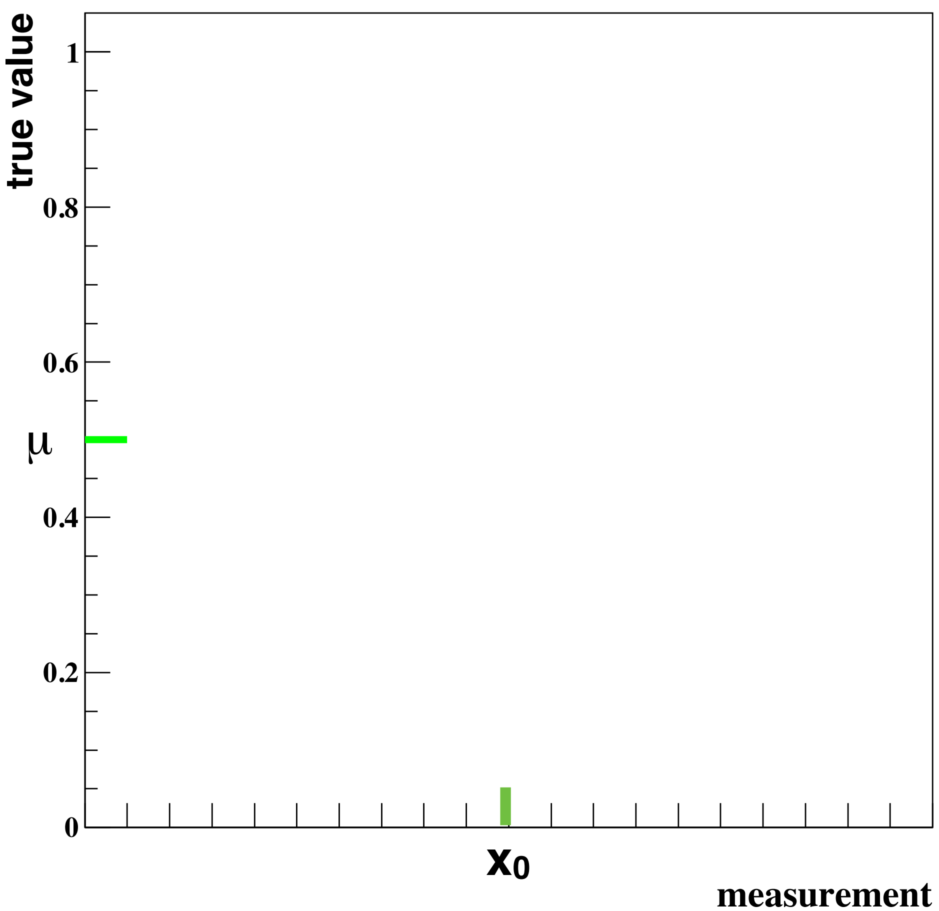
* interval $\mu^- \ldots \mu^+$ contains the true value μ with a "CL" probability

confidence belt for a Gaussian

if the producer of the cereal packet is not cheating, the weight of a packet should be normally distributed

*** back to our problem:**

— usually we buy only a packet, not the entire truck… so, what we could say about μ given that we have only one measurement x_0? now we know that the packet weight should follow a Gauss distribution



*** estimate shape of the confidence belt borders**

*** sketch the confidence belt**

confidence belt for a Gaussian

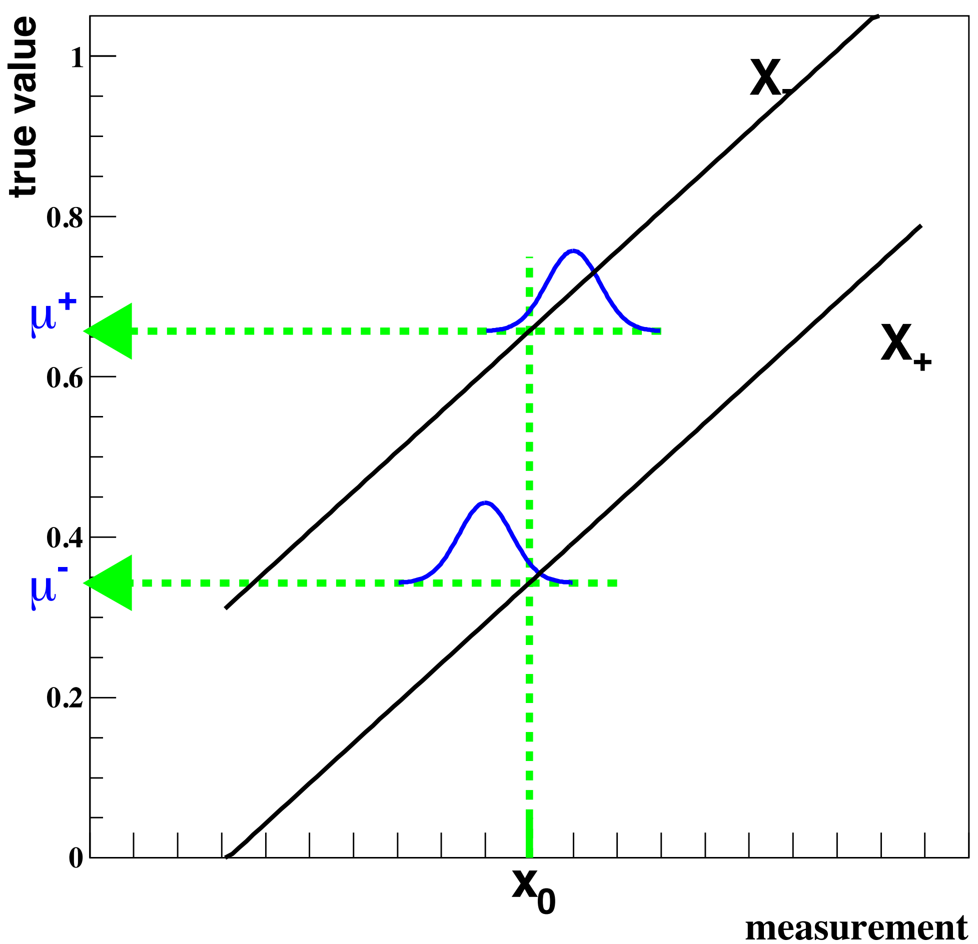
if the producer of the cereal packet is not cheating, the weight of a packet should be normally distributed

* back to our problem:

— usually we buy only a packet, not the entire truck… so, what we could say about μ given that we have only one measurement x_0? now we know that the packet weight should follow a Gauss distribution

$$\int_{x_0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu^-)^2}{2\sigma^2}} dx = \frac{1-CL}{2}$$

* equation for μ^-: requires x_0 to be "n" sigmas above μ^-



* confidence belt defined by X_- and X_+ becomes two straight lines with the corresponding number of sigmas as gradient:

$X_\pm = \mu \pm n * \sigma$ when constructed horizontally

$\mu^\pm = x_0 \pm n * \sigma$ when read vertically

with $n = 1$ for $CL = 68.27\%$, $n = 2$ for $CL = 95.45\%$

confidence intervals

* **Binomial** Confidence Intervals

* **Poisson** Confidence Intervals

* **Constrained** Confidence Intervals

binomial confidence interval

*** binomial experiments: only two possible outcomes 0 or 1**

*** discrete observed value: number of successes *m* (out of *n* trials)**

*** true value: single trial probability μ is continuous**

*** for *m* successes in *n* binomial trials the limits on the individual probabilities p_- and p_+ of the confidence interval "CL" are given by:**

$$\sum_{r=0}^{m-1} B(\mu, p_-, n) \leq \frac{\mathrm{CL}}{2}$$

$$\sum_{r=m+1}^{n} B(\mu, p_+, n) \leq \frac{\mathrm{CL}}{2}$$

with binomial distribution: $B(\mu, p, n) = \dfrac{n!}{\mu!\,(n-\mu)!}\, p^{\mu}\,(1 - p)^{n - \mu}$

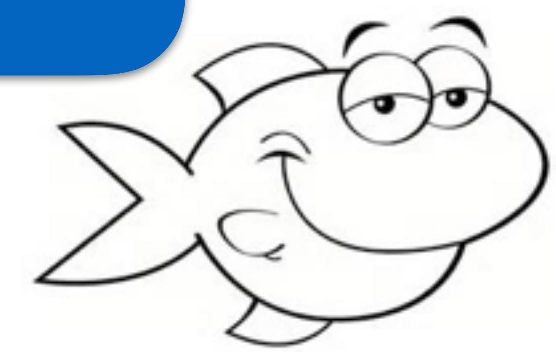
poisson confidence interval

some people remember this distribution having in mind ***Siméon Denis Poisson***

* poisson distribution approximates the binomial one for a large number of *n* trials and small probabilities *p*: $n \to \infty$ and $p \to 0$

* with probability depending on with k (number of success per interval) and λ (true expectation)

$$P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$



I remember it more like: (I catch it, I don't catch it) ^{n}

* and the limits of the confidence interval "CL" given by:

$$\sum_{r=0}^{k-1} P(r, \lambda_-) \leq \frac{CL}{2} \qquad \sum_{r=n+1}^{\infty} P(r, \lambda_+) \leq \frac{CL}{2}$$

poisson confidence interval
— proton decay —

* Super-Kamiokande (50 000 tons of water) observes less than *s* proton-decay candidate events per year

* what is the 95% CL interval for proton-decays and the proton half-life, *assuming no background* events and *s* = 1 found event per year?

* write a ROOT macro to estimate the confidence intervals of the poisson probability for *n* signal events

* estimate the time our experimentalist has to “work”

pssst, don’t disturb!
experimentalist working

poisson confidence interval — proton decay —

* **Super-Kamiokande (50 000 tons of water) observes less than *s* proton-decay candidate events per year**

* **what is the 95% CL interval for proton-decays and the proton half-life, *assuming no background* events and *s* = 1 found event per year?**

	Lower			Upper		
n	90%	95%	99%	90%	95%	99%
0	—	—	—	2.30	3.00	4.61
1	0.11	0.05	0.01	3.89	4.74	6.64
2	0.53	0.36	0.15	5.32	6.30	8.41
3	1.10	0.82	0.44	6.68	7.75	10.05
4	1.74	1.37	0.82	7.99	9.15	11.60
5	2.43	1.97	1.28	9.27	10.51	13.11
6	3.15	2.61	1.79	10.53	11.84	14.57
7	3.89	3.29	2.33	11.77	13.15	16.00
8	4.66	3.98	2.91	12.99	14.43	17.40
9	5.43	4.70	3.51	14.21	15.71	18.78
10	6.22	5.43	4.13	15.41	16.96	20.14

* **Number of protons in 50 000 tons of water: $N = 1.65 \cdot 10^{34}$**

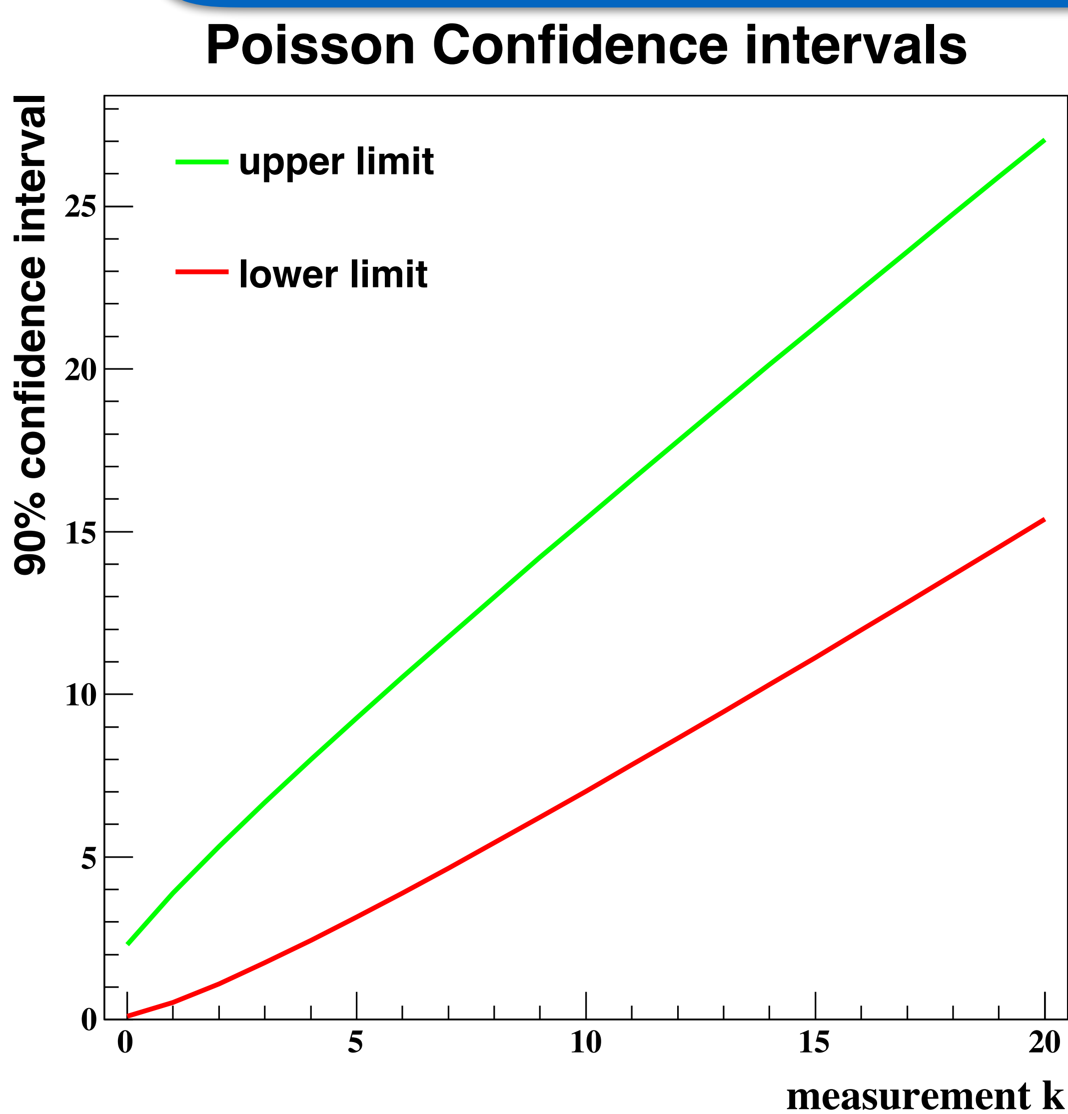
* **the 95% CL interval for 1 event: CL$_{dn}$ = 0.05, CL$_{up}$ = 4.74**

pssst, don't disturb! experimentalist still working

poisson confidence interval — proton decay —

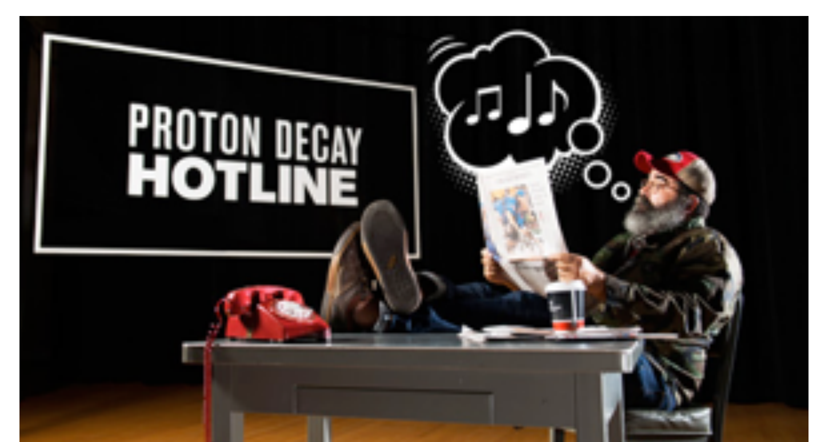
* Super-Kamiokande (50 000 tons of water) observes less than *s* proton-decay candidate events per year

* what is the 95% CL interval for proton-decays and the proton half-life, ***assuming no background*** events and *s* = 1 found event per year?



* probability for one decay/year:
$P = \mathrm{CL}\, dn/N = 3.03 \cdot 10^{-36} \ldots 2.87 \cdot 10^{-34}$

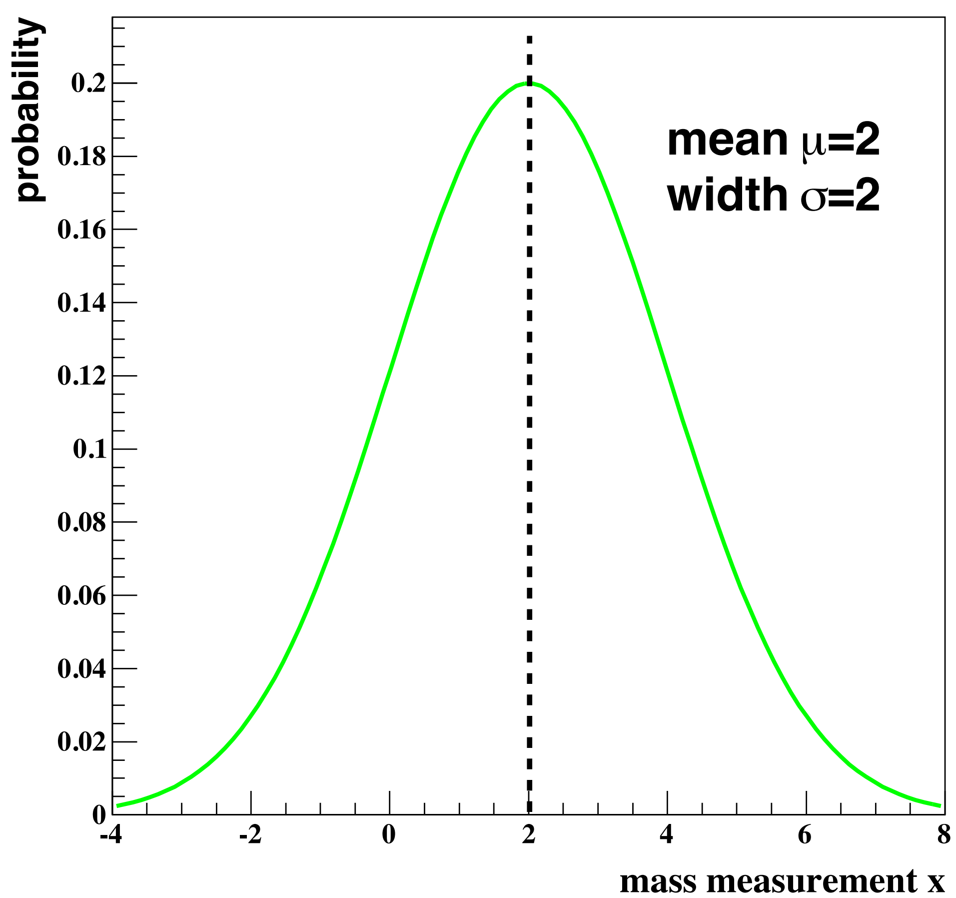
* and mean lifetime interval:
$3.48 \cdot 10^{33} < \tau = 1 / P < 3.3 \cdot 10^{35}$ years

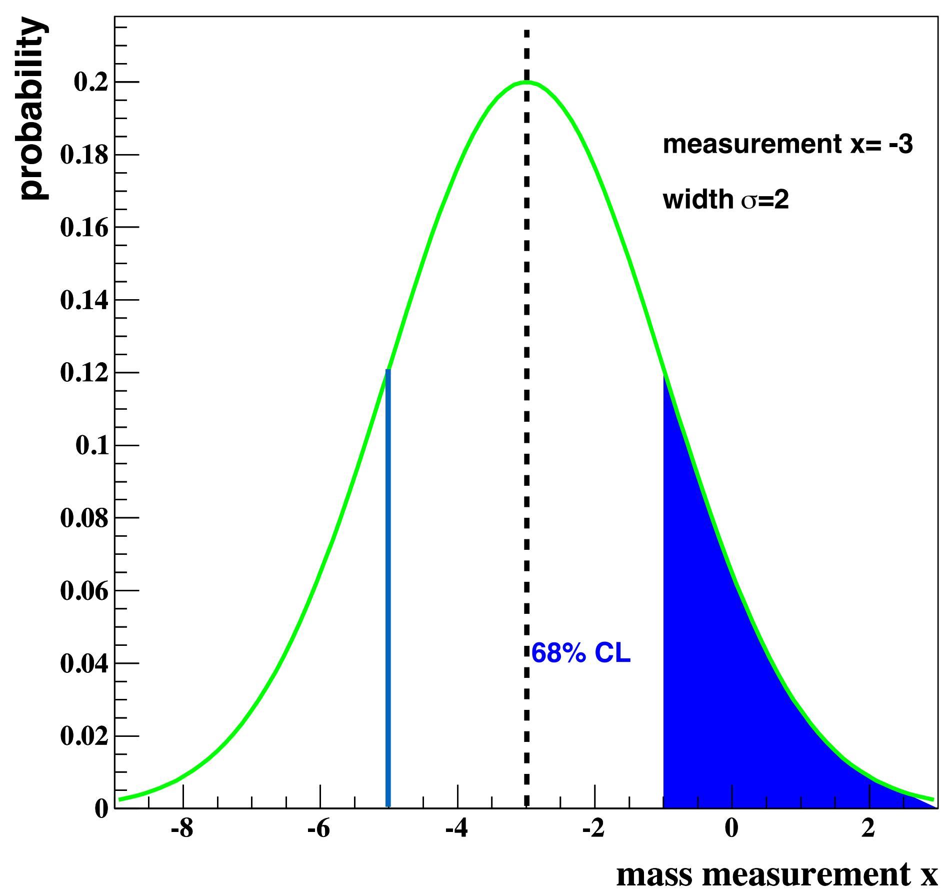


pssst, don't disturb! experimentalist (still) keeping working

constraint confidence interval

* let's assume we want to measure the mass of a "real" object, x
* some of the measurements lead to a negative upper limit - that's not "real" — it is absurd





* the mass of a "real" object is positive: the upper limit cannot be negative

* the way out: incorporate prior knowledge about the expected true value μ - Bayesian statistics

bayesian vs. frequentist

it's all about probabilities

* probability: predicts the number of expected events given

— the theory (which has parameters)

— the experimental setup (your favoured detector)

what we want to know:

— what a specific observation can say about a certain theory

* frequentist:

— give for each theory the probability to be observed, it does not give: the probability for a theory

* bayesian:

— assigns probabilities (degree of “belief”) to theories

high energy physics: use both of them, still prefer frequentist especially for discoveries

in case you wanted to know, but wan can live also without it

constraint confidence interval

* back to our problem:
let's assume we want to measure the mass of a "real" object, x

* some of the measurements lead to a negative upper limit - that's not "real" — it is absurd

* show how one can estimate the confidence interval for this kind of cases where a property is prior known

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

constraint confidence interval

* back to our problem:
let's assume we want to measure the mass of a "real" object, x

* some of the measurements lead to a negative upper limit - that's not "real" — it is absurd

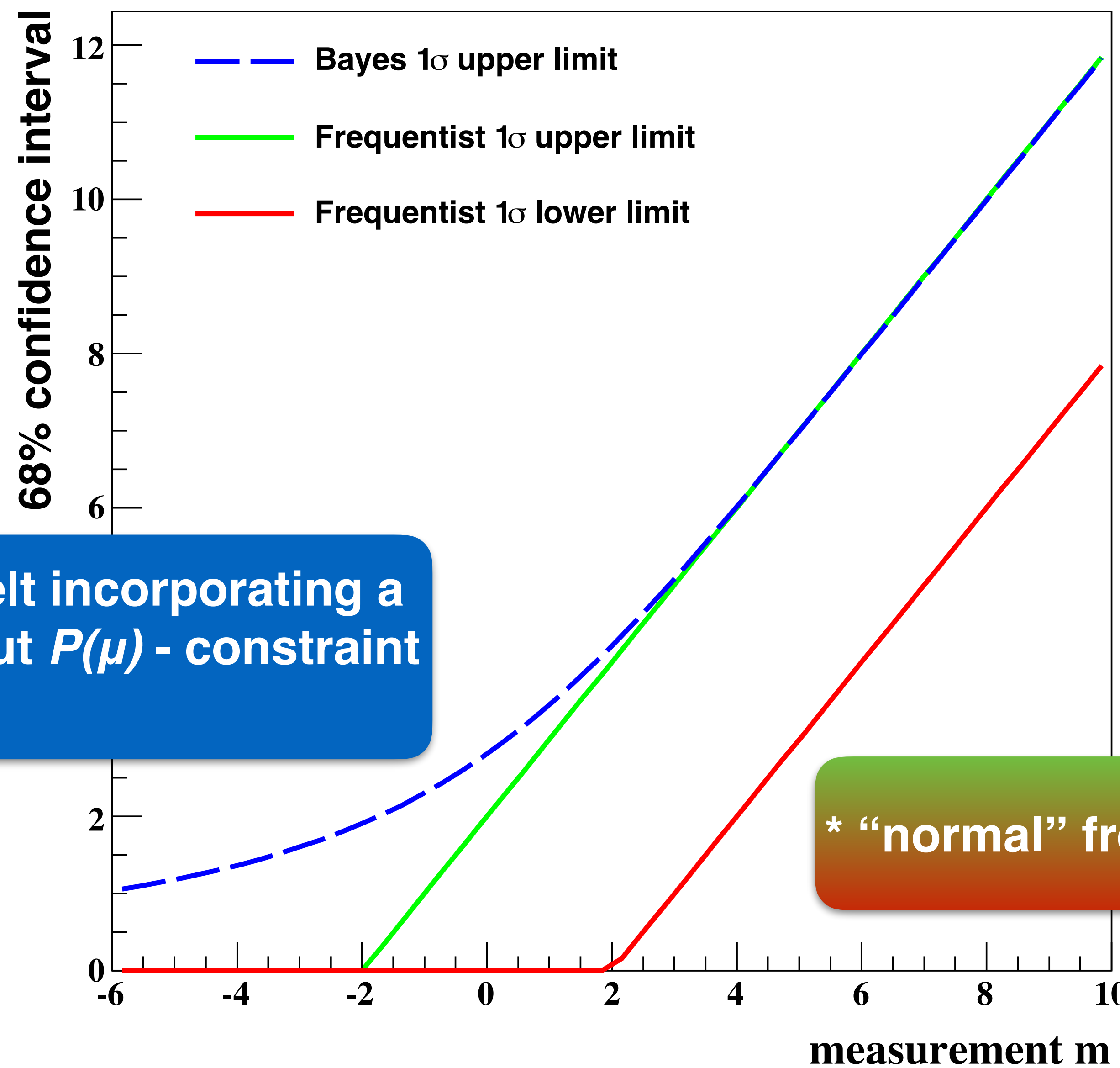
* bayesian statistics can incorporate prior knowledge about the true value *μ*

* our problem becomes: we a mass measurements of a "real" object, x, with the true value *μ constraint to be positive*

$$P(\mu_{up}|x) = \frac{P(x|\mu)}{P(x)} \cdot P(\mu) = \frac{\int_{-\infty}^{\mu_{up}} \mathrm{Gauss}(\sigma, x - x')dx'}{\int_{0}^{\infty} \mathrm{Gauss}(\sigma, x - x')dx'} \times \begin{cases} 1 & \text{for} \quad x > 0 \\ 0 & \text{else} \end{cases} = 1 - \frac{\mathrm{CL}}{2}$$

* solving the equation above for a certain "CL" one can obtain μ_{up} *(equivalent* μ_{down}*)*

frequentist vs. bayesian
— 68% confidence belt —



*** bayesian confidence belt incorporating a flat prior knowledge about *P(μ)* - constraint to positive values**

*** "normal" frequentists confidence belt**

summary — 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

*** what comes next:**
— 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

questions

*