Errors of estimators

Suppose one has n observations of a random variable x, and a hypothesis for the pdf:

$$f(x|\theta)$$

depending on an unknown parameter θ that we want to estimate, e.g. via the Maximum likelihood

$$\hat{\theta}_{MLE}(\{x_i\}) \qquad i=1,\ldots,n$$

what is the uncertainty of such estimate?

we would like to report something like

$$\hat{\theta} = 4.85 \pm 0.41$$

if the experiment were repeated many times, this would be the std. of the distribution of the estimates

Some possibilitees:

1) Analytical estimates, if the variance of $\hat{\theta}$ is analytical (or even tractable):

$$V[\hat{\theta}] = \mathbb{E}[\hat{\theta}^2] - (\mathbb{E}[\hat{\theta}])^2$$

e.g. the radioactive decays of a particle are measured: $t_1, t_2, \ldots, t_n$. The proposed distrib. is the exponential: $f(t|\tau) = \frac{1}{\tau}\exp\{-t/\tau\}$

Thu it can be checked that

$$V[\hat{\tau}] = \tau^2/n \Rightarrow \hat{\tau}^2/n$$

2) Monte Carlo method:

Simulate a large # of experiments, compute for each of them the $\hat{\theta}_m$, make a histogram, and extract the std. This is also called the "Bootstrap method".

Example: (see notebook)

3) In many practical situations the Bootstrap can be costly (MC simulations are costly in general). So it is useful to have an alternative:

The Rao - Cramér - Frechet (RCF) bound; which in practice is estimated as:

$$\hat{\sigma}^2_{\hat{\theta}} \geq \left. \frac{-1}{\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}} \right|_{\theta = \hat{\theta}}$$

$$\log \mathcal{L} = \log \prod_{i=1}^{N} p(x_i|\theta) = \sum_i \log p(x_i|\theta)$$

assume $p(x_i|\theta) = \mathcal{N}(x_i|\theta, 1)$

$$\log p(x_i|\theta) = -(x_i - \theta)^2/2 - \log\sqrt{2\pi}$$

$$\frac{\partial \log p(x_i|\theta)}{\partial \theta} = +(x_i-\theta) \Bigg\} \frac{\partial^2 \log p(x_i|\theta)}{\partial \theta^2} = -1$$

$$\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} = \sum_i \frac{\partial^2 \log p(x_i|\theta)}{\partial \theta^2} = -N$$

$$\therefore \quad \hat{\sigma}^2_{\hat{\theta}} \geq 1/N \Rightarrow \frac{1}{\sqrt{N}}$$ trend of statistical errors

4) Graphical method (cf. pg 78 Cowan)

- In the large sample limit, the pdf of the estimator $\hat{\theta}$ will be Gaussian thanks to the CLT.
  However, if this is not the case, how to proceed when reporting the uncertainties?

  ↳ A priori no problem with the above procedures, but it is not the conventional way to express it

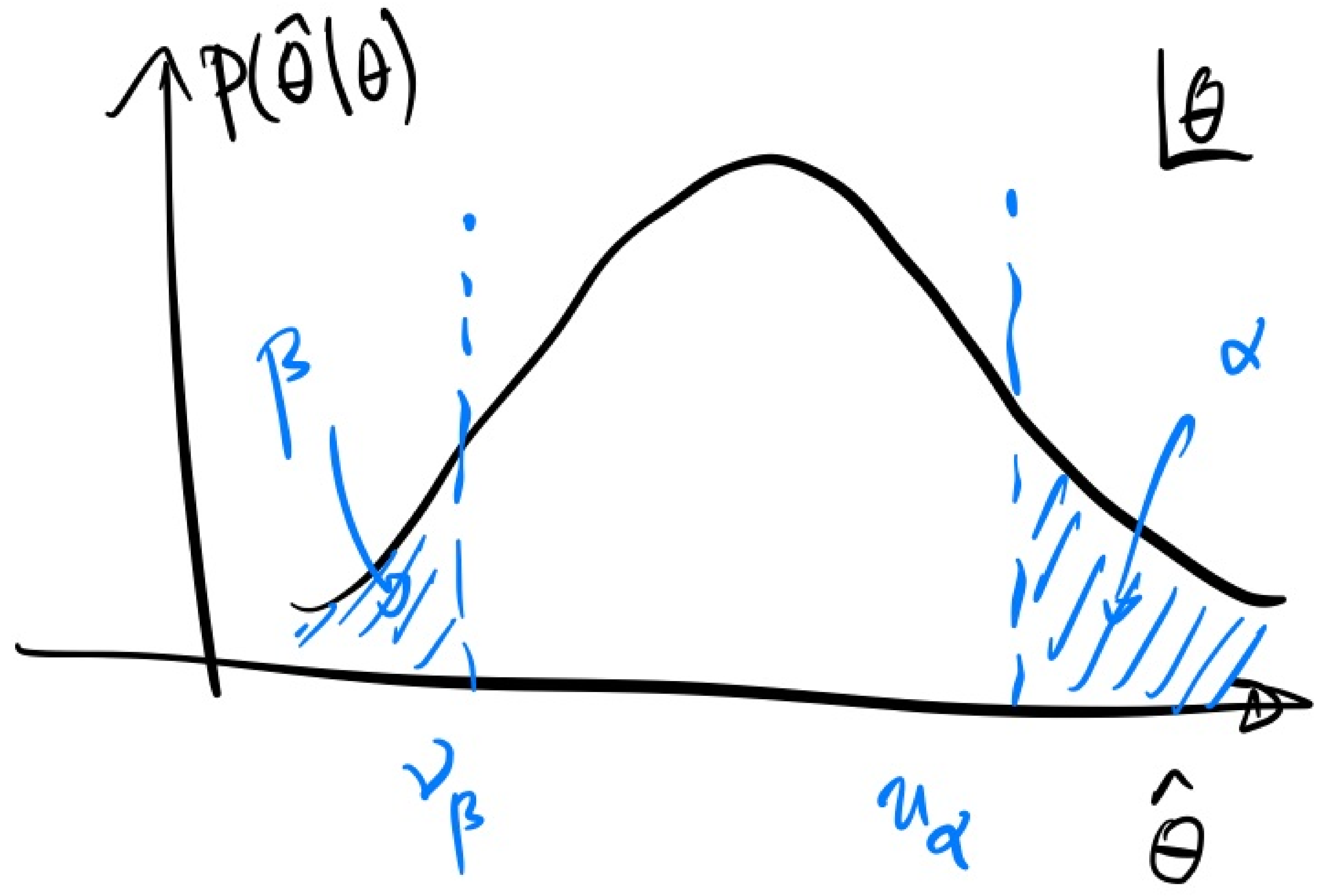
Confidence intervals

One has a series of observations $\{x_i\}_{i=1}^{N}$, following an assumed pdf $p(x|\theta)$, and we want to estimate θ.

Note that the distrib. of $\hat{\theta}_{MLE}$ will depend on the true parameter θ (since it determines the outcome of any possible experiment)

So the pdf of $\hat{\theta}$ is $p(\hat{\theta}|\theta)$

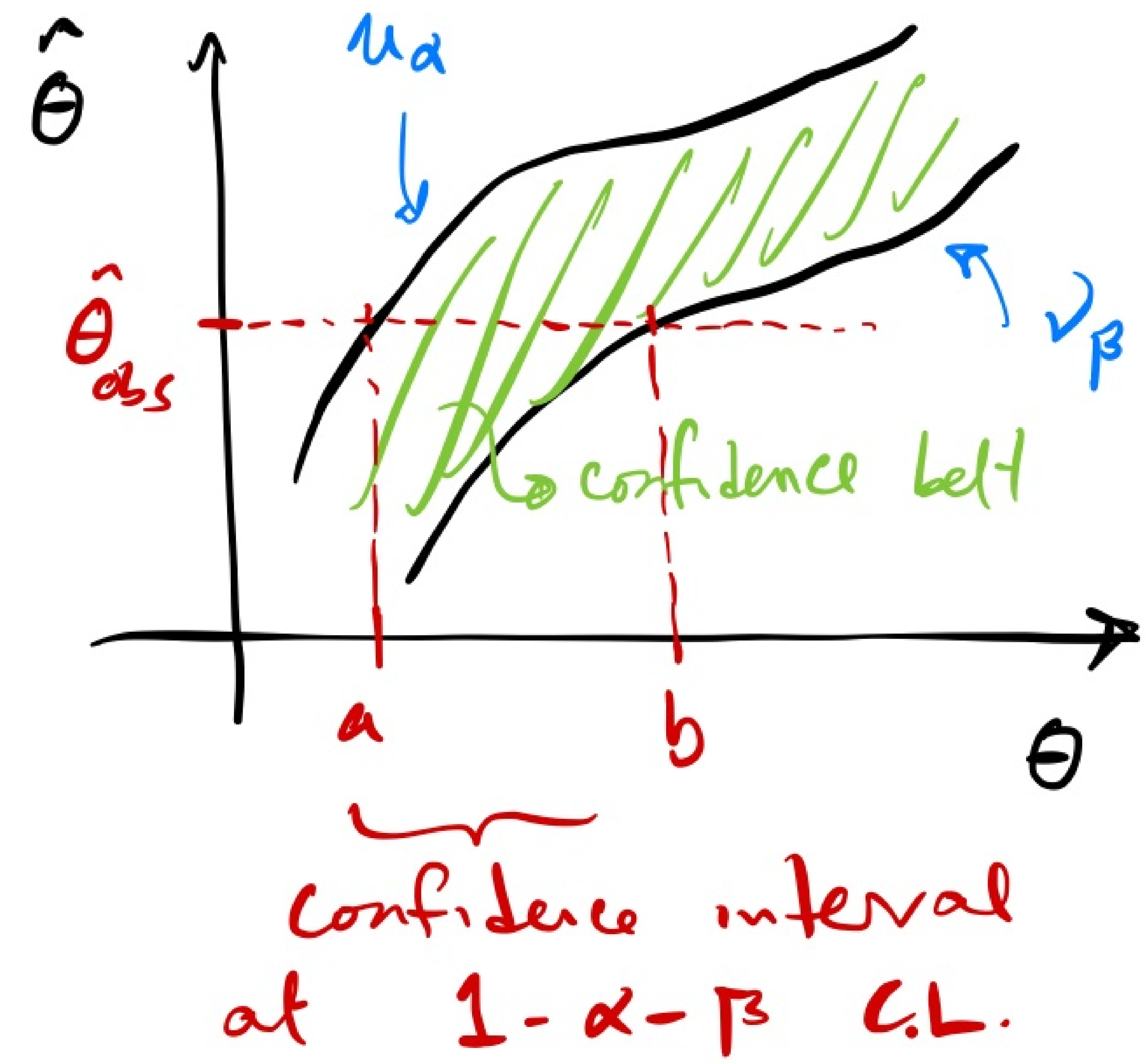
Se pueden determinar 2 valores u_α y v_β tal que



$$P(\hat\theta \geq u_\alpha(\theta)) = \int_{u_\alpha}^{\infty} p(\hat\theta|\theta)\, d\hat\theta = \alpha$$

$$P(\hat\theta \leq v_\beta(\theta)) = \int_{-\infty}^{v_\beta} p(\hat\theta|\theta)\, d\hat\theta = \beta$$

If $\hat\theta$ is a good estimator of θ, then it should happen that:



Then $\hat\theta_{MLE}$ observed will determine

$$a \equiv u_\alpha^{-1}(\hat\theta_{obs})$$

$$b \equiv v_\beta^{-1}(\hat\theta_{obs})$$

This is Neyman procedure

Equivalently:

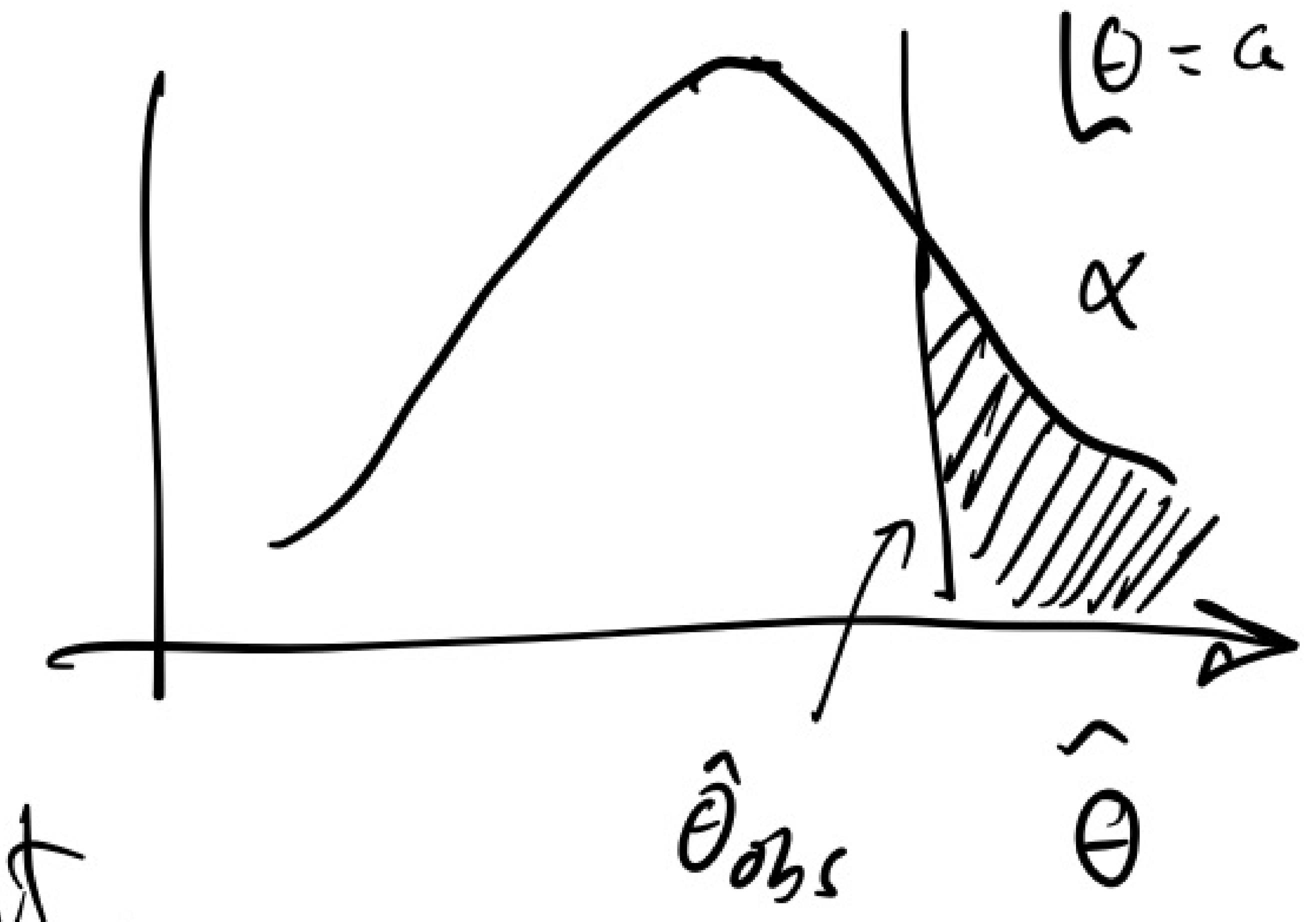
$$\left.\begin{aligned} \alpha &= \int_{\hat\theta_{obs}}^{\infty} p(\hat\theta|a)\, d\hat\theta = 1 - CDF(\hat\theta_{obs}|a) \\ \beta &= \int_{-\infty}^{\hat\theta_{obs}} p(\hat\theta|b)\, d\hat\theta = CDF(\hat\theta_{obs}|b) \end{aligned}\right\} \text{(I)}$$

to be solved e.g. numerically

of course, one-sided (limit) is also possible

So α is the fraction of times that we would get a value for $\hat{\theta} \geq \hat{\theta}_{obs}$, if the value of the true parameter $\theta = a$

Analogously for β, if $\theta = b$



- Let's look at a Poisson example

Imagine having a single measurement, giving n_{obs} (# of obs events)

What is the lower limit on λ for $(1-\alpha)\%$ CL?

$P(\lambda \geq \lambda_\alpha) = \alpha$

$$\alpha = \sum_{n=n_{obs}}^{\infty} P(n|\lambda_{low}) = 1 - \sum_{n=0}^{n_{obs}-1} P(n|\lambda_{low})$$

And the upper limit for $(1-\beta)\%$ CL?

$$\beta = \sum_{n=0}^{n_{obs}} P(n|\lambda_{up})$$

(see Notebook) see also Table 9.3 of Cowan

• Case where estimator is Gaussian ↳ common situation thanks to the CLT

$$p(\hat{\theta}|\theta) = N(\hat{\theta}|\theta, \sigma)$$

↳ assumed to be known

The cumulative distribution is

$$\Phi(\hat{\theta}|\theta,\sigma) = \int_{-\infty}^{\hat{\theta}} N(\hat{\theta}|\theta,\sigma)\, d\hat{\theta}$$

Then (I) can be expressed as:

standard Gaussian

$$\alpha = 1 - \Phi(\hat{\theta}|a,\sigma) = 1 - N_s\left(\frac{\hat{\theta}-a}{\sigma}\right)$$

$$\beta = \Phi(\hat{\theta}|b,\sigma) = N_s\left(\frac{\hat{\theta}-b}{\sigma}\right)$$

Solving for a & b:

$$N_s\left(\frac{\hat{\theta}-a}{\sigma}\right) = 1-\alpha \Rightarrow \frac{\hat{\theta}-a}{\sigma} = N_s^{-1}(1-\alpha)$$

$$a = \hat{\theta} - \underline{\underline{\sigma}} \cdot N_s^{-1}(1-\alpha)$$

σ can be replaced by $\sigma_{\hat{\theta}}$, i.e, its

$$N_s^{-1} \cdot \beta = \frac{\hat{\theta}-b}{\sigma} \Rightarrow b = \hat{\theta} - \sigma \cdot N_s^{-1}(\beta)$$

$$= \hat{\theta} + \underline{\underline{\sigma}}\, N_s^{-1}(1-\beta)$$

For central intervals, $\alpha = \beta = \gamma/2$ the quantiles $N_s^{-1}(1-\gamma/2)$ are tabulated corresponding to the C.L. = $(1-\gamma)\%$

e.g

$N_s^{-1}(1-\gamma/2)$	$1-\gamma$
1	0.6827
2	0.9544
3	0.9973

Confidence intervals using the Likelihood

Quite used in practical, even when the estimator is not Gaussian.

Consider a single parameter θ, and expand the log Likelihood in Taylor about its MLE, $\hat{\theta}$:

$$\log\mathcal{L}(\theta) = \log\mathcal{L}(\hat{\theta}) + \cancel{\left[\frac{\partial \log\mathcal{L}}{\partial\theta}\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta})}$$

zero by definition

$$+ \frac{1}{2}\left[\frac{\partial^2 \log\mathcal{L}}{\partial\theta^2}\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta})^2 + \dots$$

using the RCF bound:

$$\left.\frac{\partial^2 \log\mathcal{L}}{\partial\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{1}{\hat{\sigma}_{\hat{\theta}}^2}$$

equality holds in the large sample limit

$$= \log\mathcal{L}(\hat{\theta}) - \frac{(\theta-\hat{\theta})^2}{2\hat{\sigma}_{\hat{\theta}}^2}$$

$$\therefore \log\mathcal{L}(\hat{\theta} \pm N\hat{\sigma}_{\hat{\theta}}) = \log\mathcal{L}_{max} - N^2/2$$

Then this is true for Gaussian likelihood

however the method is used even if Likelihood is not Gaussian:

$$\log\mathcal{L}(\hat{\theta}^{+d}_{-c}) = \log\mathcal{L}_{max} - N^2/2$$

with $N = N_s^{-1}(1-\gamma/2)$ the quantile of the std Gaussian for a $CL = (1-\gamma)\%$

"likelihood interval method"

(see Notebook)

- Confidence intervals using the likelihood ratio

Consider some data D distributed according to a likelihood function having several parameters

$$\mathcal{L}(\theta_1, \theta_2, \ldots, \theta_M)$$

some of which are of interest $(\theta_1, \ldots, \theta_k)$ and some others are nuissance $(\theta_{k+1}, \ldots, \theta_M)$

we would like to create a TS for the params of interest, and let it be the likelihood ratio

$$\Lambda(\theta_1, \ldots, \theta_k \mid D) = -2 \ln \frac{\mathcal{L}(\theta_1, \ldots, \theta_k, \hat{\theta}_{k+1}, \ldots, \hat{\theta}_M)}{\mathcal{L}(\hat{\theta}_1, \ldots, \hat{\theta}_M)}$$

→ profiled likelihood

$\hat{\theta}_i$: MLE

This k-dimensional confidence interval may be difficult to obtain in general (e.g. via bootstrapping)

However, in many situations one can apply the following

Wilks theorem

the likelihood ratio $\Lambda(\theta_1, \ldots, \theta_k \mid D)$ follows, in the large sample limit, a chi-squared distrib with k d.o.fs

i.e
$$\Lambda(\theta_1, \ldots, \theta_k \mid D) \sim \chi^2_k$$

Remember: χ^2_k-distrib. is the distrib of the sum of k indep. std normal variables

i.e. $$\sum_{j=1}^{k} z_j \sim \chi^2_k$$

$$\chi^2_k(x) = \frac{1}{2^{k/2}\,\Gamma(k/2)}\, x^{k/2-1}\, e^{-x/2}$$

Then, the only thing to do is to look it up at tables!

e.g.

Consider some data, consisting of Poisson counts $\{n_1, n_2, \ldots, n_N\}$
(each component here could represent, e.g. different energy bins, or different spatial locations, ...)

Then the prob. of observing n_i is: $\Rightarrow$ $P(n_i \mid \lambda_i(s,b))$

s_{true}: unknown

here $\lambda_i = f_i^s \cdot s + f_i^b \cdot b$ are the means of the Poisson, and they are all a function of 2 type of parameters "s" and "b", where

"s": parameter of interest (signal)

$\{b\}$: nuisance parameter (bckg)

In addition, consider a bckg distributed according to the distribution:

$$b \sim N(b \mid B_i, \sigma_B)$$

assume it is known

B: true value, mean of distrib.

with $B_i = f_i^b \cdot \underbrace{b_{true}}$ unknown

Then the likelihood of the data is

$$\mathcal{L}(s,b) = \prod_{i=1}^{N} P(n_i \mid \lambda_i(s,b))\, N(b \mid f_i^b \cdot b, \sigma_B)$$

We would like to estimate s_{true} and confidence interval (and also upper limit) at 90% CL

see notebook

- <u>Limits near a physical boundary</u>

What if what one estimates is a parameter with a physical boundary, e.g. θ=mass ≥ 0 ?

It may happen that $\hat{\theta} < 0$ if the estimator is obtained as $\hat{\theta} = \underbrace{x - y}$ Gaussian variables

i.e. not positive-definite

Let $\hat{\theta}_{obs} = -2$ and $\sigma_{\hat{\theta}} = 1$

take a normally distributed estimator s.t.

$$\theta_{up} = \hat{\theta}_{obs} + \sigma_{\hat{\theta}}\, \Phi^{-1}(1-\beta)$$

for a 95% CL upper limit, $\Phi^{-1}(0.95) = 1.645$

then $\theta_{up} = -2 + 1.645 = \underbrace{-0.355}$

$< 0 \Rightarrow$ unphysical

* Two alternatives

  - Redefine θ_{up} : $\theta_{up} = \max(\hat{\theta}_{obs}, 0) + \sigma_{\hat{\theta}} \Phi^{-1}(1-\beta)$

    ↳ handy but it doesn't have the correct coverage of $(1-\beta)$% CL

  - Go Bayesian (see next section)

- Interval or upper limit? The flip-flopping policy

Imagine that you have your model of S+B and the # of expected bckg events is n_B (n_S) for the backg (signal). Upon measurement one gets n_{obs}.

e.g. less than 3σ apart

Typically, if $n_B \simeq n_{obs}$, one decides to put upper bound on n_S, otherwise if

$$n_B \gtrsim n_{obs} + 3\sigma$$

then one decides to quote an interval

"flip-flopping" policy

More concretely: Let x be the measured quantity, such that if $x \leq 3\sigma$ we quote a limit on μ, and an interval otherwise.

Let's work at 90% CL, and take $\sigma = 1$. Then, the confidence belt for μ is:

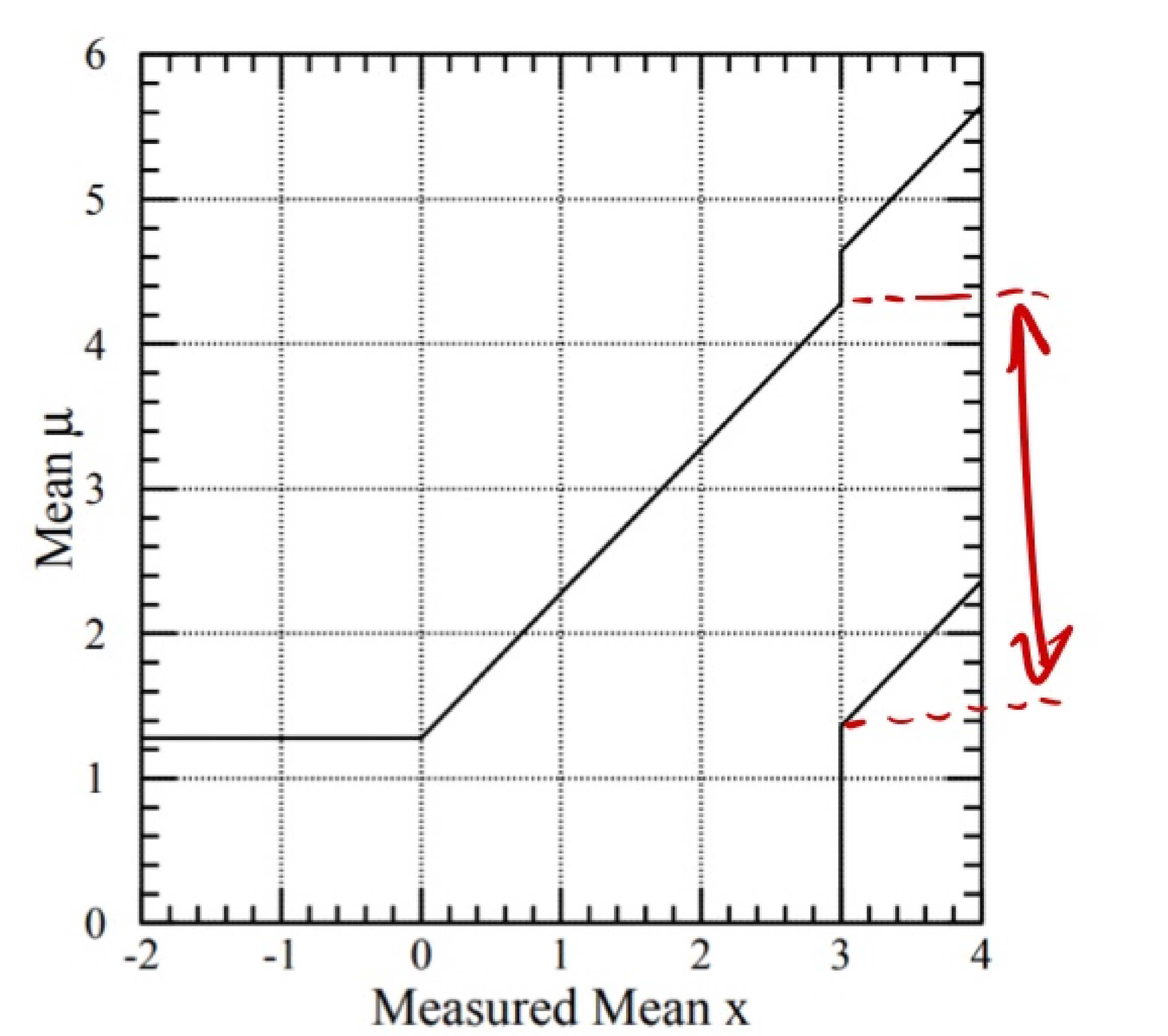


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence level. For $1.36 < \mu < 4.28$, the coverage (probability contained in the horizontal acceptance interval) is 85%.

* Problem:

Not correct coverage (85% instead of 90%)

- Feldman – Cousins intervals

The above problems are encountered in a Poisson example, with known bckg mean,

say $b = 3$ and $n = 0$

what is the upper limit @ 90% CL on signal s ?

$$\beta = 0.1 = \sum_{n=0}^{n_{obs}} P(n \mid s^{up} + b) = P(0 \mid s^{up} + 3)$$

$$0.1 = e^{-s-3} \Rightarrow -s^{up} - 3 = \ln(0.1)$$

$$s^{up} = -\ln(0.1) - 3$$

$$= 2.3 - 3 = -0.7$$

$s^{up} = -0.7 \Rightarrow$ impossible

$\Rightarrow$ empty interval !

F-C proposal : Build the Confidence belt as the region with the largest maximum likelihood ratio

* For a given s, build

$$R(n) \equiv \frac{P(n \mid s+b)}{P(n \mid \hat{s}+b)} \quad , \quad \hat{s} \equiv \max(0, n-b)$$

this is the MLE with a physical constraint

* where n is varied in order to construct the confidence belt, s.t. we progressively add the values of n corresponding to the largest $R(n)$, until the band covers at least 90% CL of the Poisson distrib.

(see Notebook)

- ## (Bayesian) Credible intervals

This makes use of the posterior distrib.

$$p(\theta \mid \text{data})$$

s.t. an interval $\theta \in [\theta_{lo}, \theta_{up}]$ is determined as

$$\boxed{1-\beta = \int_{\theta_{lo}}^{\theta_{up}} p(\theta \mid \text{data})\, d\theta}$$

Interpretation here is natural: It is the region where the true value of the parameter is contained with certain probability (degree of believe)

as opposed to:

If the experiment were to be repeated a large # of times, the would vary, covering the true value of the parameter a fraction $1-\alpha$ of times

For an upper limit for example:

$$1-\beta = \int_{-\infty}^{\theta_{up}} \underbrace{p(\theta|\vec{x})}_{\text{posteror}}\, d\theta = \frac{\int_{-\infty}^{\theta_{up}} \mathcal{L}(\vec{x}|\theta)\, \pi(\theta)\, d\theta}{\int_{-\infty}^{\infty} \mathcal{L}(\vec{x}|\theta)\, \pi(\theta)\, d\theta}$$

where the prior $\pi(\theta)$ can be e.g.

$$\pi(\theta) = \begin{cases} 0 & \text{if } \theta < 0 \\ 1 & \text{if } \theta \geq 0 \end{cases}$$

provides solution for implementing physical boundary

This has the drawbacks of being prior dependent, but ok, life is painful