A Picnic with Statistics
What does $10 \pm 2\,\mathrm{m}$ mean?

Piotr Nieżurawski
pniez@fuw.edu.pl
Physics Faculty
Warsaw University

Warsaw, 2007.05.09

Introduction

Basic information about experiment:

- Unknown value of physical parameter which we want to determine: μ
  Assumptions: this parameter has *really* one value...
- Probability density function (p.d.f.) of expected result x for given μ: $p(x|\mu)$
  Assumptions: we determine it with *very good* precision

I assume a simplified model:

- x is a continuous random variable *(warning: never true in any experiment!!!)*
- $p(x|\mu)$ is a Gaussian p.d.f. with dispersion (standard deviation) σ_0:

$$p(x|\mu) = G(x|\mu; \sigma_0)$$

One measurement x_1

In our experiment we measure value $x = x_1$. In our paper we write:

$$\mu = x_1 \pm \sigma_0$$

What does it mean?

Classical interpretation/convention:

a value of estimator of μ is x_1 and dispersion of $p(x|\mu = x_1)$ is σ_0.

Many readers will continue reasoning:

as nothing is said about $p(x|\mu = x_1)$, then it is a Gaussian p.d.f.

$$p(x|\mu = x_1) = G(x|\mu = x_1; \sigma_0)$$

To be more precise frequentists will tell that the interval

$$(\mu_{down},\ \mu_{up})\ ,$$

where

$$\begin{aligned}\mu_{down} &= x_1 - \sigma_0 \\ \mu_{up} &= x_1 + \sigma_0\ ,\end{aligned}$$

covers true value μ with confidence level $CL = P(1\sigma) \approx 68\%$.

To say it they will assume that σ_0 does not depend on μ (as we obtained in our experiment):

$$p(x|\mu) = G(x|\mu;\sigma_0)$$

Explanation of the slang:

probability on n-σ level, $P(n\sigma)$, is equal to

$$P(n\sigma) = \int_{-n}^{+n} G(x|\mu = 0; \sigma = 1)\, dx$$

Construction of frequentist confidence interval

Determine for which values of μ'

$$\int_{x:\, p(x|\mu')\geq p(x_1|\mu')} p(x|\mu')\, dx \leq CL$$

Fraction of CL experiments after applying this procedure will cover with their set of μ' values the true μ value!

What is arbitrary in this formula?

Nearly everything is up to you:

– value of CL,

– *ordering principle, i.e.* condition determining over which regions of x to integrate; here I integrate over values of x which fulfill

$$p(x|\mu') \geq p(x_1|\mu') \ ,$$

i.e. over set of x values for which value of p.d.f. $\geq$ than value of p.d.f. for result x_1.

In our case, because $p(x|\mu) = G(x|\mu;\sigma_0)$ and for Gaussian

$$G(x|\mu;\sigma_0) = G(x-\mu;\sigma_0) = G(\mu - x;\sigma_0)$$

we obtain a segment:

$$(\mu_{down},\ \mu_{up})\ ,$$

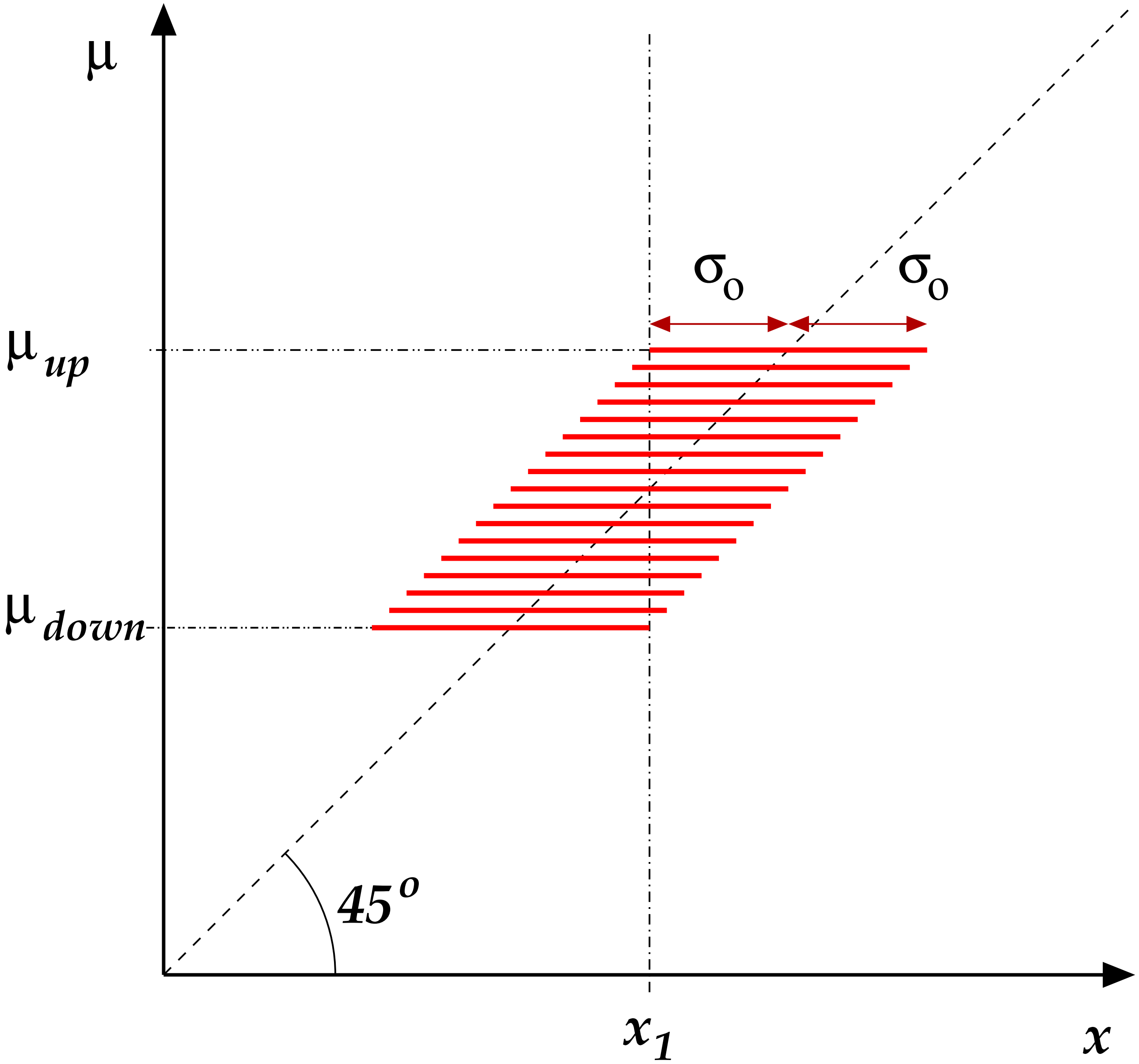
where $\mu_{down} = x_1 - n\sigma_0$ and $\mu_{down} = x_1 + n\sigma_0$,

which covers true value μ with confidence level $CL = P(n\sigma)$.

For $n = 1$ we write shortly:

$$\mu = x_1 \pm \sigma_0 \quad (\text{with } CL = 68\%)$$

Frequentist confidence interval for One measurement, $CL = P(1\sigma)$



Let's try to be better! With $CL = P(1\sigma) \approx 68\%$ we have:

$$\mu = 10 \pm 2 \text{ m}$$

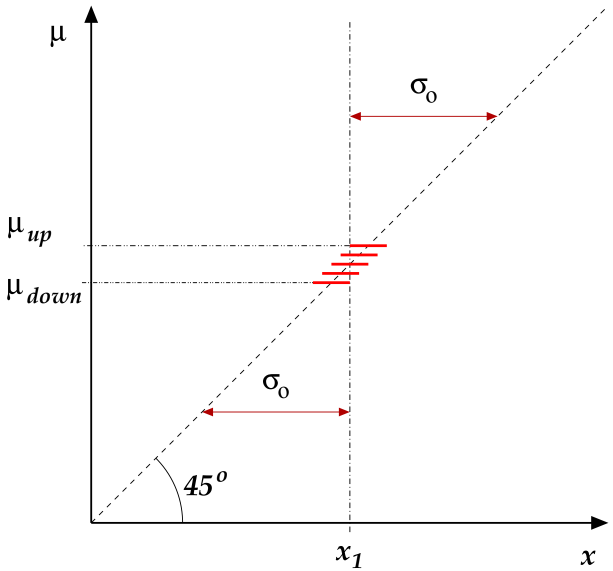
But this looks much nicer:

$$\mu = 10.00 \pm 0.02 \text{ m}$$

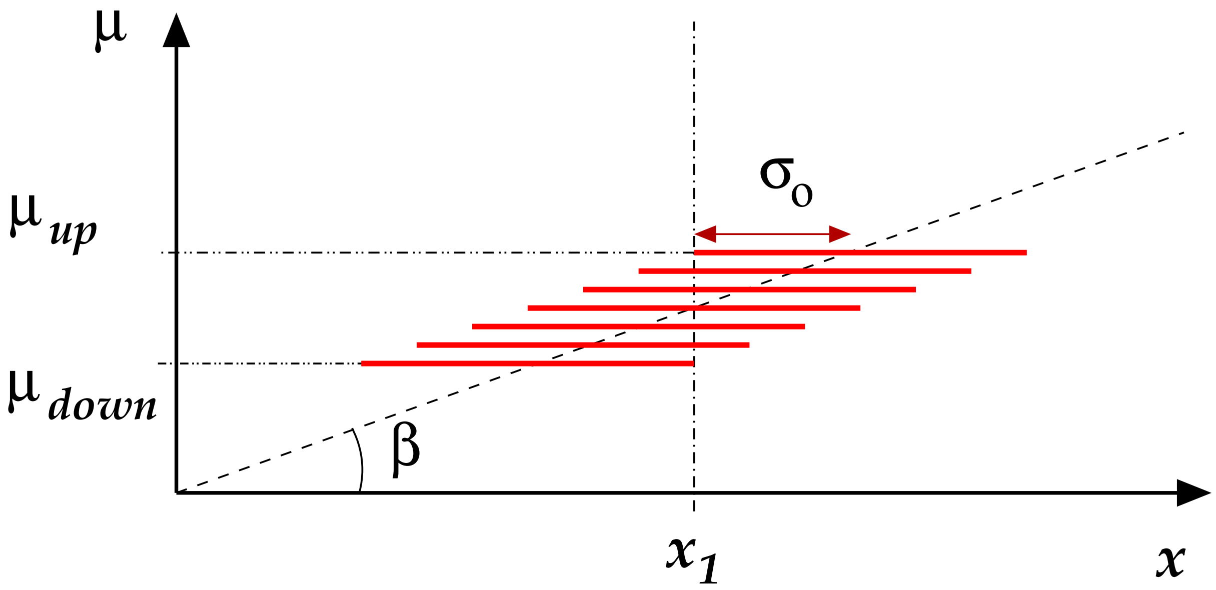
And is easy to obtain! We just assume $CL = P(0.01\sigma)$.

It's easy as well to estimate that $P(0.01\sigma) < 0.02/\sqrt{2\pi} < 0.8\%$.

Such small fraction of experiments will cover true value μ, thus we must have little confidence that our interval contains μ...



Complication I



Here we obtain $\mu_{down} = (x_1 - \sigma_0)\tan\beta$ and $\mu_{down} = (x_1 + \sigma_0)\tan\beta$, with $CL = P(1\sigma)$.

What's the problem? Measured quantity x is not the estimator of μ. But $x\tan\beta$ is. After linear transformation

$$x_N = x\tan\beta$$
$$\sigma_N = \sigma_0\tan\beta$$

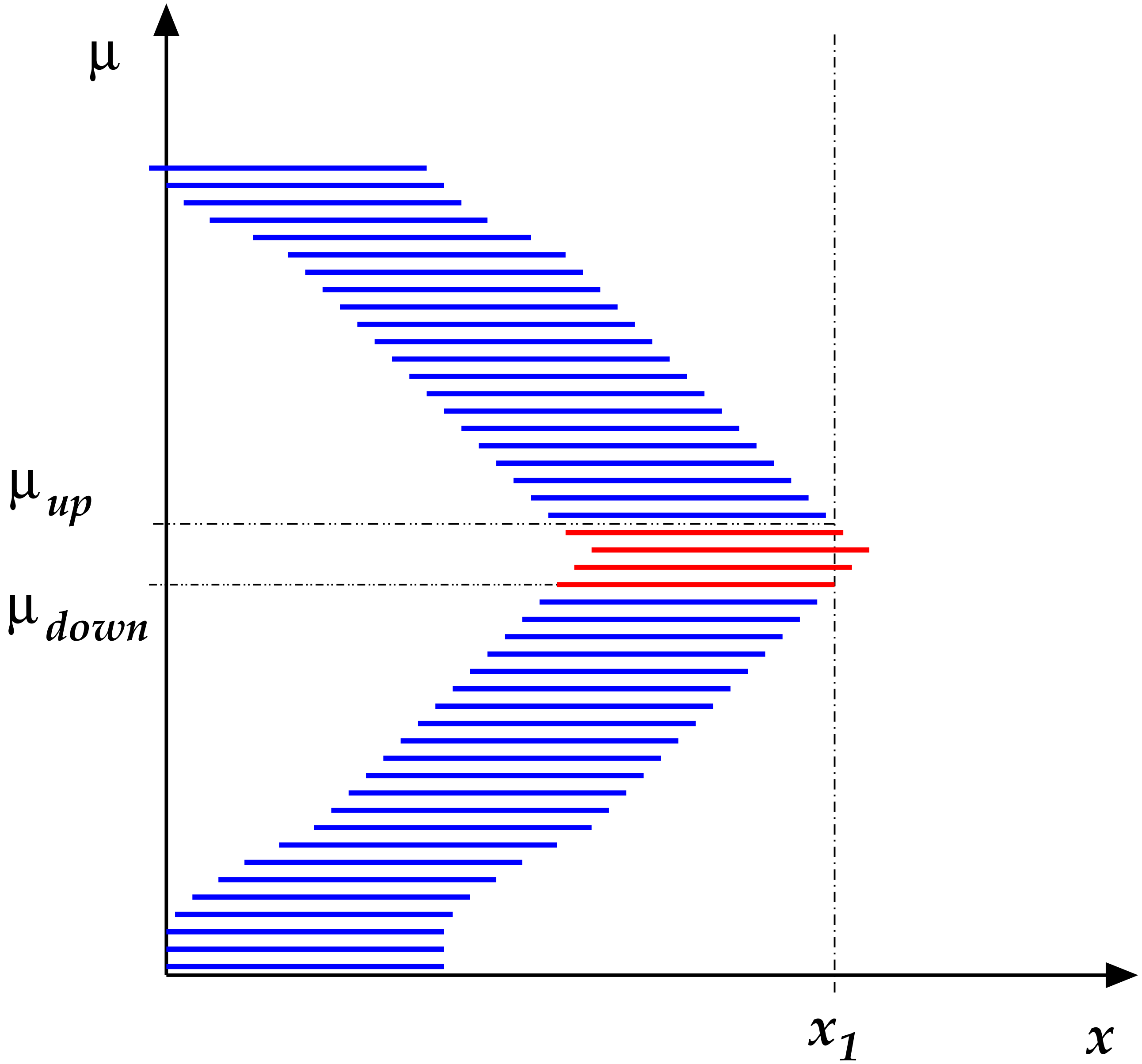
we change $p(x|\mu) = G(x|\mu\cot\beta;\sigma_0)$ into $p(x|\mu) = G(x_N|\mu;\sigma_N)$ and everything is much clearer.

In our previous examples for any assumed $CL > 0$ we always obtain one segment $(\mu_{down},\ \mu_{up})$ with non-zero length.

Now more interesting case!

Complication II

An experiment has a bit strange $p(x|\mu)$...



They will report one or two intervals for μ.

Or even no one at all (for fixed earlier CL)!

With $CL = P(1\sigma)$ they can obtain very narrow interval.

And in some cases they will get very narrow interval with $CL = P(5\sigma)$!!!

Is there really μ?

Now let's return to Gaussian p.d.fs. and read some papers about μ...

Second measurement: x_2

Other experiment reported that with $p(x|\mu) = G(x|\mu;\sigma_0)$ at $CL = P(1\sigma)$ they measured

$$\mu = x_2 \pm \sigma_0$$

How to combine both results: $x_1 \pm \sigma_0$ and $x_2 \pm \sigma_0$?

Most often used procedure – calculating weighted average – gives

$$(x_1 + x_2)/2 \;\pm\; \sigma_0/\sqrt{2}$$

But what is CL for this result?

Let's construct frequentist confidence interval. Probability density function for two independent measurements x and x':

$$p(x, x'|\mu) = G(x|\mu;\sigma_0)G(x'|\mu;\sigma_0)$$

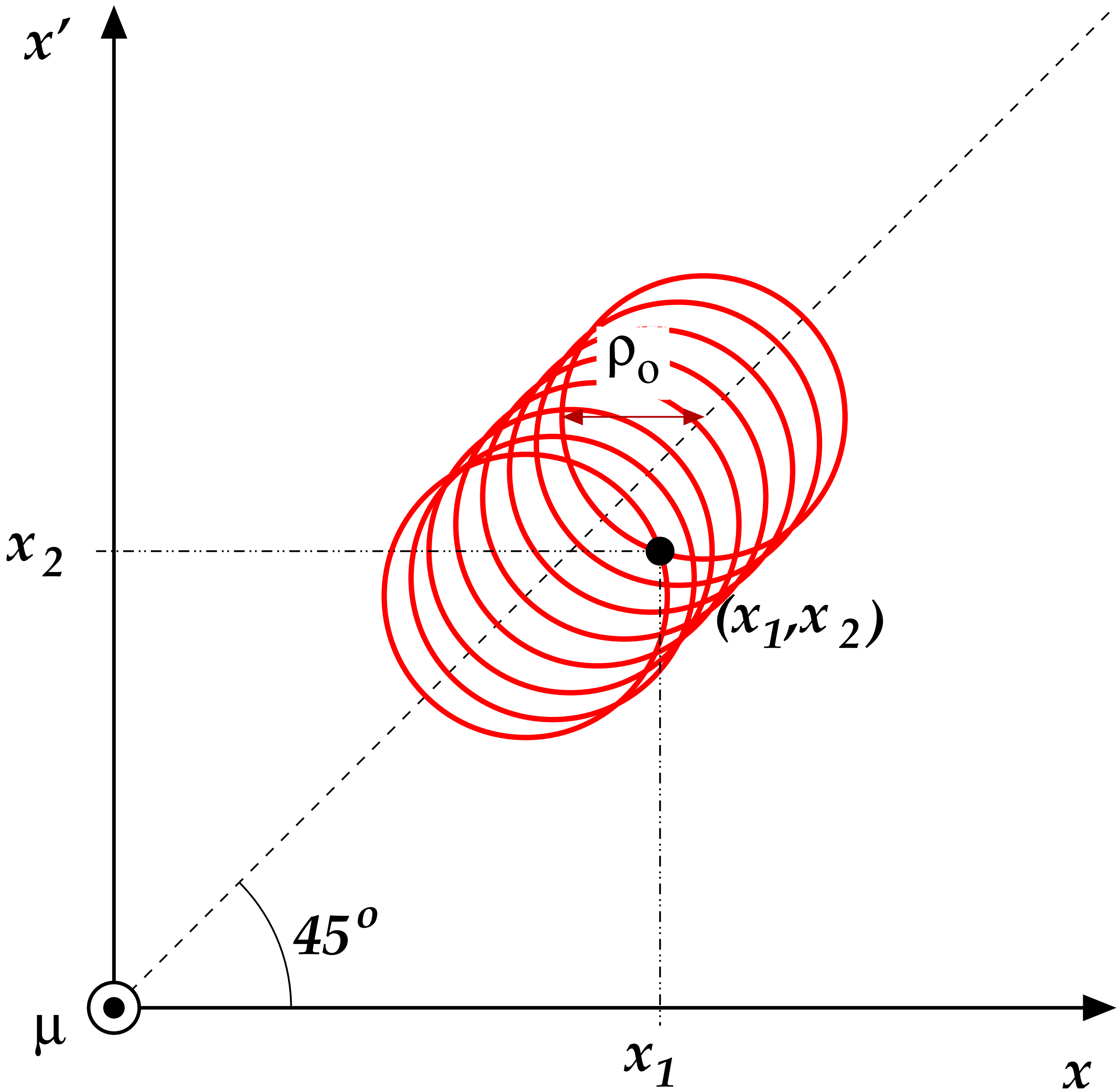
We determine for which values of μ'

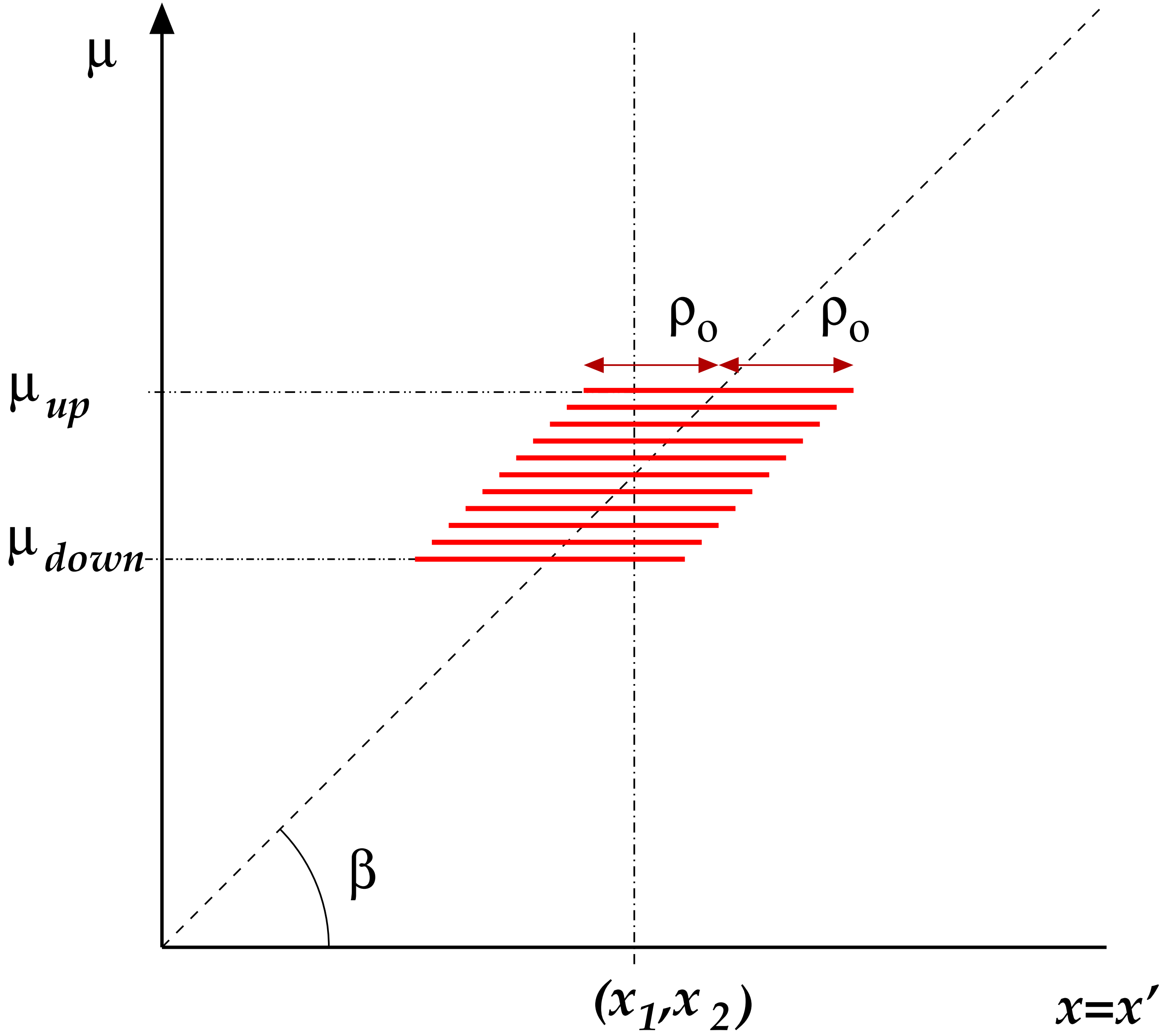
$$\int_{x,x':\,p(x,x'|\mu') \geq p(x_1,x_2|\mu')} p(x, x'|\mu')\,dx\,dx' \leq CL$$

Given CL we can obtain one of two results (Exercise: Prove it on the figure.):

1) a segment $(\mu_{down},\ \mu_{up})$ with the middle point equal to $(x_1 + x_2)/2$ (value of estimator!),

2) empty set!!! (in this case we can increase CL).





$$\tan\beta = \mu/(\mu^2+\mu^2)^{1/2} = 1/\sqrt{2}$$

Simple, lucky case:

both experiments have measured the same value: $x_1 = x_2$

Then, after substituting variables $x = \rho\cos\phi + \mu$ and $x' = \rho\sin\phi + \mu$, we easily calculate (and infer from figure) that with given CL we obtain $\mu_{down} = x_1 - \rho_0/\sqrt{2}$ and $\mu_{up} = x_1 + \rho_0/\sqrt{2}$ where

$$\rho_0 = \sigma_0\sqrt{-2\ln(1-CL)}$$

ρ_0	$\sigma_0/2$	$\sigma_0/\sqrt{2}$	σ_0	$\sqrt{2}\sigma_0$	$1.515\,\sigma_0$	$2\sigma_0$
$\rho_0/\sqrt{2}$	$\sigma_0/(2\sqrt{2})$	$\sigma_0/2$	$\sigma_0/\sqrt{2}$	σ_0	$1.515\,\sigma_0/\sqrt{2}$	$\sqrt{2}\sigma_0$
CL	0.118	0.221	0.393	0.632	0.683	0.865

With $CL = P(1\sigma)$ we obtain broader interval than for 1 measurement (around 1.07 times)! The length of the interval for given CL will increase with the number of measurements.

Weighted-average-based interval in *ab-ovo*-approach gives coverage with only $CL \approx 40\%$. One obtains limits $\mu_{down} = x_1 - \sigma_0/\sqrt{2}$ and $\mu_{up} = x_1 + \sigma_0/\sqrt{2}$ with $CL = P(1\sigma)$ if 1-dimensional p.d.f. of new random variable, $\bar{x} = (x + x')/2$, is used. (Exercise: Prove it.) But weighted average in the following case will work as well...

Simple, strange case:

second experiment measured value $x_2 = x_1 + 4\sigma_0$.

Then we have to choose $CL > CL_{min} = 1 - e^{-4} \approx 98\%$ to obtain non-zero segment of μ'! Probability of *better* result is $\geq CL_{min}$ for *ANY* value of μ.
The safe conclusion: we are CL_{min} confident that those two results are not compatible and further experiments are necessary. Probably in at least one of the experiments $p(x|\mu)$ is wrongly determined or we do not understand what μ is... Or we were very unlucky.

We reject with CL_{min} the hypothesis that these two experiments measured the same parameter μ.

Weighted-average method still gives you $\mu_{down} = x_1 + 2\sigma_0 - \sigma_0/\sqrt{2}$
and $\mu_{up} = x_1 + 2\sigma_0 + \sigma_0/\sqrt{2}$ with 1-dimensional confidence $CL = P(1\sigma)$... **:-)**
This is trustworthy result only if in the (x, x', μ) space CL_{min} is low!

Conclusion: It is necessary to know value and method of obtaining CL to judge such results as 10 ± 2 m.

Bayesian confidence interval

After first measurement x_1 we calculate *degree of belief* for μ:

$$p(\mu|x_1) = N p(x_1|\mu) p(\mu) \ ,$$

where N is a normalization constant,

$$N^{-1} = \int p(x_1|\mu) p(\mu) \, d\mu \ ,$$

and $p(\mu)$ is a *prior* p.d.f. describing our belief about true value of μ.

Let's assume that

$$p(\mu) = G(\mu|\mu_P; \sigma_P)$$

After simple calculation we obtain weighted-average-like result:

$$p(\mu|x_1) = G(\,\mu\,|\ \ (\sigma_P^2 x_1 + \sigma_0^2 \mu_P)/(\sigma_P^2 + \sigma_0^2);\ \ \sigma_P \sigma_0 / \sqrt{\sigma_P^2 + \sigma_0^2}\ )$$

At the beginning we assume very large σ_P in comparison to σ_0, *i.e.* we do not know where to expect μ. Then after one measurement:

$$p(\mu|x_1) = G(\mu\,|\,x_1;\ \sigma_0)$$

Before second measurement our degree of belief about μ is equal to $p(\mu) = p(\mu|x_1) = G(\mu \,|\, x_1;\, \sigma_0)$. After second measurement:

$$
\begin{aligned}
p(\mu|x_2, x_1) &= N' p(x_2|\mu)p(\mu) = N' p(x_2|\mu)p(\mu|x_1) = \\
&= G(\,\mu \,|\; (x_2 + x_1)/2;\;\; \sigma_0/\sqrt{2}\,)
\end{aligned}
$$

And from condition

$$\int_{\mu_{down}}^{\mu_{up}} p(\mu|x_2, x_1)\, d\mu = CL$$

one obtains limits (with additional, arbitrary requirement of symmetry)

$$
\begin{aligned}
\mu_{down} &= (x_1 + x_2)/2 - \sigma_0/\sqrt{2} \\
\mu_{up} &= (x_1 + x_2)/2 + \sigma_0/\sqrt{2}
\end{aligned}
$$

with $CL = P(1\sigma)$.

The same *precision* with the same CL will be reported for two very different situations: $x_2 = x_1$ and $x_2 = x_1 + 4\sigma_0$.

Of course if somebody is narrow-minded, then with $\sigma_P = \sigma_0/100$ will obtain:

$$\begin{aligned} p(\mu|x_1) &\approx G(\mu \,|\, \mu_P;\, \sigma_P) \\ p(\mu|x_2) &\approx G(\mu \,|\, \mu_P;\, \sigma_P) \end{aligned}$$

It is time to finish this picnic... Before the battle begins.

Conclusions

Report CL and method of calculating it if conventional interpretation is false.

Be open-minded if using Bayesian method.

With frequentist method you can determine with the same procedure if result is compatible with an assumed framework (CL_{min}). Not always results with high CL and narrow intervals are trustworthy.

Each method has arbitrary parameters and sub-procedures:

- frequentist: CL, ordering principle
- Bayesian: CL, ordering principle, prior p.d.f.

Usually both approaches are mixed and many approximations are involved.
Mainly we use just combinations of mean and dispersion estimators...

For your 'free' time...

Exercise: What is the result for cross section, σ, in all known to you approaches if for measured number of events, x_N, and measured luminosity, x_L, p.d.fs. $p(x_N|N)$ and $p(x_L|L)$ are known where N and L are true values of number of events and luminosity (in ideal world $N = L\sigma$).