Statistical Methods in Particle Physics

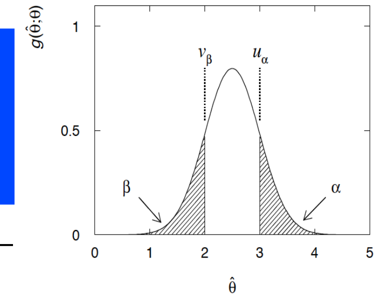
Lecture 12

January 16, 2012

Silvia Masciocchi, GSI Darmstadt
s.masciocchi@gsi.de

Winter Semester 2011 / 12

Outline



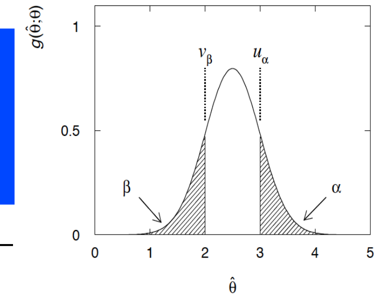
Interval estimation

- The standard deviation as statistical error
- Classical confidence intervals
  - For a parameter with a Gaussian distributed estimator
  - For the mean of a Poisson distribution
- Limits near a physical boundary

Discussion:

- Statistical and systematic errors
- Blind analysis

The standard deviation



We have seen methods for estimating properties of probability density functions (pdf's) and ways to obtain the variance of the estimators.

Suppose the result of an experiment is an estimate of a certain parameter:

n observations of random variable x

Hypothesis for the pdf f(x;θ), θ unknown parameter

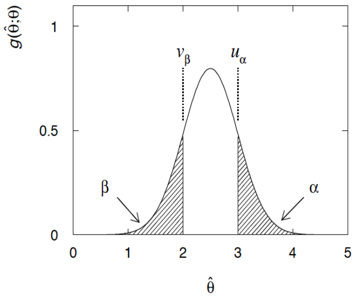
From x_1, ..., x_n build the function $\hat{\theta}(x_{;}, .., x_n)$ e.g. max. likelihood

→ Determine the estimator $\hat{\theta}_{obs}$ (value actually observed) and its standard deviation $\hat{\sigma}_{\hat{\theta}}$

The variance (or equivalently its square root, the standard deviation) of the estimator is a measure of how widely the estimates would be distributed if the experiment were to be repeated many times with the same number of observations per experiment

Standard deviation σ → statistical error or standard error

Statistical error



In reporting the measurement of θ as:

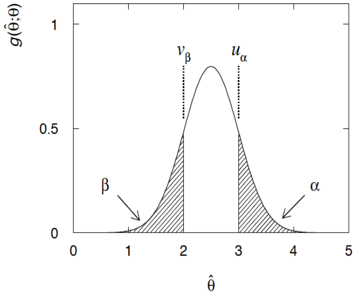
$$\hat{\theta}_{\text{obs}} \pm \hat{\sigma}_{\hat{\theta}}$$

one means that repeated estimates all based on n observations of x would be distributed according to a pdf $g(\hat{\theta})$ centered around some true value θ and true standard deviation, which are estimated to be

$$\hat{\theta}_{\text{obs}} \quad \text{and} \quad \hat{\sigma}_{\hat{\theta}}$$

- For most practical estimators, the sample pdf g becomes approximately Gaussian in the large sample limit
- If more than one parameter is estimated, the pdf becomes a multidimensional Gaussian characterized by a covariance matrix V
- The standard deviation, and in case the covariance matrix, tell everything how repeated estimates would be distributed

Confidence interval



If the form of the estimator pdf $g(\hat{\theta})$ is not Gaussian, then the 'standard deviation' definition of statistical error bars does not hold!

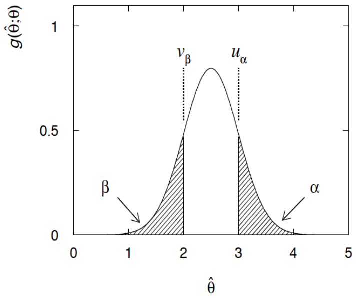
In such cases, one usually reports **confidence intervals**
→ an interval reflecting the statistical uncertainty of the parameter
(very often: asymmetric errors)

Such confidence intervals should:

- communicate objectively the result of the experiment;
- have a given probability of containing the true parameter;
- provide information needed to draw conclusions about the parameter possibly incorporating stated prior beliefs.

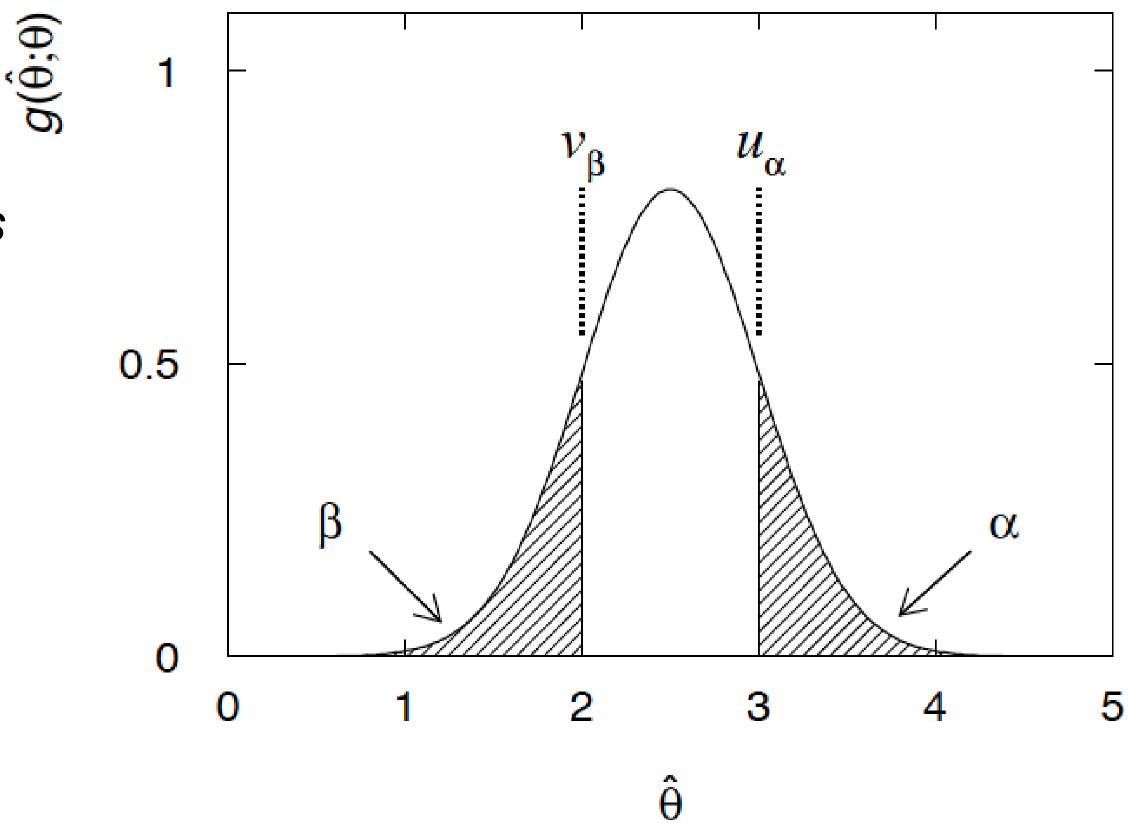
Special case: estimate limits of parameters near a physically excluded region (e.g. an observed event rate consistent with zero)

Frequentist confidence interval

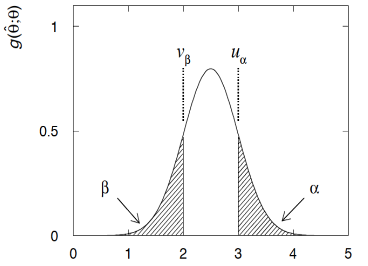


Consider the estimator $\hat{\theta}$ for a parameter θ, and an estimate $\hat{\theta}_{obs}$
The sampling distribution for θ is $g(\hat{\theta};\theta)$

By means of e.g. an analytical calculation or a Monte Carlo study, one knows g, which contains the true value θ as parameter.
That is, the real value of θ is not known, but for a given value of θ, one knows what the pdf of $\hat{\theta}$ would be

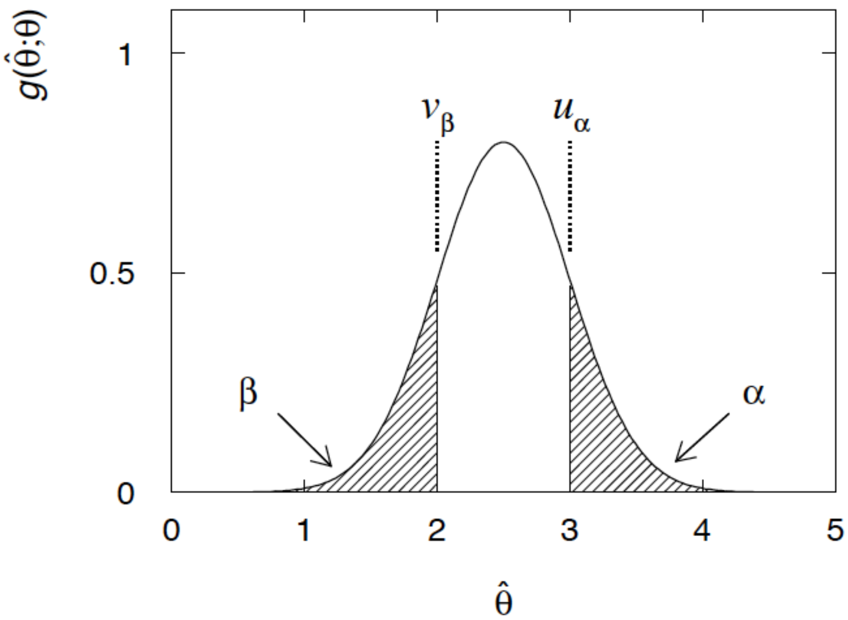


Frequentist confidence interval



From $g(\hat{\theta};\theta)$ one can determine the value u_α such that there is a fixed probability α to observe $\hat{\theta} \geq u_\alpha$:

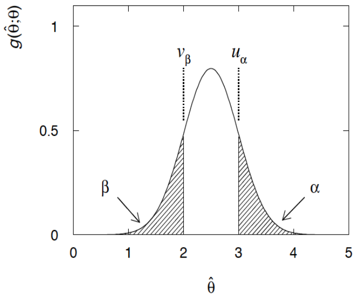
$$\alpha = P(\hat{\theta} \geq u_\alpha(\theta)) = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta};\theta)\, d\hat{\theta}$$



And the value v_β such that there is the probability β to observe $\hat{\theta} \leq v_\beta$:

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

Confidence belt

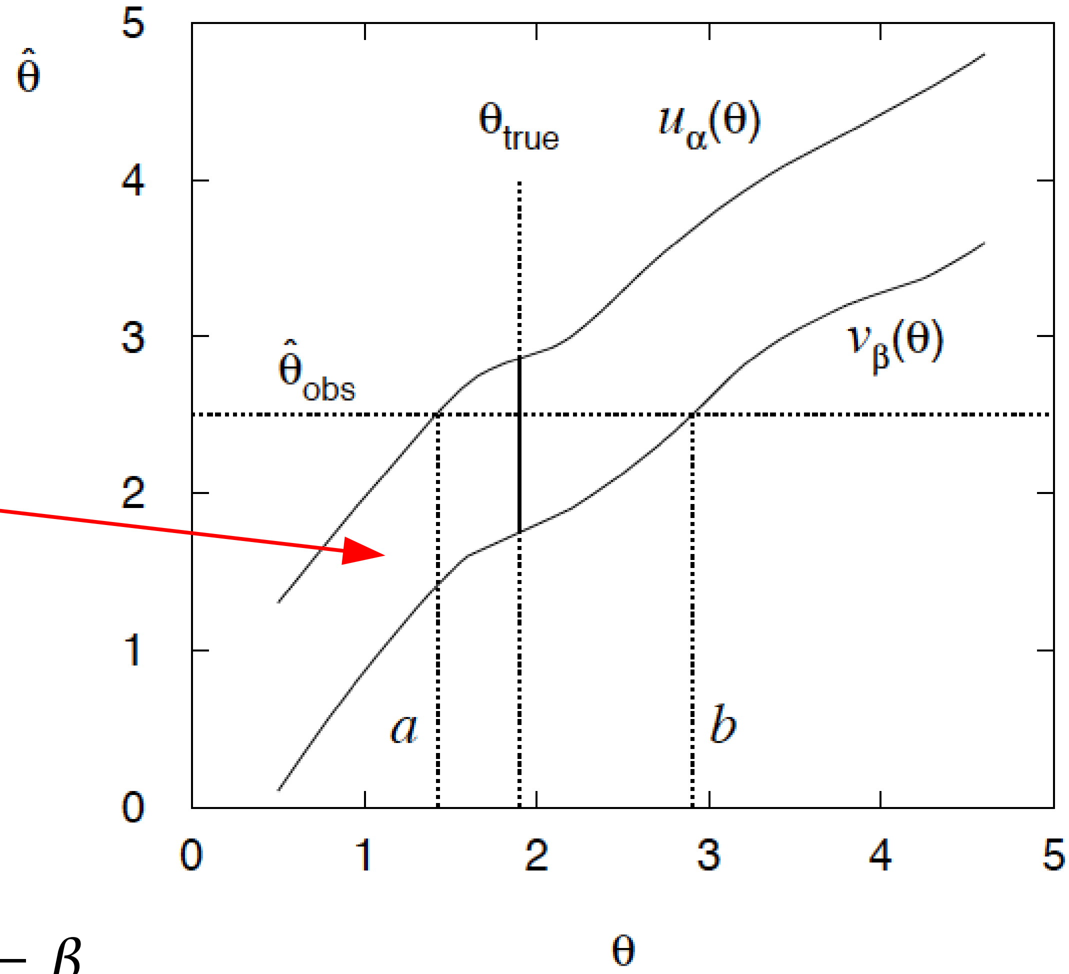


See how the functions $u_\alpha(\theta)$ and $v_\beta(\theta)$ can be as a function of the true value of θ

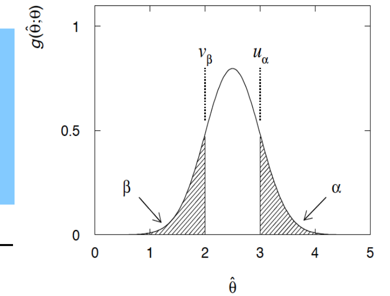
The region between the two curves is called the **confidence belt**

The probability for the estimator to be inside the belt, regardless of the true value of θ, is:

$$P\left(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)\right) = 1 - \alpha - \beta$$



Confidence belt

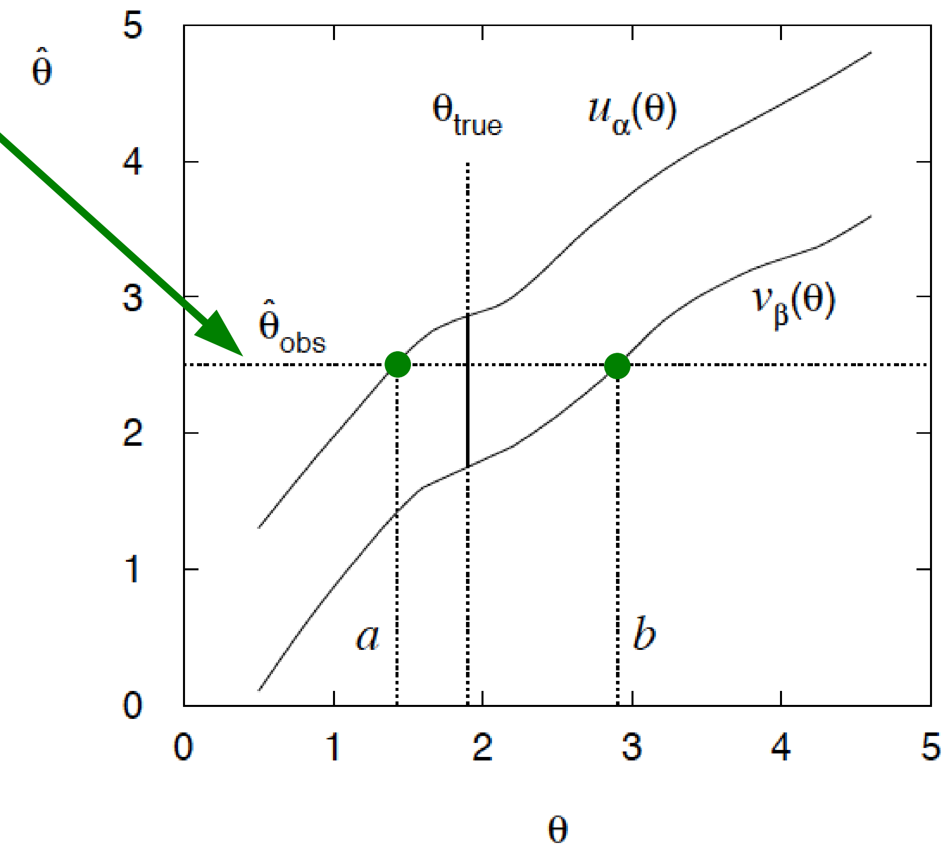


For the value of the estimator actually found in the experiment $\hat{\theta}_{\text{obs}}$

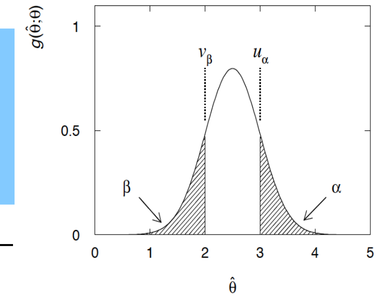
find the points where that intersects the confidence belt ●

this determines the points a and b

The interval [a,b] is called a **confidence interval** at a **confidence level** (or coverage probability) of $1 - \alpha - \beta$



Confidence interval

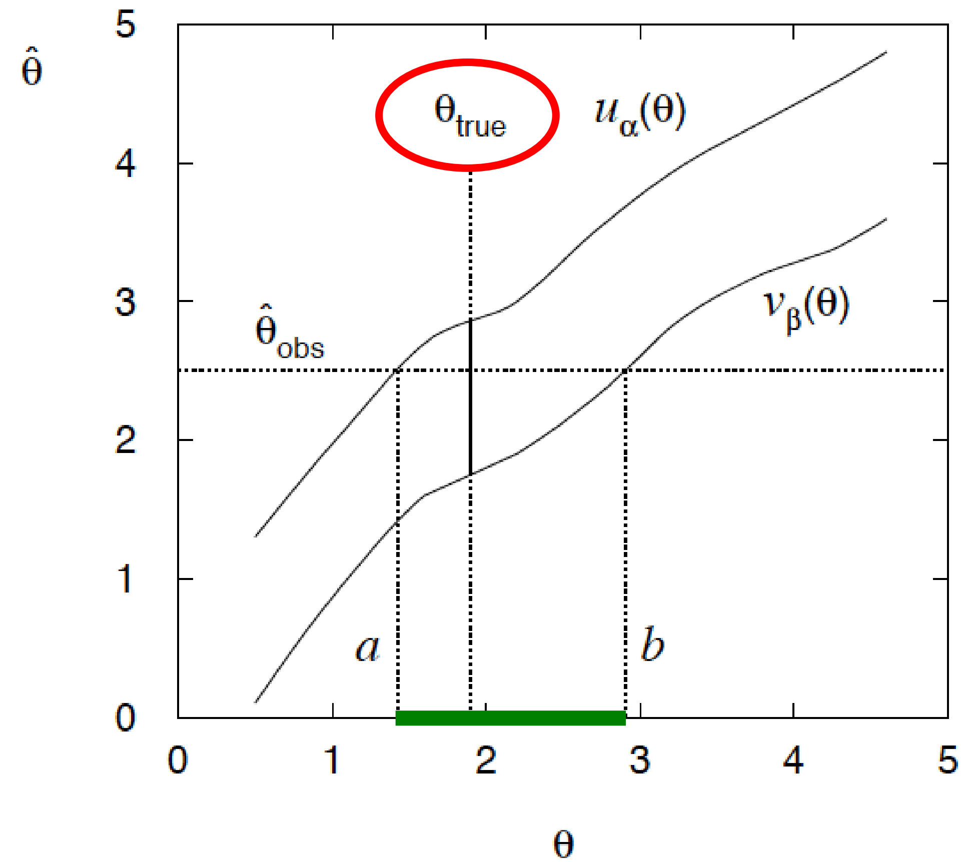


The interval [a,b] is called a **confidence interval** at a **confidence level** of $1 - \alpha - \beta$

Means that:

If the experiment were repeated many times, the interval [a,b] would include the true value of the parameter θ in a fraction $1 - \alpha - \beta$ of the experiments

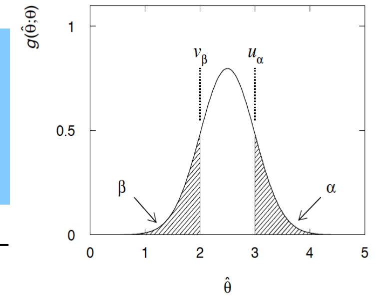
Also: $1 - \alpha - \beta$ is the probability for the interval to cover the true value of the parameter



Quote as: $\hat{\theta}^{+d}_{-c}$

where $c = \hat{\theta} - a$, $d = b - \hat{\theta}$ are usually displayed as error bars

One-sided and central confidence intervals



- Sometimes ONLY specify α OR β

  → **one-sided confidence interval** or limit

  That is, the value a represents a lower limit on the parameter θ such that a ≤ θ with the probability 1 – α

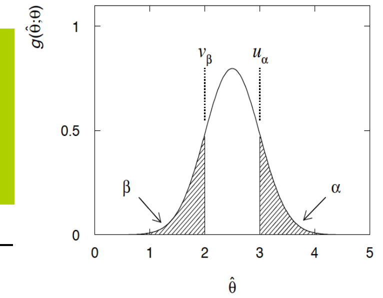
  Similarly, b represents an upper limit on θ such that P(θ ≤ b) = 1 – β

- One often chooses α = β = γ/2 giving a so-called **central confidence interval** with probability 1 – γ

  A central confidence interval does not necessarily mean that a and b are equidistant from the estimated $\hat{\theta}$, but only that the probabilities α and β are equal

  In high energy physics, the error convention is to take the 68.3% central confidence interval (see later)

Confidence interval by inverting a test



Confidence intervals for a parameter θ can be found by defining a **test** of the hypothesized value θ (do this for all θ):

- Specify values of the data that are 'disfavoured' by θ (critical region) such that:

  P(data in critical region) ≤ γ

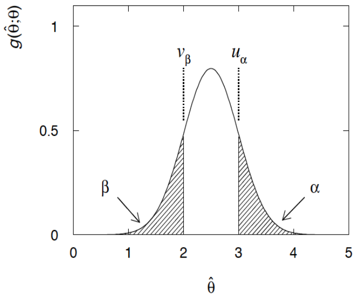
  for a specified γ, e.g. 0.05 or 0.1
- Invert the test to define a confidence interval as:

  set of θ values that would NOT be rejected in a test of size γ (the confidence level is 1-γ)

The interval will cover the true value of θ with probability ≥ 1 – γ.

Equivalent to a confidence belt construction. The confidence belt is acceptance region of a test

Confidence interval and p-value



Equivalently we can consider a significance test for each hypothesized value of θ, resulting in a p-value, p_θ

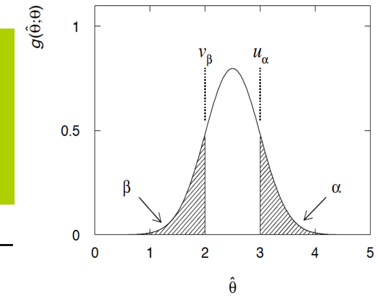
If $p_\theta < \gamma$, then reject θ

The confidence interval at CL = 1 – γ consists of those values of θ which are not rejected !

E.g. un upper limit on θ is the greatest value for which $p_\theta \geq \gamma$

In practice find by setting $p_\theta = \gamma$ and solve for θ

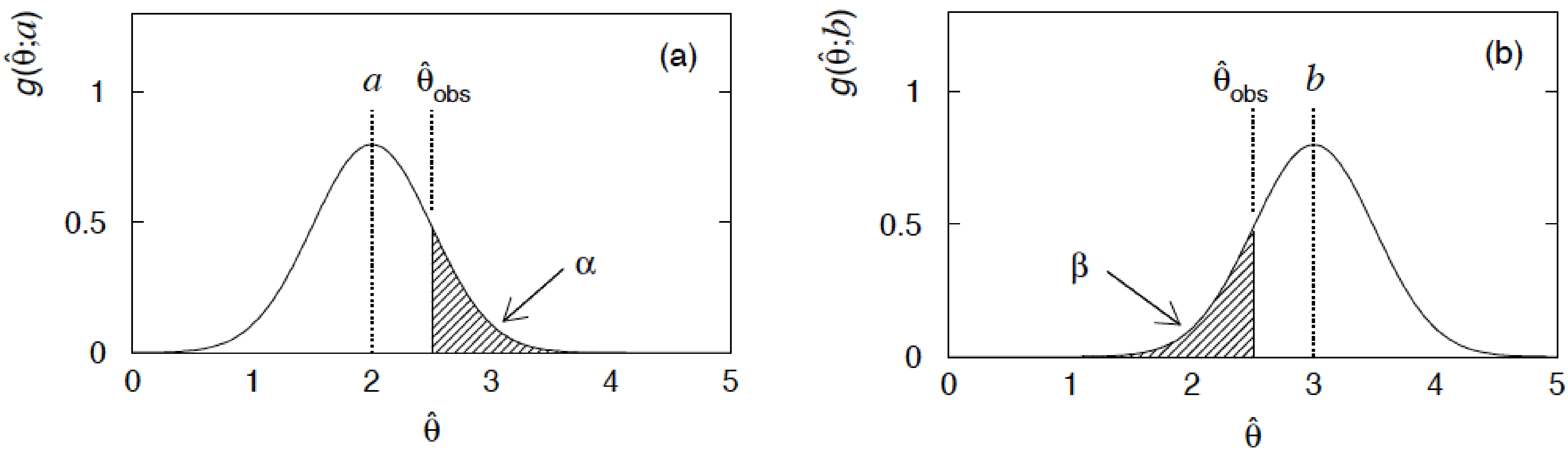
Confidence intervals in practice



Usually we do not construct confidence belts, but solve:

$$\alpha = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta};\theta)\, d\hat{\theta} = \int_{\hat{\theta}_{obs}}^{\infty} g(\hat{\theta};a)\, d\hat{\theta}$$

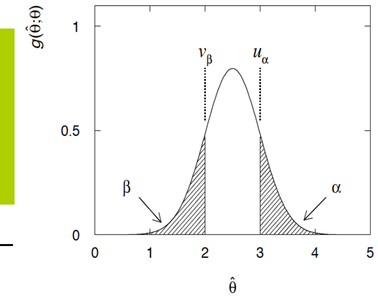
$$\beta = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta};\theta)\, d\hat{\theta} = \int_{\infty}^{\hat{\theta}_{obs}} g(\hat{\theta};b)\, d\hat{\theta}$$



→ a is hypothetical value of θ such that $P\,(\hat{\theta} > \hat{\theta}_{obs}) = \alpha$

→ b is hypothetical value of θ such that $P\,(\hat{\theta} < \hat{\theta}_{obs}) = \beta$

Meaning of the confidence interval



NOTE !!! the interval is random, the true θ is an unknown constant

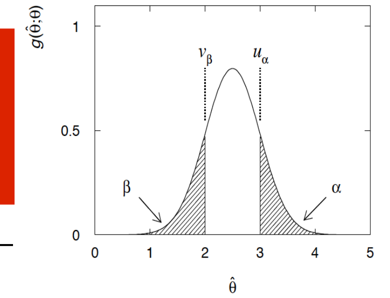
Often report interval [a,b] as

$$\hat{\theta}^{+d}_{-c} \quad \text{where} \quad c = \hat{\theta} - a, \quad d = b - \hat{\theta}$$

So, what does $\hat{\theta} = 80.25^{+0.31}_{-0.25}$ mean?

- It does NOT mean: $P(80.00 < \theta < 80.56) = 1 - \alpha - \beta$
- But rather: repeat the experiment many times with the same sample size, construct interval according to the same prescription each time, in $1 - \alpha - \beta$ of experiments, interval will cover θ

Confidence interval for Gaussian



Consider a Gaussian distributed estimator:

$$g(\hat{\theta};\theta) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(\frac{-(\hat{\theta}-\theta)^2}{2\sigma_{\hat{\theta}}^2}\right)$$

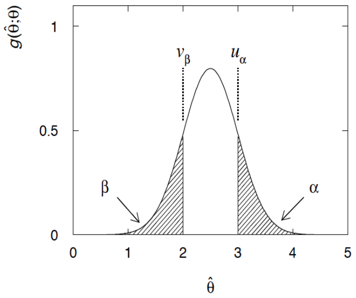
with mean θ and standard deviation $\sigma_{\hat{\theta}}$

It has the cumulative distribution of $\hat{\theta}$:

$$G(\hat{\theta};\theta,\sigma_{\hat{\theta}}) = \int_{-\infty}^{\hat{\theta}} \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(\frac{-(\hat{\theta}'-\theta)^2}{2\sigma_{\hat{\theta}}^2}\right) d\hat{\theta}'$$

This is a commonly occurring situation since, according to the **central limit theorem**, any estimator that is a linear function of a sum of random variables becomes Gaussian in the large sample limit.

Confidence interval for Gaussian



To find the confidence interval for θ, solve for a and b:

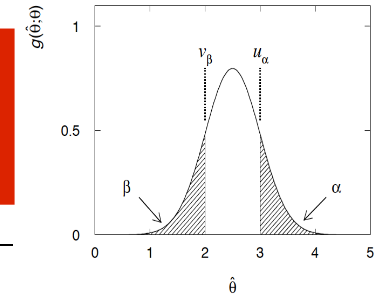
$$\alpha = 1 - G(\hat{\theta}_{\mathrm{obs}}; a, \sigma_{\hat{\theta}}) = 1 - \Phi\left(\frac{\hat{\theta}_{\mathrm{obs}} - a}{\sigma_{\hat{\theta}}}\right)$$

$$\beta = G(\hat{\theta}_{\mathrm{obs}}; b, \sigma_{\hat{\theta}}) = \Phi\left(\frac{\hat{\theta}_{\mathrm{obs}} - b}{\sigma_{\hat{\theta}}}\right)$$

where G is the cumulative distribution for $\hat{\theta}$ and

$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x'^2/2}\, dx'$ is the cumulative for the standard Gaussian

Confidence interval for Gaussian



Solving for a and b:

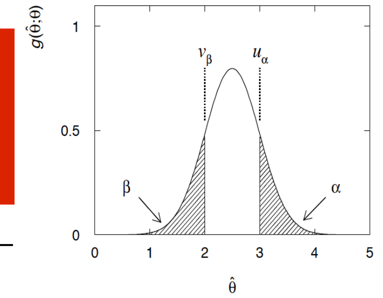
$$a = \hat{\theta}_{\text{obs}} - \sigma_{\hat{\theta}}\, \Phi^{-1}(1-\alpha)$$

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

Φ^{-1} = quantile of standard Gaussian (inverse of cumulative distribution, use ROOT)

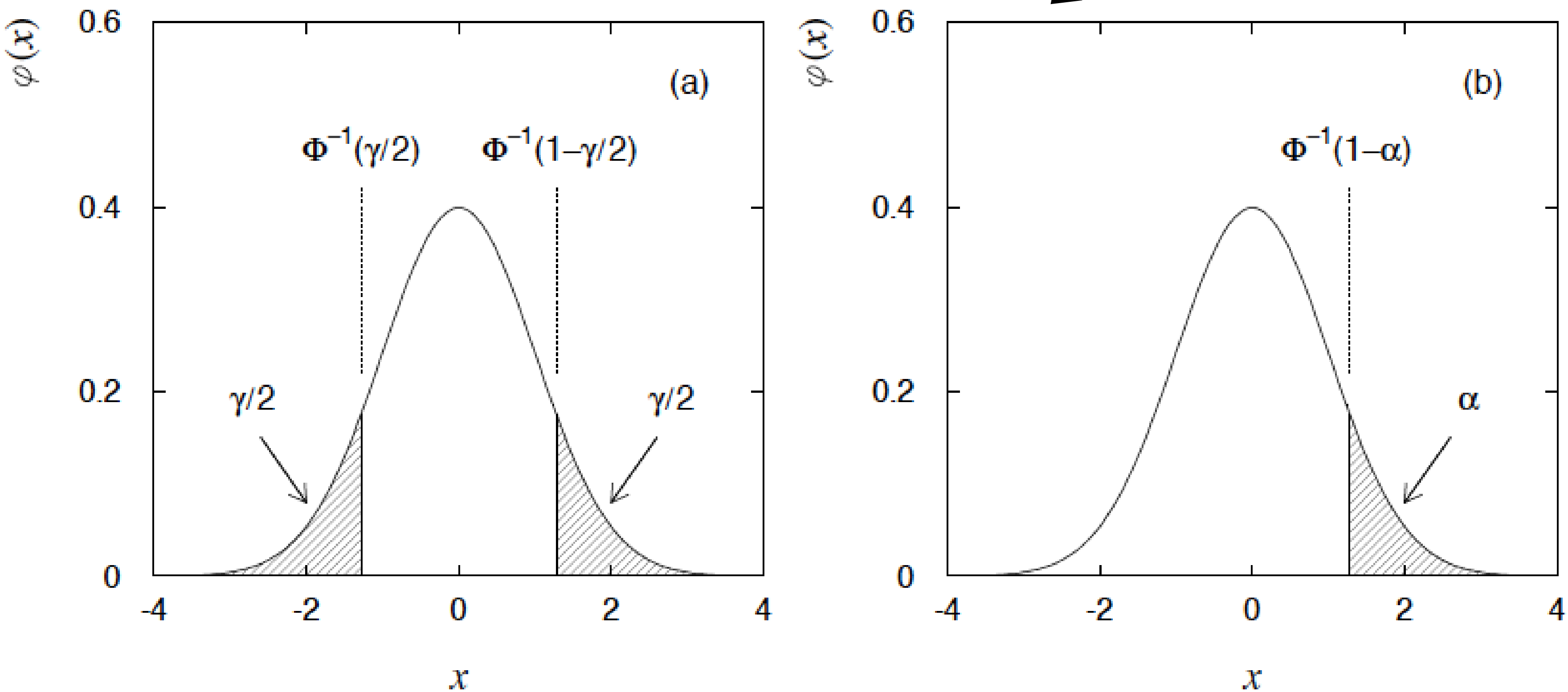
→ $\Phi^{-1}(1-\alpha), \Phi^{-1}(1-\beta)$ give how many standard deviations a and b are from $\hat{\theta}$

Quantiles of the standard Gaussian

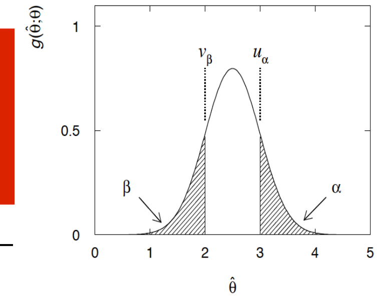


When we have a Gaussian estimator, to have a central confidence interval or a one-sided limit, we need to know the quantiles shown here:

$\alpha = \beta = \gamma/2$



Quantiles of the standard Gaussian



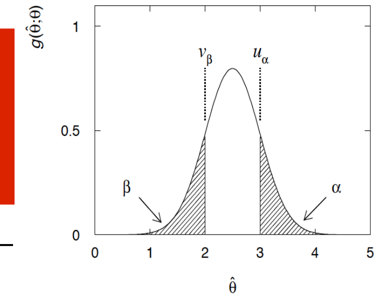
Typically, take a round number for the quantile (NUMBER OF SIGMAS !!!)

central		one-sided	
$\Phi^{-1}(1-\gamma/2)$	$1-\gamma$	$\Phi^{-1}(1-\alpha)$	$1-\alpha$
1	0.6827	1	0.8413
2	0.9544	2	0.9772
3	0.9973	3	0.9987

Or a round number for the coverage probability:

central		one-sided	
$1-\gamma$	$\Phi^{-1}(1-\gamma/2)$	$1-\alpha$	$\Phi^{-1}(1-\alpha)$
0.90	1.645	0.90	1.282
0.95	1.960	0.95	1.645
0.99	2.576	0.99	2.326

Gaussian estimator: summary



For the conventional 68.3% central confidence interval, one has:

$\alpha = \beta = \gamma/2$

With $\Phi^{-1}(1-\gamma/2) = 1$

i.e. a 1 σ error bar.

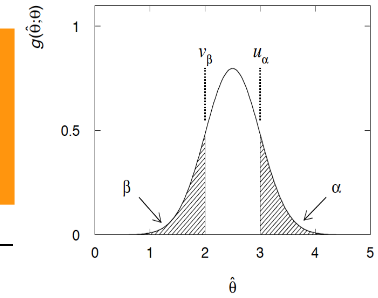
This results in the simple prescription:

$$[a,b] = [\hat{\theta}_{\text{obs}} - \sigma_{\hat{\theta}},\, \hat{\theta}_{\text{obs}} + \sigma_{\hat{\theta}}]$$

The final result of the measurement of θ is then simply reported as:

$$\hat{\theta}_{\text{obs}} \pm \sigma_{\hat{\theta}}$$

Confidence interval for mean of Poisson distr



Suppose n is Poisson distributed, the estimator: $\hat{\nu} = n$

Estimate is $\hat{\nu}_{obs} = n_{obs}$

$$P(n;\nu) = \frac{\nu^n}{n!} e^{-\nu}, \qquad n = 0, 1, \ldots$$

Minor problem: for fixed α, β, the confidence belt does not exist for all ν

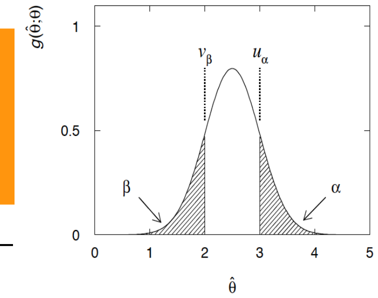
Just solve:

$$\alpha = P(\hat{\nu} \geq \hat{\nu}_{obs}; a) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{a^n}{n!} e^{-a}$$

$$\beta = P(\hat{\nu} \leq \hat{\nu}_{obs}; b) = \sum_{n=0}^{n_{obs}} \frac{b^n}{n!} e^{-b}$$

for a and b

Confidence interval for mean of Poisson distr



Use the trick:

$$\sum_{n=0}^{n_{obs}} \frac{\nu^n}{n!} e^{-\nu} = 1 - F_{\chi^2}(2\nu; n_d = 2(n_{obs}+1))$$

where F_{χ^2} is the cumulative chi-square distribution for n_d degrees of freedom.

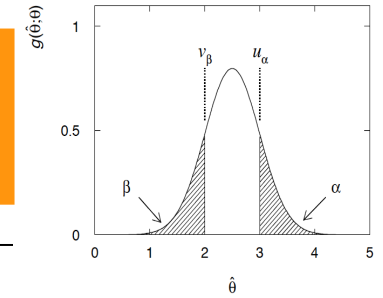
Find:

$$a = \frac{1}{2} F_{\chi^2}^{-1}(\alpha; n_d = 2n_{obs})$$

$$b = \frac{1}{2} F_{\chi^2}^{-1}(1-\beta; n_d = 2(n_{obs}+1))$$

where $F_{\chi^2}^{-1}$ is the quantile of the chi-square distribution

Confidence interval for mean of Poisson distr



An important case: $n_{obs} = 0$

$$\beta = \sum_{n=0}^{0} \frac{b^n e^{-b}}{n!} = e^{-b} \quad \rightarrow \quad b = -\log\beta$$

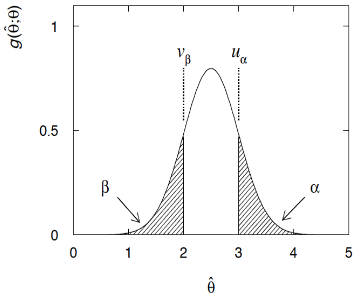
Calculate an upper limit at confidence level (1-β) = 95%

$$b = -\log(0.05) = 2.996 \approx 3$$

Useful table:

n_{obs}	lower limit a			upper limit b		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0	–	–	–	2.30	3.00	4.61
1	0.105	0.051	0.010	3.89	4.74	6.64
2	0.532	0.355	0.149	5.32	6.30	8.41
3	1.10	0.818	0.436	6.68	7.75	10.04
4	1.74	1.37	0.823	7.99	9.15	11.60
5	2.43	1.97	1.28	9.27	10.51	13.11

Limits near a physical boundary



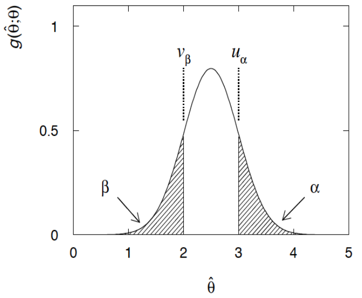
Often the purpose of an experiment is to search for a new effect, the existence of which would imply that a certain parameter is not equal to zero. For example, the existence of the Higgs.

If the data yield a value of the parameter significantly different from zero, then the new effect has been **discovered**, and the parameter's value and a confidence interval to reflect its error are given as the result.

If, on the other hand, the data result in a fitted value of the parameter that is consistent with zero, then the result of the experiment is reported by giving an **upper limit** on the parameter (a similar situation occurs when absence of the new effect corresponds to a parameter being large or infinite; one then places a lower limit).

The procedure to set limits is very delicate and can present serious difficulties (estimators which can take on values in the excluded region, negative mass of a particle, negative number of events, etc).

Setting limits on Poisson parameter



Consider the case of finding $n = n_s + n_b$ events where

n_b events from known processes (background)

n_s events from a new process (signal)

are Poisson random variables with means s and b.

Therefore $n = n_s + n_b$ is also Poisson distributed, with mean s+b

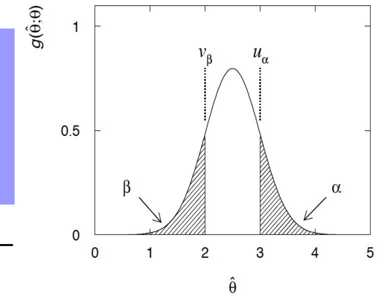
Assume b is known.

Suppose we are searching for evidence of the signal process, but the number of events found is roughly equal to the expected number of background events, e.g. b = 4.6 and we observe $n_{obs} = 5$ events.

The evidence for the presence of signal events is not statistically significant

→ set an upper limit on the parameter s

Upper limit for Poisson parameter



Find the hypothetical value of s such that there is a given small probability, say γ=0.05 to find as few events as we did or less:

$$\gamma = P(n \le n_{obs}; s, b) = \sum_{n=0}^{n_{obs}} \frac{(s+b)^n}{n!} e^{-(s+b)}$$

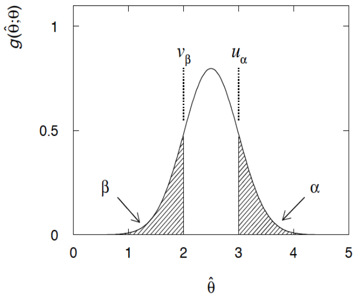
Solve numerically for $s = s_{up}$

This gives an **upper limit on s at a confidence level of (1-γ)**

Example (see page before):

Suppose b = 0 and we find n_{obs} = 0. For (1-γ) = 0.95, $s_{up} \approx 3$

Calculating Poisson parameter limits

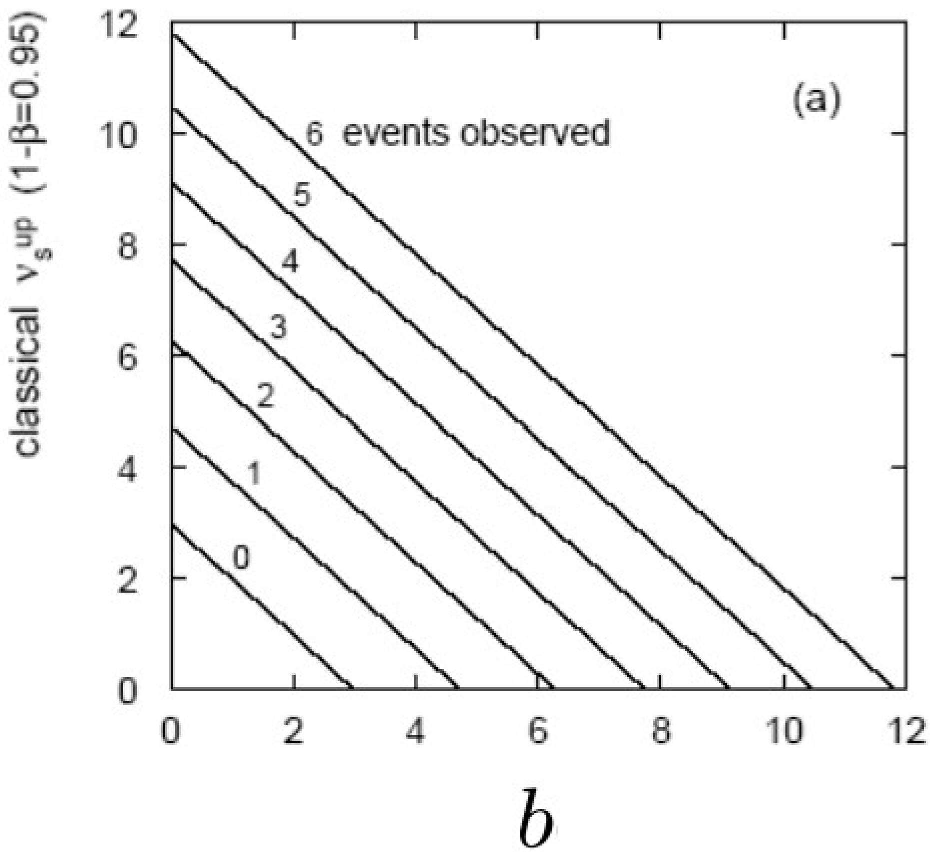


To solve for s_{lo}, s_{up}, we can exploit the relation to the χ^2 distribution (see page 22)

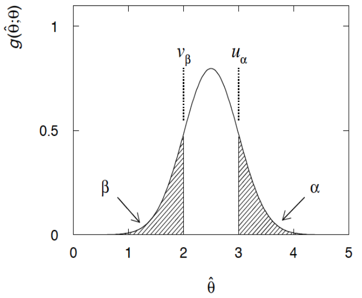
$$s_{lo} = \frac{1}{2} F^{-1}_{\chi^2}(\alpha; 2n) - b$$

$$s_{up} = \frac{1}{2} F^{-1}_{\chi^2}(1-\beta; 2(n+1)) - b$$

For low fluctuation of n, this can give negative result for s_{up}
i.e. confidence interval is empty



Limits near a physical boundary



Suppose for example b = 2.5 and we observe n = 0.

If we choose CL = 0.9, we find from the formula for s_{up}:

$$s_{up} = -0.197 \quad (CL = 0.90)$$

Physicist:

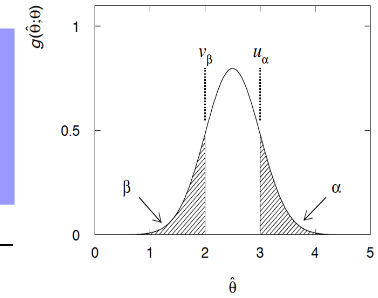
We already knew s ≥ 0 before we started; cannot use negative upper limit to report a result!

Statistician:

The interval is designed to cover the true value only 90% of the time: This was clearly not one of those times.

Not uncommon dilemma when limit of parameter is close to a physical boundary!

Expected limit for s = 0



Physicist: I should have used CL = 0.95 → then $s_{up} = 0.496$

Even "better": for CL = 0.917923 we get $s_{up} = 10^{-4}$!

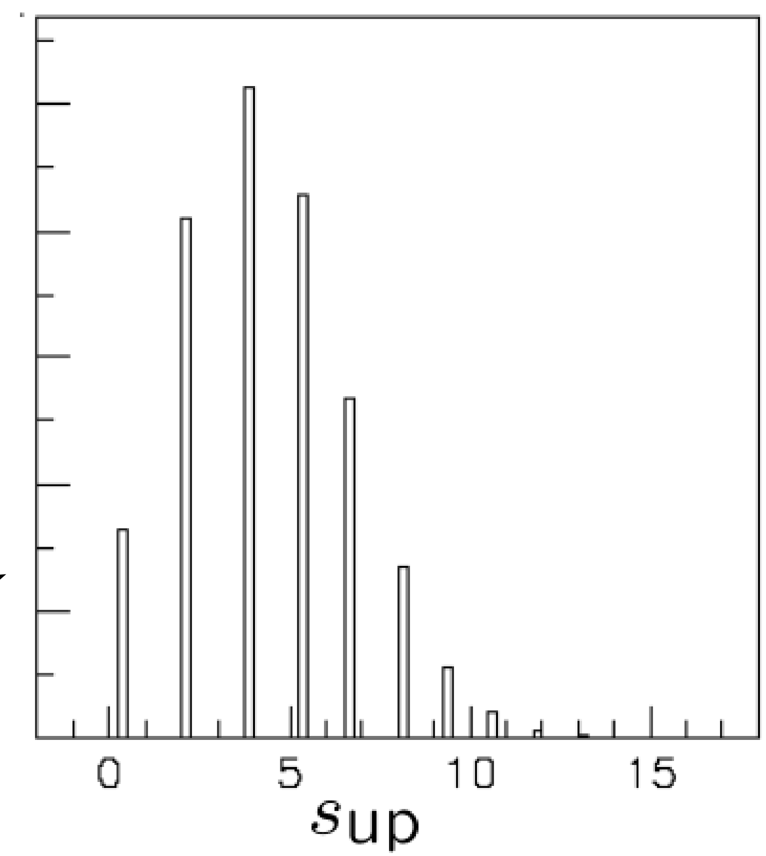
Forget it !

Reality check: with b = 2.5, typical Poisson fluctuation in n is at least $\sqrt{2.5} = 1.6$

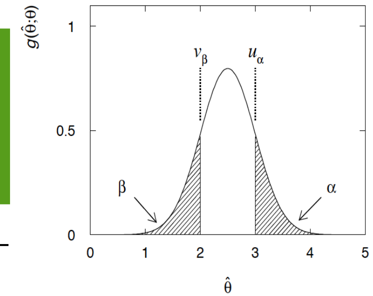
How can the limit be so low?

Look at the mean limit for the no-signal hypothesis (s=0) (sensitivity)

Distribution of 95% CL limits with b = 2.5 and s = 0. Mean upper limit is 4.44



Approximate confidence intervals from lnL or chi2



Recall the trick to estimate $\sigma_{\hat{\theta}}$ if ln L(θ) is parabolic:

$$\ln L(\hat{\theta} \pm N\,\sigma_{\hat{\theta}}) = \ln L_{\max} - \frac{N^2}{2}$$

CLAIM: this still works even if ln L is not parabolic, as an approximation for the confidence interval.

i.e. use

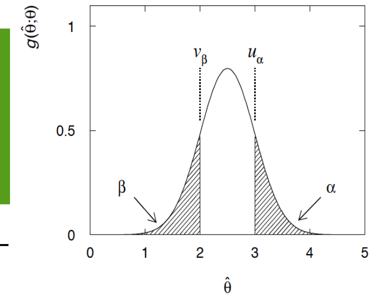
$$\ln L(\hat{\theta}^{+d}_{-c}) = \ln L_{\max} - \frac{N^2}{2}$$

$$\chi^2(\hat{\theta}^{+d}_{-c}) = \mathrm{chi2}_{\min} + N^2$$

where $N = \Phi^{-1}(1 - \gamma/2)$

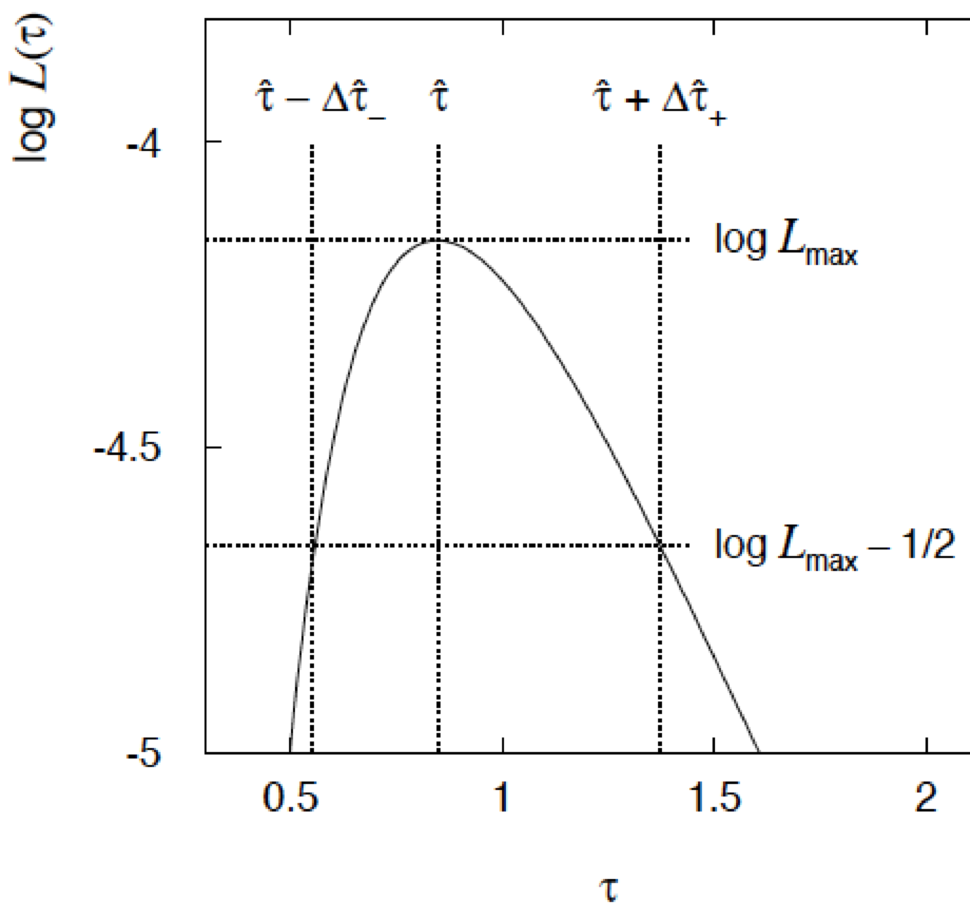
is the quantile of the standard Gaussian corresponding to the CL 1-γ.

For example: N = 1 → 1-γ = 0.683



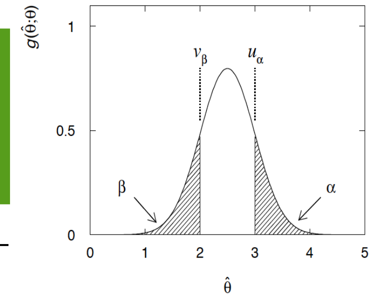
Approximate confidence intervals from lnL or chi2

Exponential example (see lecture 10): take interval where ln L is within 1/2 of the maximum → approximation of 68.3% confidence interval



$$\hat{\tau} = 0.85^{+0.52}_{-0.30}$$

For the non classical cases ...

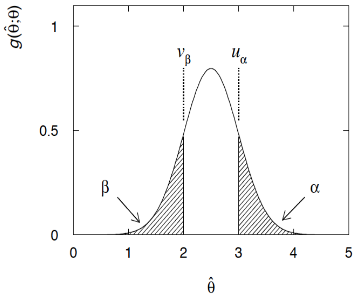


In many practical applications, estimators are Gaussian distributed (at least approximately). In this case the confidence interval can be determined easily.

Similarly is for estimators with a Poisson distribution.

But even in the other cases, a simple approximate technique can be applied using the likelihood function (or equivalently the function).

Discussion



- Statistical and systematic uncertainties

- Blind analysis