INTERVAL ESTIMATION

- Goal. Confidence interval.

In interval estimation, we want to find the range $\vartheta_a \le \vartheta \le \vartheta_b$ which contains the true value ϑ_0 with probability β.

Such interval is called "confidence interval" with probability content β.

Typically, we choose $\beta = 68.3\%$ and call it 1 standard deviation error. However, the 68% interval corresponds to ± 1 STD only for a Gaussian distr.

Given an observation x from a PDF $f(x|\vartheta)$, the probability content β of the ~~range~~ region $[a, b]$ in <u>x-space</u> is:

$$\beta = P(a \le x \le b) = \int_a^b f(x|\vartheta)\, dx$$

If $f(x|\vartheta)$ and the parameter ϑ are known, one can always compute β given a and b.

If the parameter ~~ϑ~~ ϑ is unknown, we need to find another variable

$$z = z(x, \vartheta)$$

such that the PDF of z is independent of ϑ: ~~$f(z|\vartheta) = f($~~

$$f(z|x, \vartheta) = f(z|x)$$

If this can be found, we can find the optimal range $[\vartheta_a, \vartheta_b]$ <u>in ϑ space</u> such that:

$$P(\vartheta_a < \vartheta < \vartheta_b) = \beta$$

This interval $[\vartheta_a, \vartheta_b]$ is called "confidence interval".

A method which yields ~~an~~ such an interval $[\vartheta_a, \vartheta_b)$ is said to possess the property of <u>coverage</u>.

Notice that: $\rightarrow$ ϑ_0 is an unknown constant

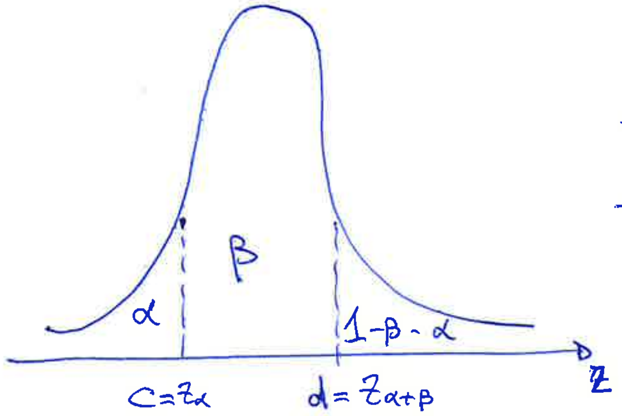
$\rightarrow$ ϑ_a and ϑ_b are functions of x, not of ϑ.

- Confidence intervals for the mean of a Gaussian

For any Gaussian, we can re-define: $z = \frac{x-\mu}{\sigma}$

~~and at~~ which is a standard-normal variable: $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$

Estimating the interval $[c, d]$ so that $P(c \le z \le d) = \beta$ is equivalent to finding $[z_\alpha, z_{\alpha+\beta}]$:



→ There's infinite choices of the interval!

→ A standard choice is the central interval symmetric around zero, so that:

$$\alpha = \frac{1-\beta}{2}$$

$\beta = \frac{1-\alpha}{2}$	z_α	$z_{\alpha+\beta}$	
0.6827	-1	$+1$	$\to \pm 1\sigma$
0.9	-1.65	$+1.65$	
0.95	-1.96	$+1.96$	
0.9545	-2	$+2$	$\to \pm 2\sigma$
0.9973	-3	$+3$	$\to \pm 3\sigma$

- Confidence intervals for several parameters

Suppose we have an n-dimensional Gaussian:

$$f(\vec{x}\,|\,\vec{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|C|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\theta})^T C^{-1} (\vec{x}-\vec{\theta})\right)$$

Each x_i is normal, therefore $Q(\vec{x},\vec{\theta}) = (\vec{x}-\vec{\theta})^T C^{-1} (\vec{x}-\vec{\theta})$ is a $\chi^2(n)$ distribution, and does not depend on $\vec{\theta}$:

$$Q(\vec{x},\vec{\theta}) = Q(\vec{x})$$

- Second derivative matrix

Assume $\vec{x}$ has an n-dim Gaussian PDF.

One can prove that the n-dim covariance matrix C can be obtained from the inverse of the 2^nd order partial derivative matrix of $-\ln\mathcal{L}$:

$$C^{-1}_{ij} = -\frac{\partial \ln \mathcal{L}(\vec{x}|\vec{\vartheta})}{\partial\vartheta_i\,\partial\vartheta_j}$$

This covariance matrix gives an n-dim elliptic contour with the correct coverage only if the PDF is exactly Gaussian!

This is the "standard" classical method used to compute uncertainties in common fitting algorithms, e.g. Migrad/Hesse of Minuit/ROOT.

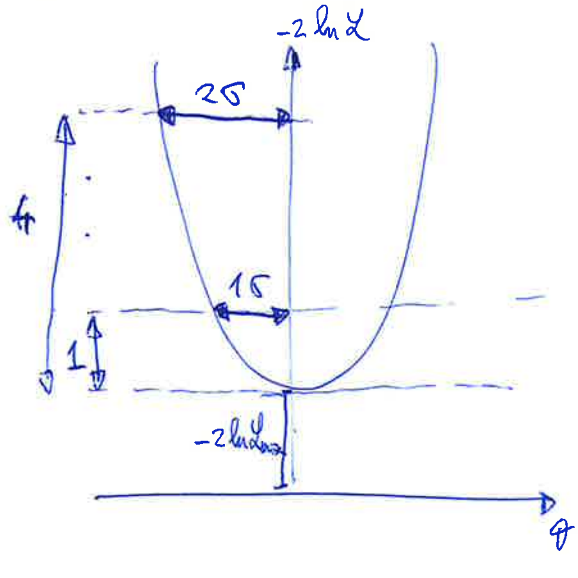
- Log-Likelihood scan

Another common method consists in taking a scan of $-2\ln\mathcal{L}$ around its minimum value, $-2\ln\mathcal{L}_{max}$.

→ For a Gaussian 1-dim distribution:

$$\ln\mathcal{L}(x|\mu) = \ln c - \frac{(\mu-x)^2}{2\sigma^2} \quad \Rightarrow \text{parabola in } \mu$$

$$-2\ln\mathcal{L} = -2\ln\mathcal{L}_{max} + \frac{(\mu-x)^2}{\sigma^2}$$



The intercept at $-2\ln\mathcal{L} = -2\ln\mathcal{L}_{max} + 1$ provides the $\pm 1\sigma$ interval.

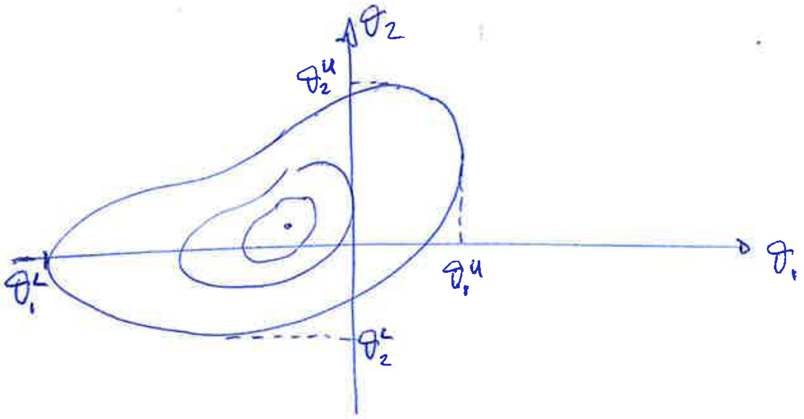
The intercept at $+4$ provides the $\pm 2\sigma$, and so on

- $\mathcal{L}$ scan in dim > 1 : Profile likelihood

For dim > 1 we have: $\ln \mathcal{L}_{\vec{\theta}}(\vec{x}|\vec{\theta}) = \ln \mathcal{L}_{max} - \frac{1}{2}\chi^2_\beta(\cancel{H})$

$$\hookrightarrow -2\ln \mathcal{L}(\vec{x}|\vec{\theta}) = -2\ln \mathcal{L}_{max} + \chi^2_\beta(\cancel{H})$$

In principle, we can compute the contours:



Notice that: → The inner contour is more nearly elliptical than the outer ones

→ The coverage is improved with respect to the Gaussian approximation

→ This is still an approximation valid for large n.

- Variance of transformed variables, aka Error propagation

Suppose we have a variable $\vec{x}$ with a given PDF$_x$ and mean $\vec{\mu}$ an variance $V(\vec{x})$

Suppose we want to compute the variance of the transformed variable $\vec{y} = \vec{y}(\vec{x})$,

and that the function can be expanded in Taylor series around $\vec{\mu}$:

$$\vec{y}(\vec{x}) = \vec{y}(\vec{\mu}) + \sum_i (x_i - \mu_i)\frac{\partial y}{\partial \mu_i} + \ldots$$

The expectation value of y is: $\bar{y} = y(\vec{\mu})$

The variance is: $V(y) = E[y - E(y)]^2$

$$\simeq E\left[\sum_i (x_i - \mu_i)\frac{\partial y}{\partial x_i}\right]^2$$

$$\simeq \sum_i \sum_j \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j} E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

$$\simeq \sum_{i,j} \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j} \mathrm{Cov}(x_i, x_j)$$

- Confidence intervals for any PDF: Ordering rules

The approximation of $-2\ln\mathcal{L}$ with a Gaussian or its excursion around its minimum guarantee an exact coverage only for a small set of cases, and in particular for large n.

Here we'll see a general approach.

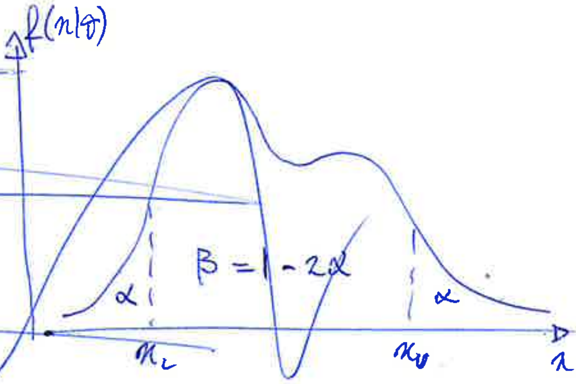
Suppose we have a variable x with PDF $f(x|\theta)$.

In general, f is not symmetric, so we need to decide how to compute an interval corresponding to some predefined probability content β.

→ ~~Central interval~~

$$\cancel{[x_L(\theta), x_U(\theta)] = [\bar{x}(\theta)-\delta, \bar{x}(\theta)+\delta]}$$

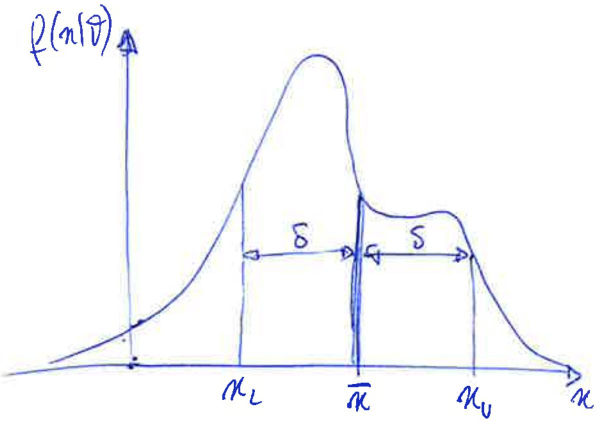
$$\cancel{\int_{-\infty}^{x_L} f(x|\theta)\,dx = \int_{x_U}^{+\infty} f(x|\theta)\,dx = \frac{1-\beta}{2}}$$



→ Central interval

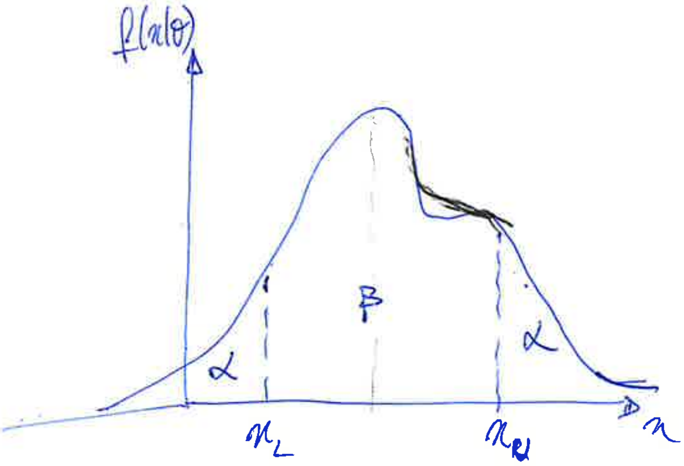
$$[x_L(\theta), x_U(\theta)] = [\bar{x}(\theta)-\delta, \bar{x}(\theta+\delta)]$$

$\bar{x}$ could be the mean or the mode



→ Equal areas

$$\int_{-\infty}^{x_L} f(x|\theta)\,dx = \int_{x_U}^{+\infty} f(x|\theta)\,dx = \frac{1-\beta}{2}$$



$\boxed{\text{Lect. 9}}$

- Neyman confidence belt

Take a variable x with PDF $f(x|\theta)$ and θ unknown.

~~Assume x could be an estimator of the parameter θ.~~

Suppose $t(x)$ is some function of the data.

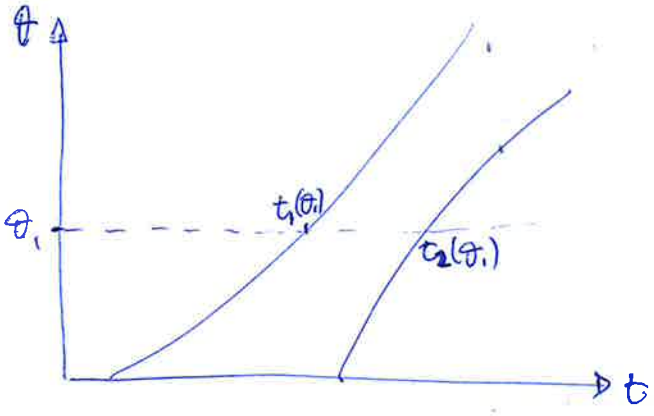
We can write: $\beta = P(t_1 \le t \le t_2) = P(t_1(\theta) \le t \le t_2(\theta))$.

$$\stackrel{!}{=} \int_{t_1}^{t_2} f(t|\theta)\, dt$$

Assume that we have a way to determine t_1 and t_2 for each value of θ.

Such values form two curves in the (t,θ) space: ~~These~~

From the PDF of t: $f(t(x)|\theta)$ we get t_1 and t_2

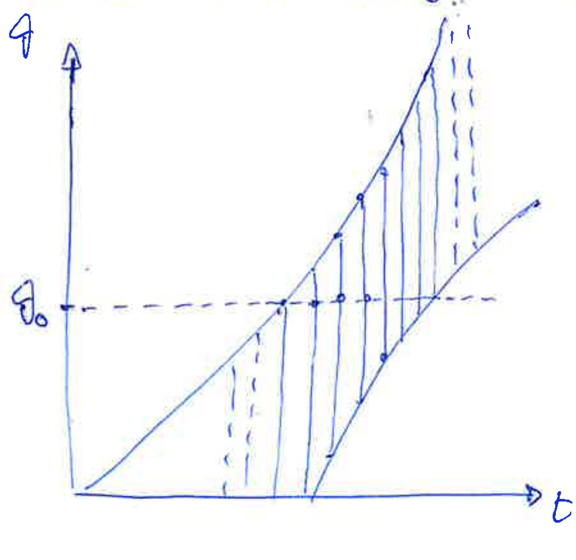


The ~~...~~ space between these curves is called the "Neyman confidence belt".

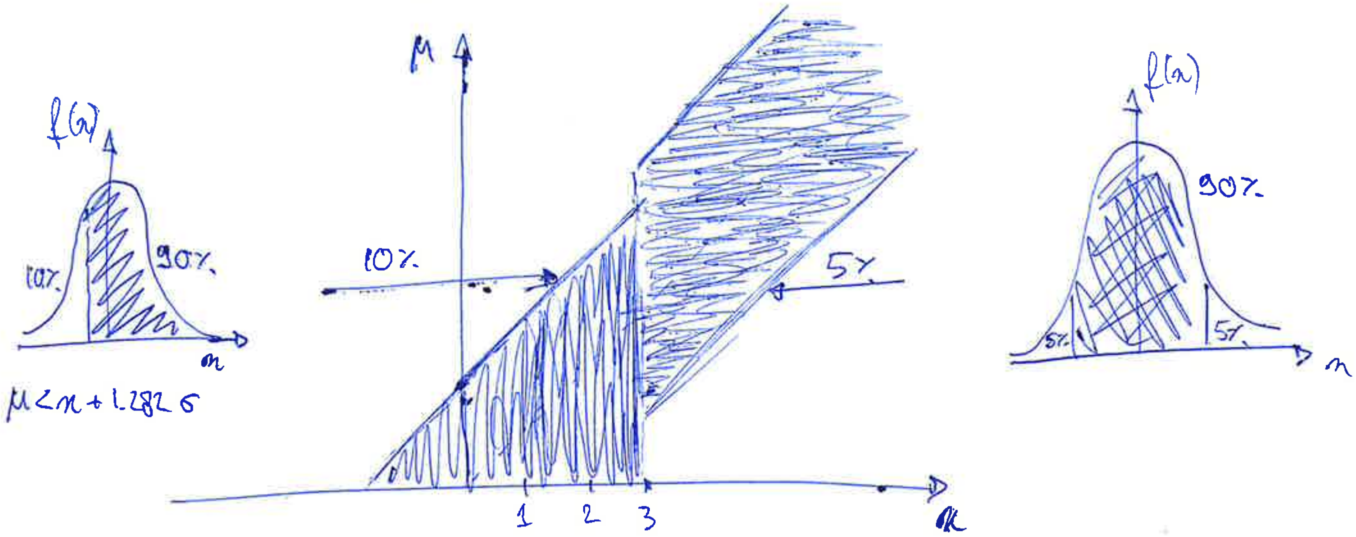
From this belt, ~~we want to~~ given a specific measured value t_0, we want to extract an interval on the parameter θ.

Suppose θ_0 is the true, unknown value of θ.

If we repeat the measurement many times, a fraction β of the measurements will fall in $[t_1(\theta_0), t_2(\theta_0)]$ by definition:



The corresponding belt would be:



For some values of μ, e.g. $\mu = 2.5$, we have a coverage of 85% only!

Notice that: → The coverage is a property of the method, not of ~~the~~ a particular interval

→ The flip-flopping issue arises from the fact that our ordering rule depends on the outcome of the measurement.
Feldman and Cousins showed that this should not be done!

<u>Example: Flip-flopping</u>

- Examples → FC Belt for Gaussian
  → FC belt for electron neutrino mass

- ~~Numerical~~ Analytical FC belt calculation for Gaussian

Recall the flip-flopping case, where we had x with a PDF:

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right)$$

The value $\hat{\mu}$ that maximizes $f(x|\mu)$ given some measured x is:

$$\hat{\mu}(x) = \max\{x, 0\}$$

The PDF for x, using the max-$\mathcal{L}$ estimate for μ is:

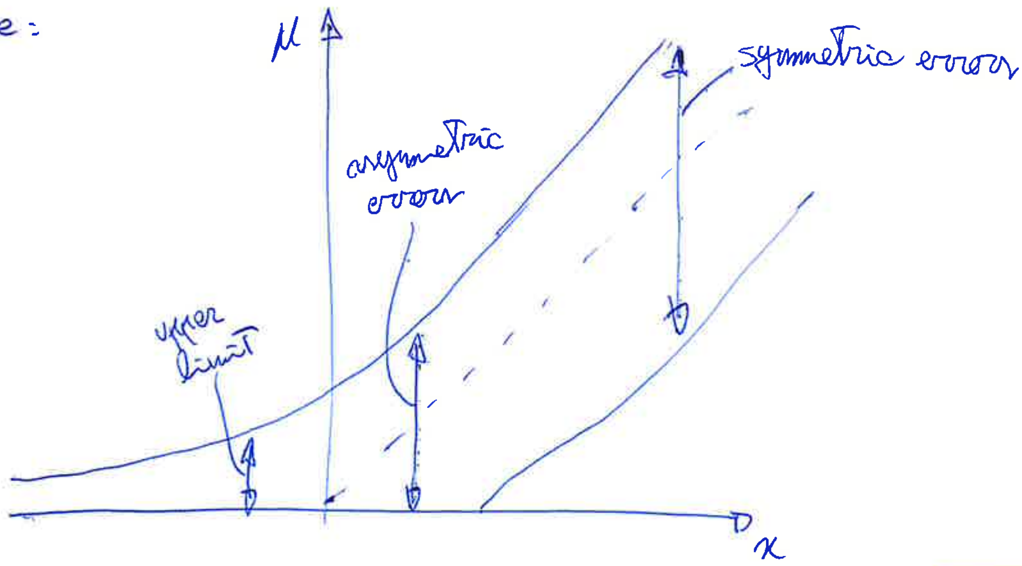
$$f(x|\hat{\mu}(x)) = \begin{cases} \frac{1}{\sqrt{2\pi}} & \text{if } x \geq 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) & \text{if } x < 0 \end{cases}$$

The likelihood ratio becomes:

$$\lambda(x|\mu) = \frac{f(x|\mu)}{f(x|\hat{\mu}(x))} = \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2}\right) & \text{if } x \geq 0 \\ \exp\left(x\mu - \frac{\mu^2}{2}\right) & \text{if } x < 0 \end{cases}$$

At this point, we can find the interval $[\mu_1, \mu_2]$ numerically for any value of μ.

The result will be:



If (1) and (2) give different results:

→ If the number of parameters is small (~ 2), Feldman-Cousins is easy to implement.

→ If the number of parameters is $\gtrsim 3$, the FC might become very complicated or CPU intensive, but in this situations typically the profile-$\mathcal{L}$ method provides good coverage.

→ Coverage calculation

In any case, one should make sure the method provides the desired coverage!

This can be done as follows:

1) Assume values for each parameter, generate $\gtrsim 10^4$ Toy-MC experiments, count how many times the confidence interval covers the true value of each parameter.

2) Repeat for different values of the parameters.