SYSTEMATIC UNCERTAINTIES

- Goal is to address ~~answer~~ the following questions:
  - What are systematic uncertainties?
  - How do we correctly evaluate them?
  - How do we know if all systematics have been accounted for?
  - How do we treat correlations between systematic uncertainties?
  - How do we combine them?

~~In truth, no sy~~

In practice, no systematic and coherent treatment is available nor possible. We can only refer to knowledge, experience, common sense and intuition.

This is more a draft of a cookbook rather than a mathematical explanation. I will ~~give you~~ explain the main ~~cook~~ cooking ~~the~~ techniques and explain a few recipes, but then you have to explore on your own.

- Definitions of systematic uncertainties

~~a) Systematic~~

Possible definitions:

a) Systematic uncertainties are all uncertainties that are not directly due to the statistics of the data

↳ With this definition, the trigger efficiency or detector acceptance obtained from MC simulations would be systematic errors even though they have a statistical behaviour.

b) Systematic uncertainties are measurement errors which are not due to statistical fluctuations of real or simulated data samples

↳ In this case the trigger efficiency or detector acceptance would be treated as a statistical uncertainty.

$\boxed{\text{Syst 1}}$

- List of typical ~~syst~~ sources of systematics:
  - Badly known (geometrical) detector acceptance
  - " " trigger efficiency
  - reconstruction efficiency
  - detector calibration
  - detector resolution
  - Badly known background
  - Imperfect or overapproximated simulation (for speed reason)
  - Imperfect theoretical model implemented in simulations
  - Uncertainties in input parameters of the analysis (e.g. cross sections, branching ratios, lifetimes, luminosity, ...)
  - Computational and software errors
  - Personal biases towards a specific result

Everything else is an unknown unknown that affect the result to an unknown extent.

↳ Our goal is also to identify such unknown unknowns, or to quantify their relevance.

- Detection of possible sources of systematics

a) Top-down approach: Think of all possible sources of systematics, think of anything that could have gone wrong

↳ Experience and discussion with colleagues is crucial!

b) Bottom-up approach: scrutinously check your analysis for internal consistency by comparing data with (Toy) MC simulations, or by dividing data into non-overlapping subsamples and checking whether they are compatible with each other.

→ Usually, we use a combination of the two methods.

- Example: ~~Th~~ Space dependence of CUORE BI

~~Detector~~ Experiment made of ~1000 detectors.
Data taking divided in ~30 datasets of ~1 month duration.

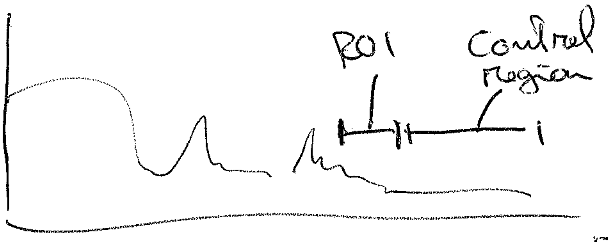
Question: is there any space dependence of the measured bkg rate?

↳ Is a single expectation value for the bkg rate enough to model the bkg rate of all 1000 detectors?

Method: 1) Compute average bkg rate

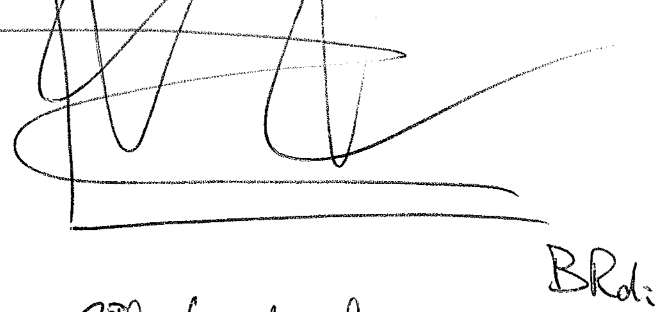
$$BR = \sum_{\text{detector } i} \sum_{\text{dataset } d} \frac{n_{di}}{m T_{di} \Delta E}$$

n_{di} = number of counts in control region
m = detector mass
T_{di} = dataset duration
ΔE = width of control region



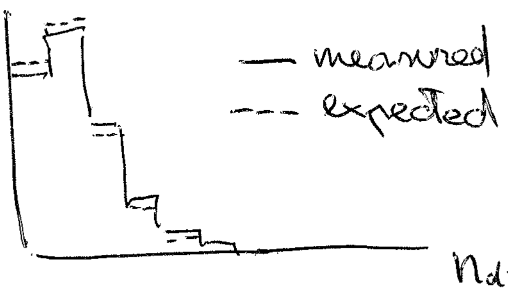
~~2) Produce distribution of bkg rate for each detector-dataset:~~

~~$BR_{di} = \frac{n_{di}}{m T_{di} \Delta E}$~~



2) Compare distribution of n_{di} ~~[illegible]~~ num (made of ~30 000 entries) with ~~Poisson distribution~~ combination of corresponding Poisson distributions, ~~all with~~ produced by setting

$$\lambda_{di} = BR \cdot m \cdot T_{di} \cdot \Delta E$$ with same BR for all!



Syst 3

4) If the two distributions agree we are happy, otherwise we have an indication ~~xx~~ of the existence of a space-dependence, ~~xx~~ and at that point we need to identify it.

↳ Split the detector in 2 ~~xx~~ subgroups based on their position within CUORE, and repeat the analysis. Good luck!

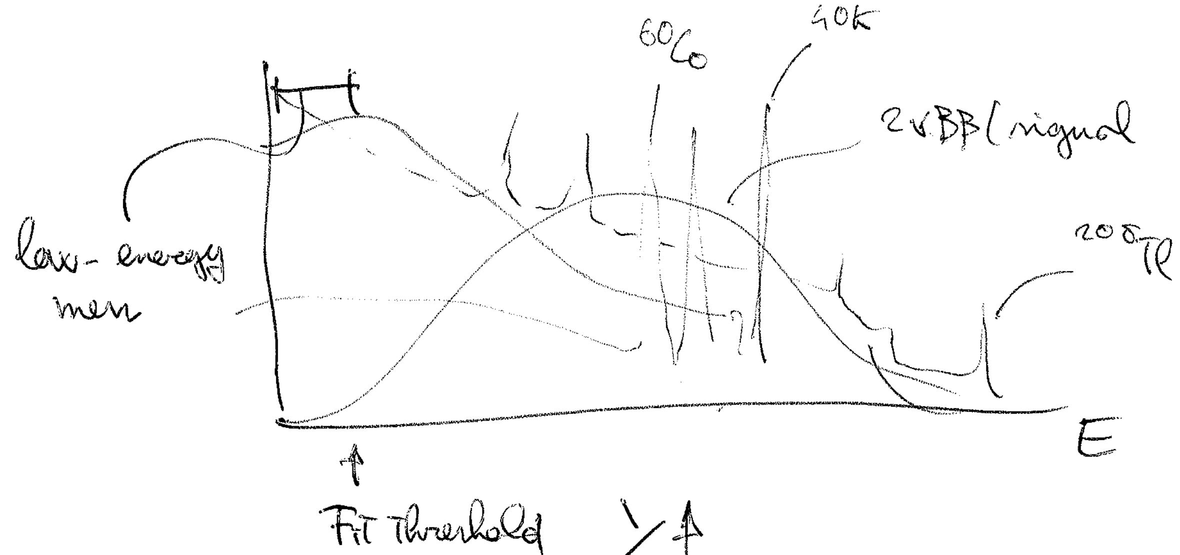
~~• Example: background systematic~~

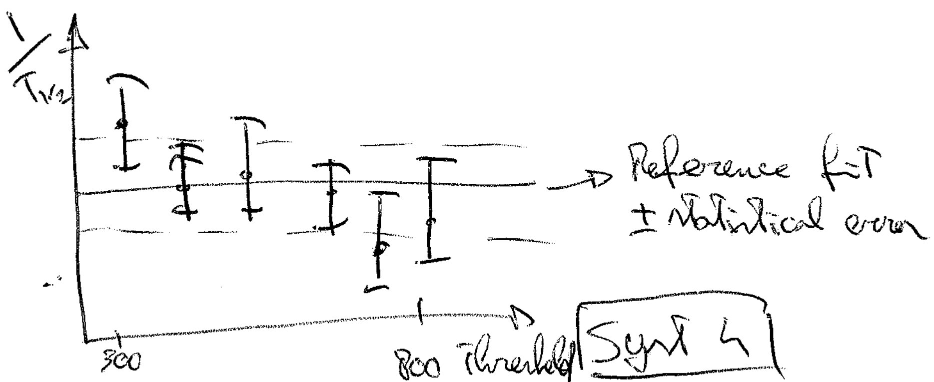
– Example: Evaluating the result in intervals of an analysis parameter

~~Rationale: the result stick~~

Rationale: ~~If~~ The result should be "independent" of the fit range of the parameters.

→ In the CUORE $2\nu\beta\beta$ decay measurement, we ~~xx~~ fitted the entire spectrum with a combination of MC-simulated spectra and extracted the $2\nu\beta\beta$ half-life. ~~Then we repeated~~
Since the low-Energy part was not well described, we varied the fit threshold and ~~xx~~ studied the effect on $T_{1/2}$

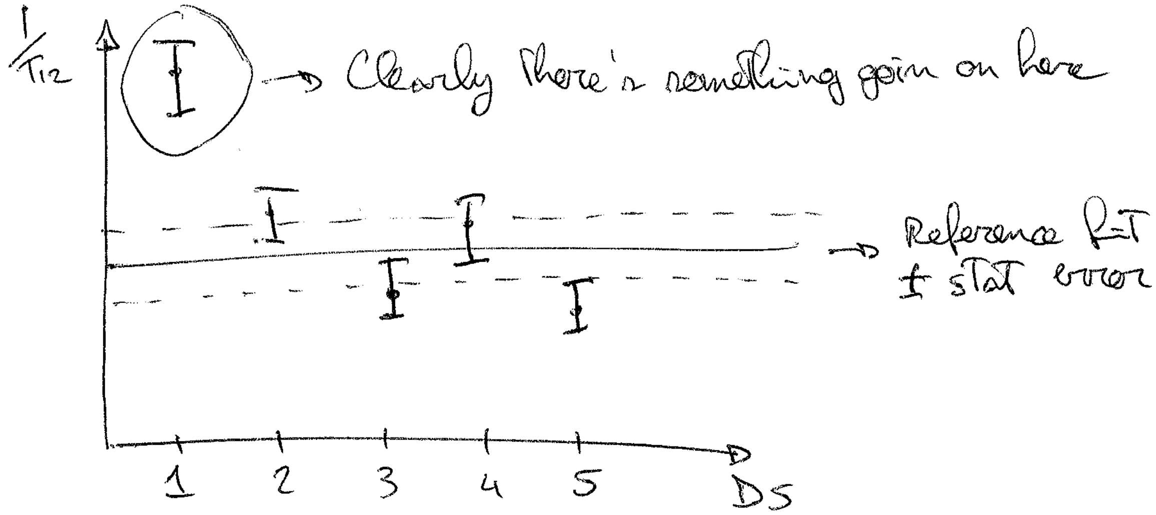




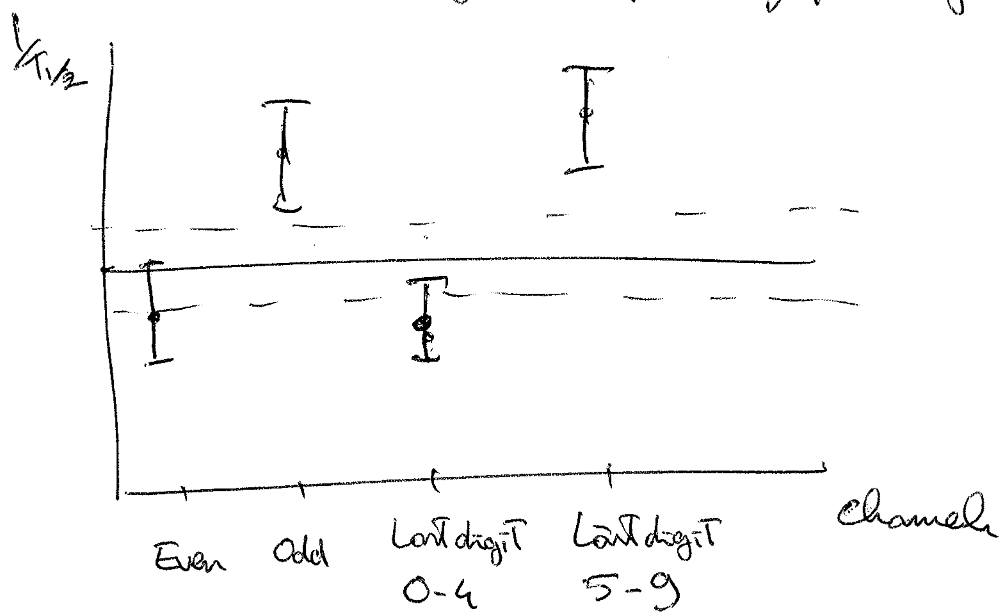
$\boxed{\text{Syst 4}}$

- Example: Splitting data into independent subsets

→ In CUORE, we repeat the fit for each DS separately, obtaining:



→ We also split the detector geometrically, with a non-sense splitting to ~~evaluat~~ evaluate the magnitude of ~~an~~ any possible geometric systematic:



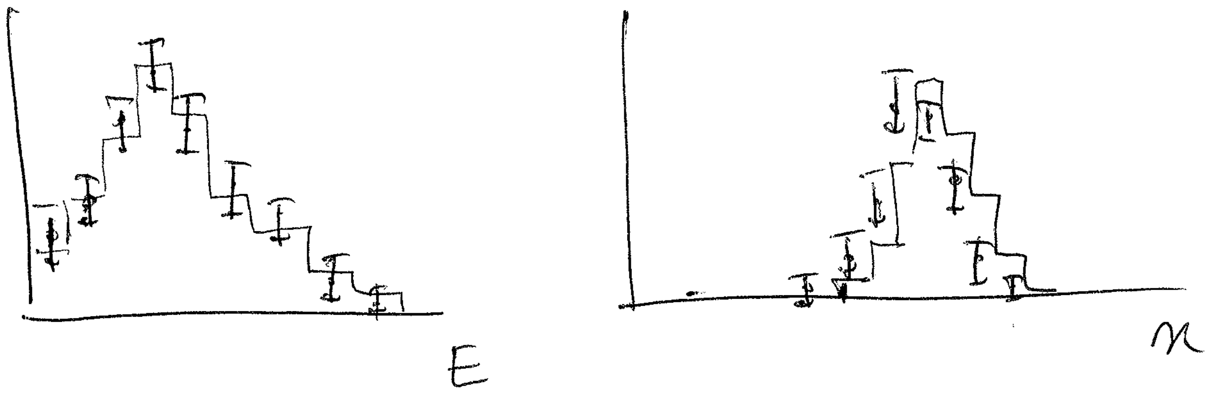
- Example: ~~Background systematics~~ Detector acceptance

Systematics ~~du~~ due to detector acceptance are generally a result of a poor MC description of the detector, in particular of detector inefficiencies (e.g. blind spots or dead layers) or misalignments.

↳ Think of dead layer in ~~Ge~~ germanium detector

↳ Think of blind spot in tracking detector

One trick is to compare the distribution of as many variables as possible. For example, if we ~~l~~ are interested in the energy deposition in a tracking detector, we might find an indication for the presence of a systematic in the distribution of the hit position, rather than in the ~~measured~~ energy spectrum.



- Example: Background systematics

- Example : Analysis software

Most analyses involve complex software developed ~~specifically~~ for the specific situation. It is crucial to make sure there is no bug!

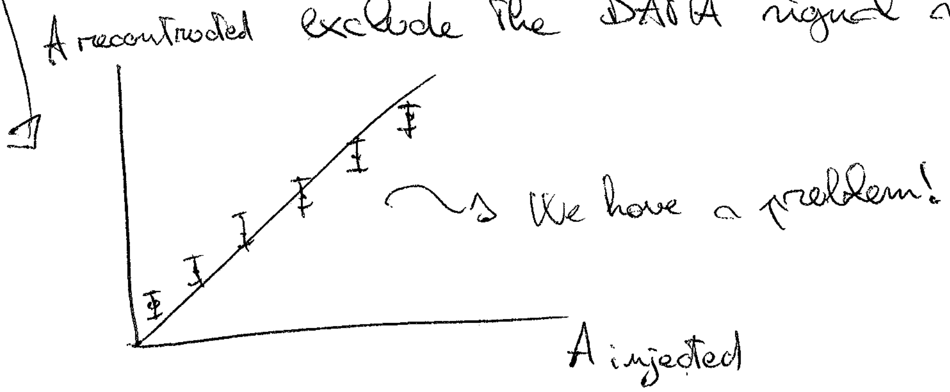
Solutions: 1) Perform a closure test using ~~artificially~~ data or Toy MC to make sure the code does what it is supposed to do!

2) Repeat the test varying the statistics of the fake data.

3) Study the fit bias as a function of the statistics, as a problem might arise only with very low or very high stat

4) Be critical of your result. Does it make sense from the Physics point of view?

↳ Don't make like your colleague who discovered Dark Matter on ~~dat~~ the same data that were used to exclude the DATA signal at 3σ!



Syst 7

- Quantification of systematic uncertainties

→ No general procedure, just many ways to perform valid estimations

→ If we really want to give a general approach, we want to:

1) Identify the source of systematic uncertainty
2) ~~Quantify~~ Vary the source by a "reasonable" amount
3) Quantify the variation on the result ~~[illegible]~~

The art lies in the definition of "reasonable".

- Example: External input parameters

If the analysis depends on an external input parameter ~~θ~~ with a known uncertainty σ_θ, we can just vary it by $\pm\sigma_\theta$ and repeat the analysis.

~~Take $y = f(x, \theta, \ldots)$~~
~~$\hat{x} =$ expectation value for x~~
~~Then: $\hat{y} = f$~~

Take: φ = parameter of interest $= f(\underset{\text{data}}{x}, \theta, \nu)$ → other nuisance parameters

$\hat{\theta}$ = expectation value of $\hat{\theta}$ from independent measurement

Then: $\hat{\varphi} = f(x, \hat{\theta}, \nu)$

$$\sigma^{+}_{\varphi,\text{syst}} = f(x, \hat{\theta} + \sigma_\theta, \nu)$$

$$\sigma^{-}_{\varphi,\text{syst}} = f(x, \hat{\theta} - \sigma_\theta, \nu)$$

↳ often times this can be obtained by a simple error propagation.

- ~~Example~~ Example: Tolerances

Suppose we don't know the standard deviation σ_θ of a parameter θ, but just an ~~minimum~~ allowed range $[\theta_{min}, \theta_{max}]$.
This would be the case of a hit on a scintillator strip.
~~The sd~~ For a uniform distribution, the standard deviation is:

$$\sigma_\theta = \frac{\sigma_{max} - \sigma_{min}}{\sqrt{12}} \simeq 0.29 (\sigma_{max} - \sigma_{min})$$

which is ~40% smaller than the naive half-range!

- Example: Small systematics

If a systematic effect gives - even in the worst case scenario - an effect that is much smaller than the statistical uncertainty, don't waste time in calculating the corresponding systematic error!

- Example: background estimation

Suppose we take the following data, consisting of a ROI with signal + background and some sidebands with background only:

We have 2 options:

a) Side band subtraction → does not rely on possibly incorrect MC, but does not predict ~~the~~ bkg shape in ROI

b) Model bkg with MC → does not provide bkg normalization

Solution: combine the 2 methods if ~~a~~ a) is not enough, meaning if the bkg shape cannot be parameterized.

$\boxed{\text{Syst 9}}$

– Example: Detector resolutions

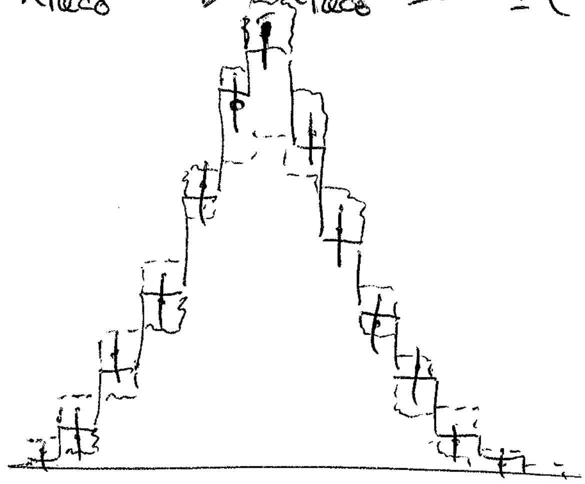
Usually, MC simulations are not good at reproducing resolutions, because they are the result of (strong) approximations.

Solution: evaluate the min and max mis-estimation of resolution on whichever observable x. These will result in a factors $(1+k_+)$ and $(1-k_-)$.

Then, the reconstructed values x_{reco} can be modified to:

$$x_{reco} \rightarrow x_{reco} \pm k_\pm (x_{reco} - x_{True})$$

↳ from simulation



– Example: Theory uncertainties

Suppose your result relies on some input from theory, which is computed using some model.

The theoretical input depends on:

a) The level of approximations used in the model (think of higher-order loops)

b) The parameterization used in the model itself (think of the various nuclear models)

Solution: Quote the spread between models, or ± the spread.
Educated guess helps a lot here!

↳ Example: $M_{\beta\beta}$ and NME

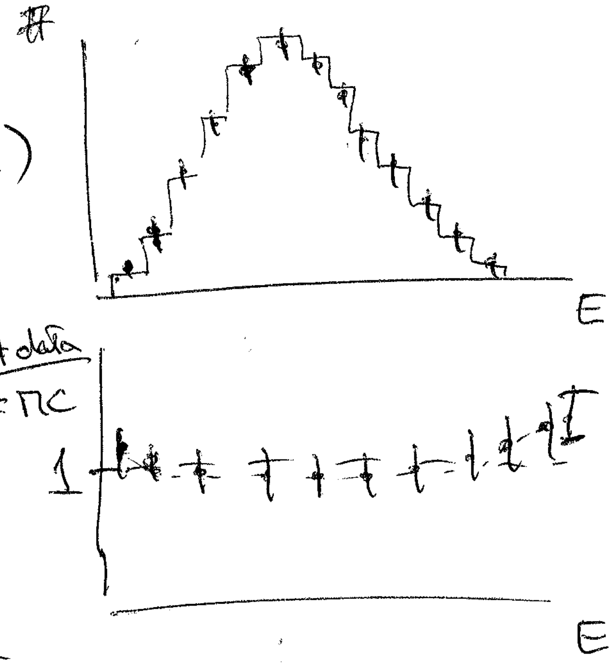
$$\frac{1}{T_{1/2}} = G \, g_A^4 \, M^2 \, \frac{m_{\beta\beta}^2}{m_e^2}$$

G = phase space → computed precisely
g_A = axial coupling → fixed
$m_{\beta\beta}$ = parameter of interest
m_e = electron mass
M = Nuclear Matrix Element

→ A single limit on $T_{1/2}$ is reflected in a "range of limits" on $m_{\beta\beta}$!

- Example: Discrepancy between data and simulation

Suppose that despite all efforts, we get stuck with a discrepancy between data and MC. How do we quantify the discrepancy?



Compute fraction of data ~~events~~ f laying in badly described region (10%)

Compute relative difference d between data and MC (e.g. 20%).

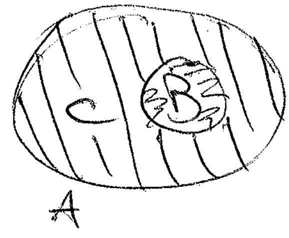
⇒ The systematic uncertainty will be $\pm f \cdot d = \pm 0.2\%$.

→ Notice that we are relying on the fact that the MC normalization in the "well described" region is correct!

- Example: CuT variations

Trick: split data in ~~non overl~~ disjoint subsamples

↳ Start from A and restrict to B



Data sample A → $x_A \pm \sigma_A$

↳ statistical uncertainty of A

Data sample B → $x_B \pm \sigma_B$ ⎤
C → $x_C \pm \sigma_C$ ⎦ → Statistically uncorrelated!

x_A ~~should~~ corresponds to the weighted average of x_B and x_C:

$$\bar{x} = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} = \frac{\frac{x_B}{\sigma_B^2} + \frac{x_C}{\sigma_C^2}}{\frac{1}{\sigma_B^2} + \frac{1}{\sigma_C^2}} = x_A$$

Knowing $x_A, x_B, \sigma_A, \sigma_B$, we can find ~~$x_C$~~ and σ_C:

$$x_C = \frac{\sigma_B^2 x_A - \sigma_A^2 x_B}{\sigma_B^2 - \sigma_A^2} \qquad \frac{1}{\sigma_C^2} = \frac{1}{\sigma_A^2} - \frac{1}{\sigma_B^2}$$

The difference between the uncorrelated x_B and x_C is:

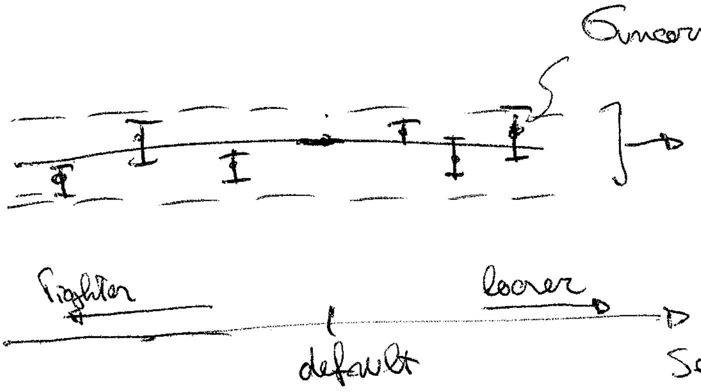
$$x_C - x_B = \frac{\sigma_B^2 x_A - \sigma_A^2 x_B}{\sigma_B^2 - \sigma_A^2} - \frac{(\sigma_B^2 - \sigma_A^2) x_B}{\sigma_B^2 - \sigma_A^2} = \sigma_B^2 \cdot \frac{x_A - x_B}{\sigma_B^2 - \sigma_A^2}$$

Normalizing by its STD:

$$\frac{x_C - x_B}{\sqrt{\sigma_C^2 + \sigma_B^2}} = \sigma_B^2 \cdot \frac{x_A - x_B}{\sigma_B^2 - \sigma_A^2} \cdot \frac{1}{\sqrt{\sigma_B^2 + \frac{\sigma_A^2 \sigma_B^2}{\sigma_A^2 - \sigma_B^2}}} = \frac{x_A - x_B}{\sqrt{\sigma_B^2 - \sigma_A^2}}$$

⇒ The uncorrelated error is: $\sigma_{uncorr}^2 = |\sigma_B^2 - \sigma_A^2|$

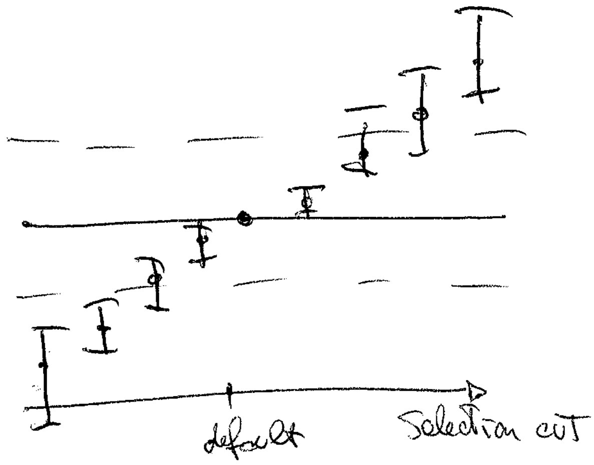
If we plot the uncorrelated error as a function of the cut variation:



→ Default ± stat error

Possible outcomes of cut variation:

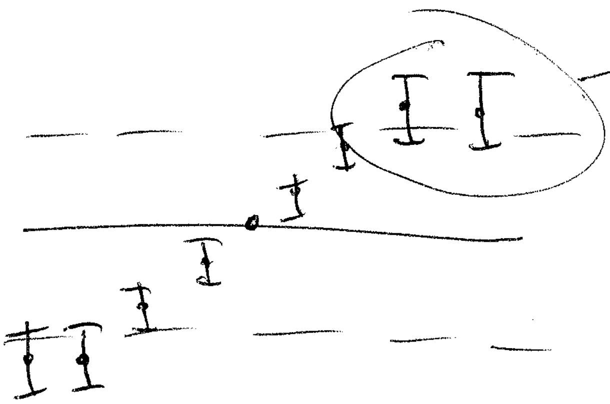
a)



→ Not possible to quote an uncertainty

→ We have a problem that we need to understand!

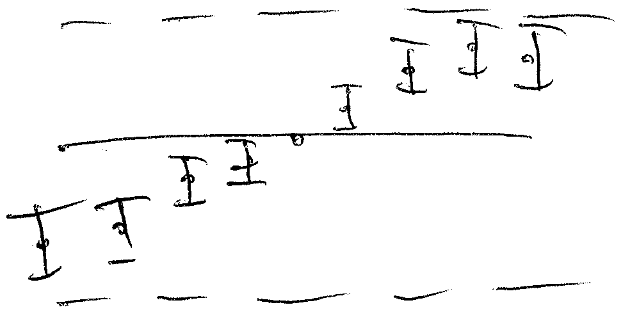
b)



→ ~~Do the~~ Does the trend really stabilize or is it due to a fluctuation?

If yes, quote the maximum variation, otherwise go back to a

c)



→ Quote max variation

→ If we mis-estimate the systematic, it's still smaller than the statistical uncertainty

- Combination of systematic uncertainties

→ If all systematic uncertainties are independent from each other, add them in quadrature:

$$\sigma_{TOT}^2 = \sum_{syst\ i=1}^{n} \sigma_{syst\ i}^2$$

→ If they are correlated, compute and account for the correlation:

$$\sigma_{TOT}^2 = \sum_i \sigma_i^2 + 2 \sum_{i<j} \rho_{ij} \sigma_i \sigma_j$$

Good luck with it!

- Avoiding ~~biases~~ human biases and blind analyses

1) Do not expect a certain result

Suppose you measure something that was already measured before

a) If your measurement agrees, you're happy

b) If you have a small ~~bu~~ but significant disagreement, you look for possible explanations of the discrepancy and try to make your result compatible with the previous one.
You will not look for effects making the discrepancy even bigger...

c) If the discrepancy is big ($>5\sigma$), you look for mistakes in both your and the original analysis...

2) Do not ~~tu~~ tune selection criteria using the same data used for the final measurement

⇒ Blind your data! (blind, not dumb!)