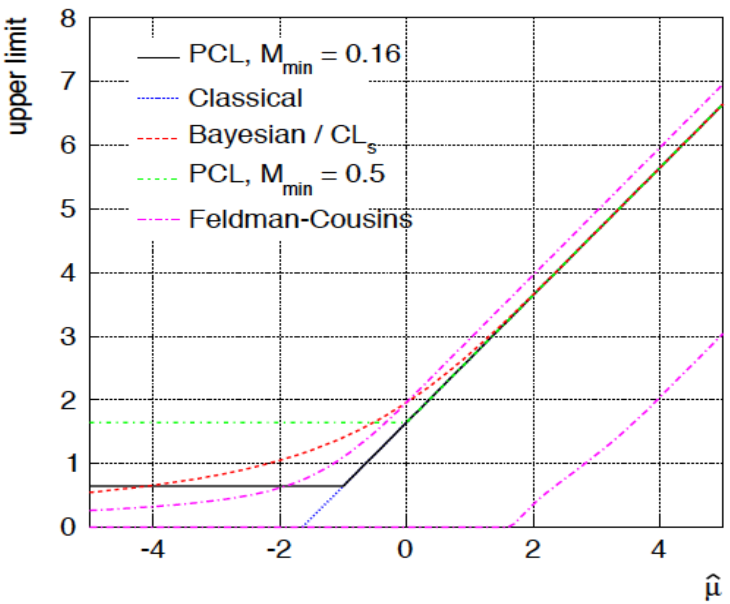
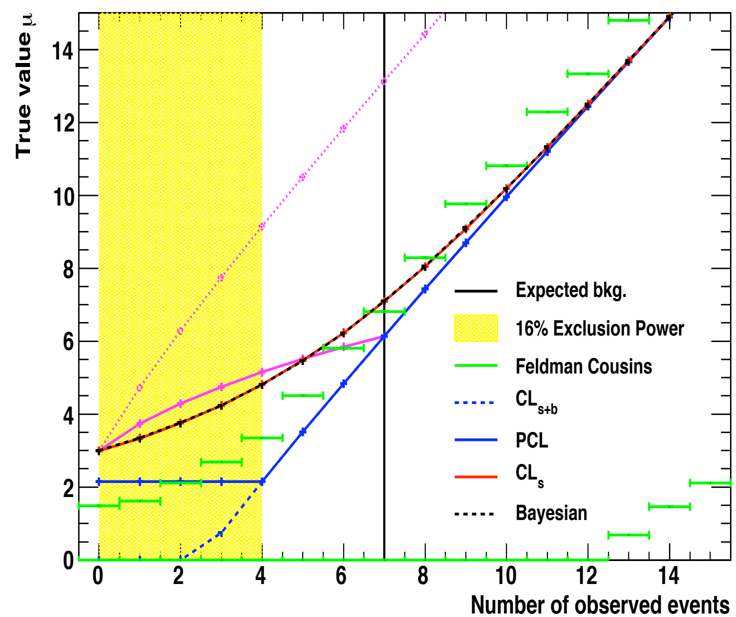
Confidence Intervals and Limits for Pedestrians

Markus Schumacher

GK Lecture, Freiburg, 26 - 28 September 2017





Goal of the lecture: understand the content and interpretation of the two figures

Outline

Lecture 1: Basics (26.9.)

- Motivation
- Frequentist and Bayesian Probability
- Parameter Estimation from Maximum Likelihood
- Frequentist Confidence Intervals a la Neyman and Coverage
- Bayesian Credibility Interval from Likelihood Principle

Lecture 2: Limits for Gaussian Probability Distribution (27.9)

- Connection of Frequentist Limit to Frequentist Hypothesis Test
- Limits close to physical boundary
- Frequentist and Bayesian Limits
- Modified Frequentist: CL_s Method and Power Constrained Limit (PCL)
- Unified Approach, Feldman- Cousins Intervals (FCL)

Lecture 3: Limits for Poisson Distribution (28.9.)

- Confidence Intervals
- Limits close to physical boundary
- Frequentist, Bayesian, PCL, CL_s, FC Limits

Outline

Lecture 1: Basics (26.9.)
- Motivation
- Frequentist and Bayesian Probability
- Parameter Estimation from Maximum Likelihood
- Frequentist Confidence Intervals a la Neyman and Coverage
- Bayesian Credibility Interval from Likelihood Principle

Lecture 2: Limits for Gaussian Probability Distribution (today)
- Connection of Frequentist Limit to Frequentist Hypothesis Test
- Limits close to physical boundary
- Frequentist and Bayesian Limits
- Modified Frequentist: CL_S Method and Power Constrained Limit (PCL)
- Unified Approach, Feldman- Cousins Intervals (FCL)→ 28.9

Lecture 3: Limits for Poisson Distribution (28.9.)
- Confidence Intervals with and w/o background
- Limits close to physical boundary
- Frequentist, Bayesian, PCL, CL_s, FC Limits

Interpretation of CI: Frequentist and Bayesian

CI: Attempt for a probability statement connecting measurement with true value

Frequentist:
- objects to / can not make probability assignment to true values
- construct a confidence interval CI [a,b] at xy% CL from data in such a way that in a sequence of repeated identical measurements the fraction xy% of such intervals contains the true value
- no statement about true value in a single experiment
- ”the coverage probability of the interval is XY %“
- no problems with “empty” intervals: $m_\nu^2 < -1\ eV^2$, $s < -0.3$ @95% CL

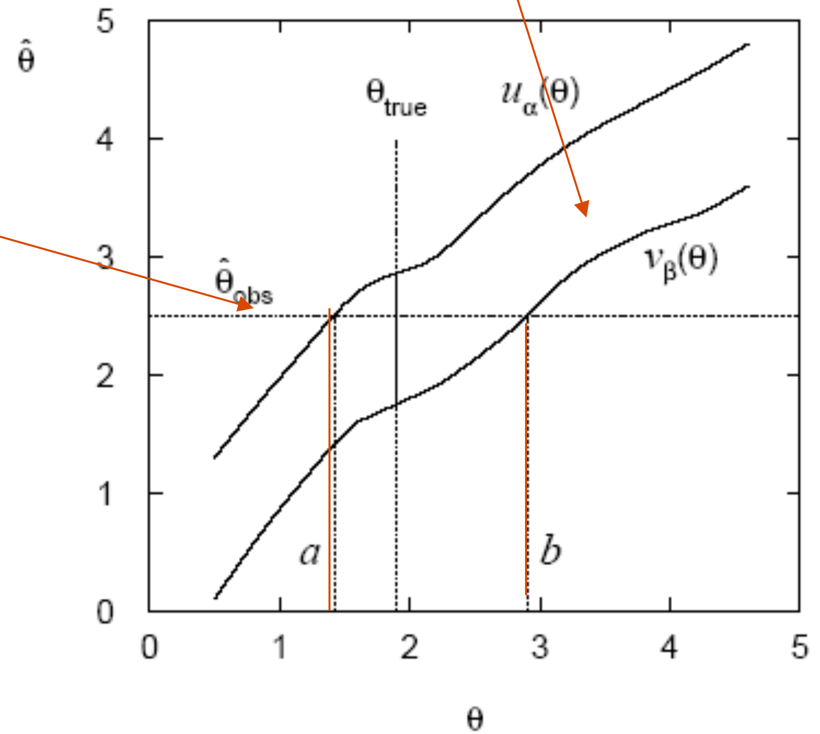
Bayesian:
- wants to make statement about probability of true value from single measurement
- credibility interval / Bayesian confidence interval [a,b] at xy% CL
- probability / degree of belief that true values lies in [a,b] is xy%
- coverage and outcome of not observed experiments not interesting
- all information is in observed likelihood function → likelihood principle
- „empty“ intervals are meaningless in Bayesian interpretation but are avoided by an appropiate prior probability

Classical Frequentist Intervals

Region btw. $u_\alpha(\theta)$ and $v_\beta(\theta)$ is the confidence belt $P(l_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)) = 1 - \alpha - \beta$

Boundaries of confidence interval given by intersect of observed value with confidence belt → [a,b]

$$a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta})$$
$$b(\hat{\theta}) \equiv l_\beta^{-1}(\hat{\theta}).$$



For all possible true values θ holds:

$$\hat{\theta} \geq u_\alpha(\theta) \Leftrightarrow a(\hat{\theta}) \geq \theta \qquad P(a(\hat{\theta}) \geq \theta) = \alpha$$
$$\hat{\theta} \leq l_\beta(\theta) \Leftrightarrow b(\hat{\theta}) \leq \theta, \qquad P(b(\hat{\theta}) \leq \theta) = \beta$$

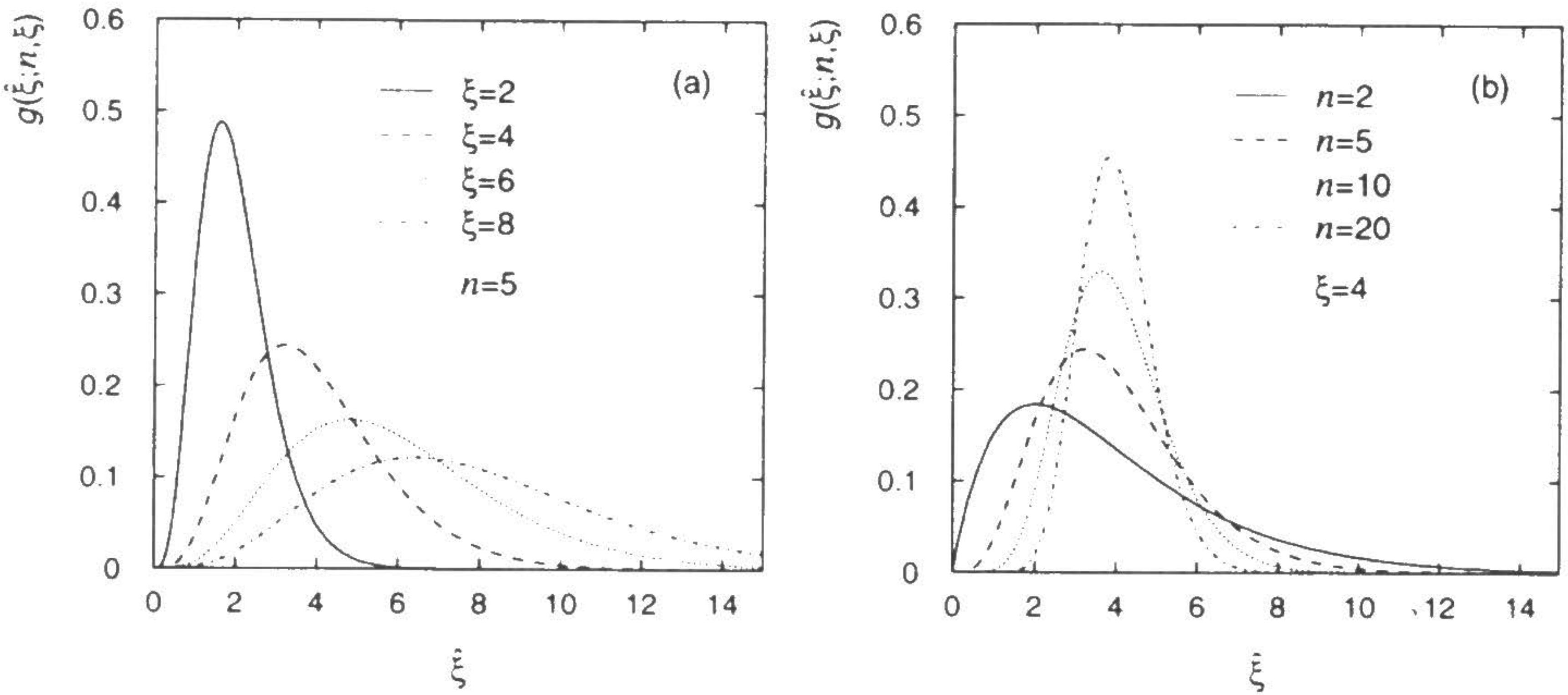
$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

Correct coverage by construction

Calculation of confidence belt very CPU intensive

Construction of CI for Exponential PDF

ML-Schätzer Estimator = arithmetic mean of lifetimes
PDF for ML estimator is special case of gamma function



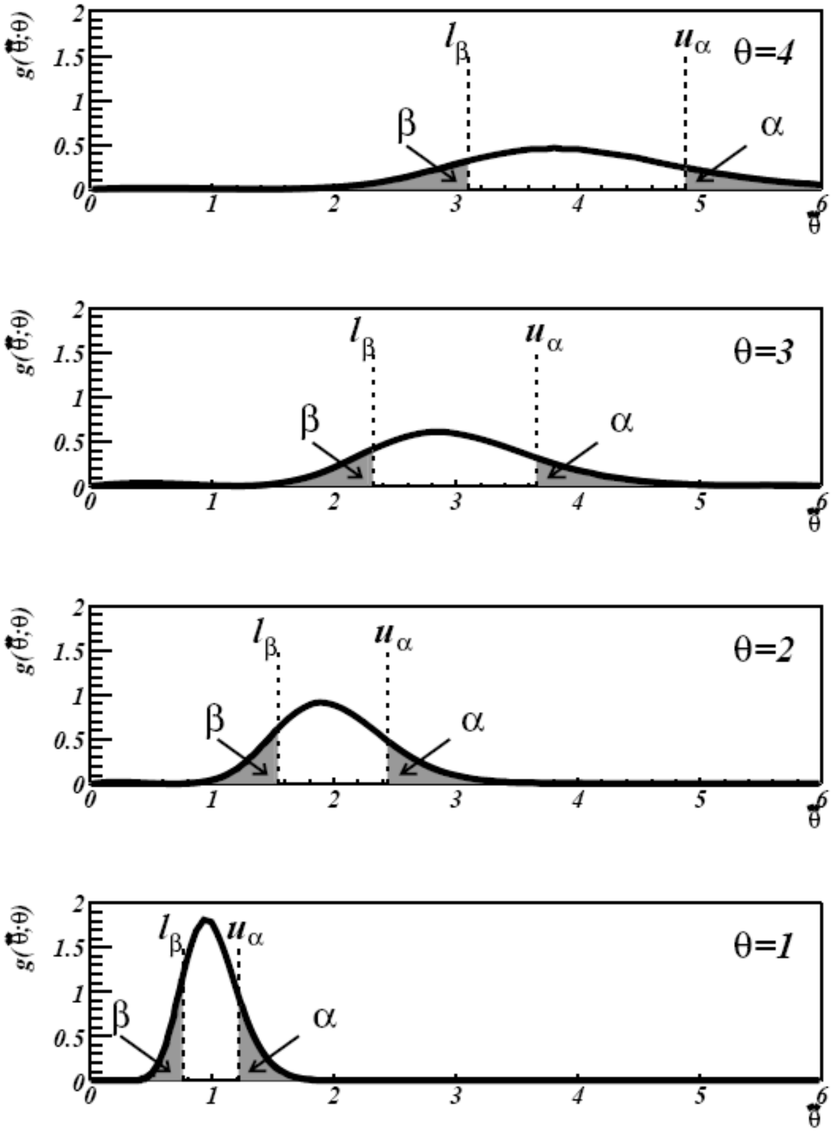
PDF for N_{SP}=5 various true lifetimes

for true lifetime = 4,various N_{SP}

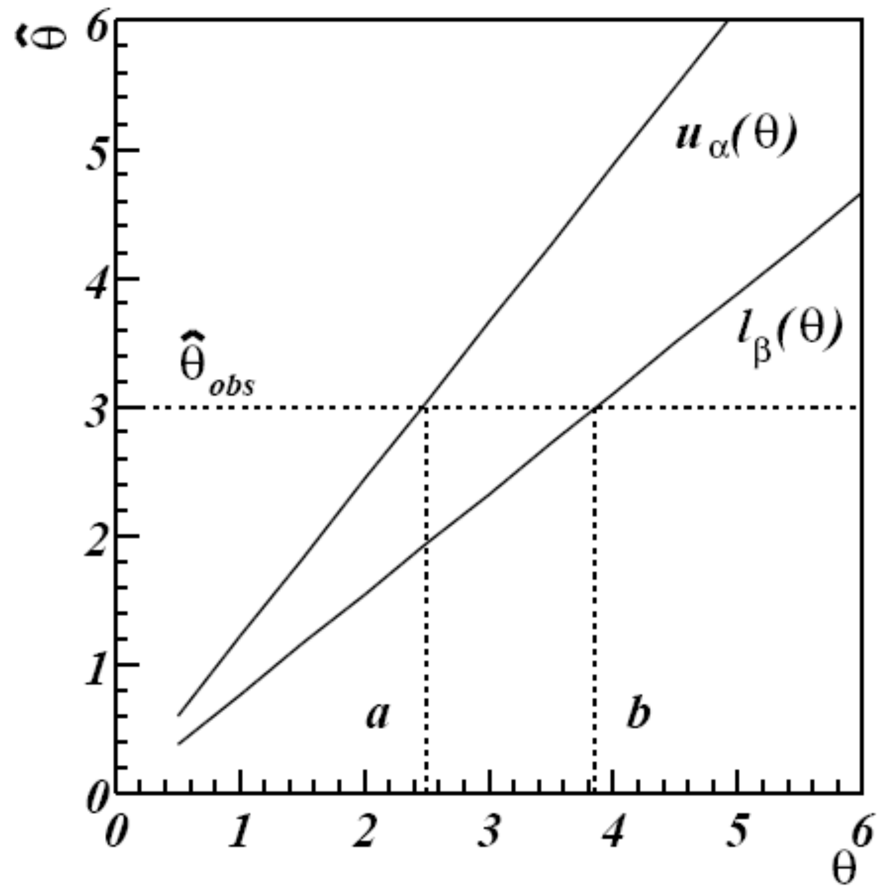
for $N_{SP} \to \infty$ PDF converges to Gaussian PDF due to Central Limit Theorem

Construction of CI for Exponential PDF (2)

PDF for ML-estimator for various true life times for sample size $N_{SP} = 20$



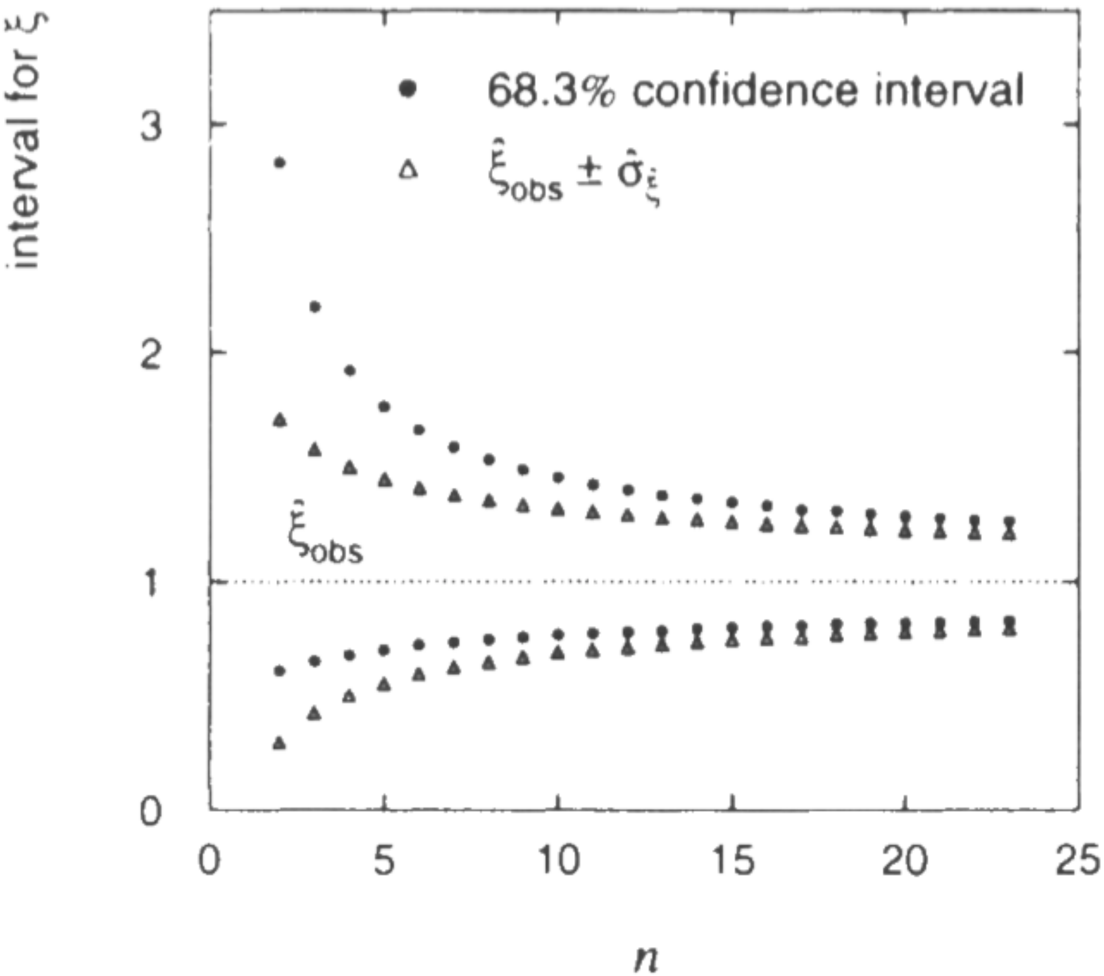
Confidence belt: u_α and l_β



Confidence interval [a;b]

Construction of CI for Exponential PDF (3)

Comparison of CI from estimator ± 1 standard deviation (triangles)
and from correct Nyman Construction (points)



for n = N_{SP} → ∞ both CI get identical as PDF for estimator → Gauss-PDF f_{Gauss}

for finite/small n = N_{SP}

- correct Neyman CI longer
- coverage of naive CI smaller than claimed CLI

Basics of Hypothesis Tests

Null hypothesis H_0 : hypothesis which you try to falsify / reject
(one can not verify / approve hypothesis)

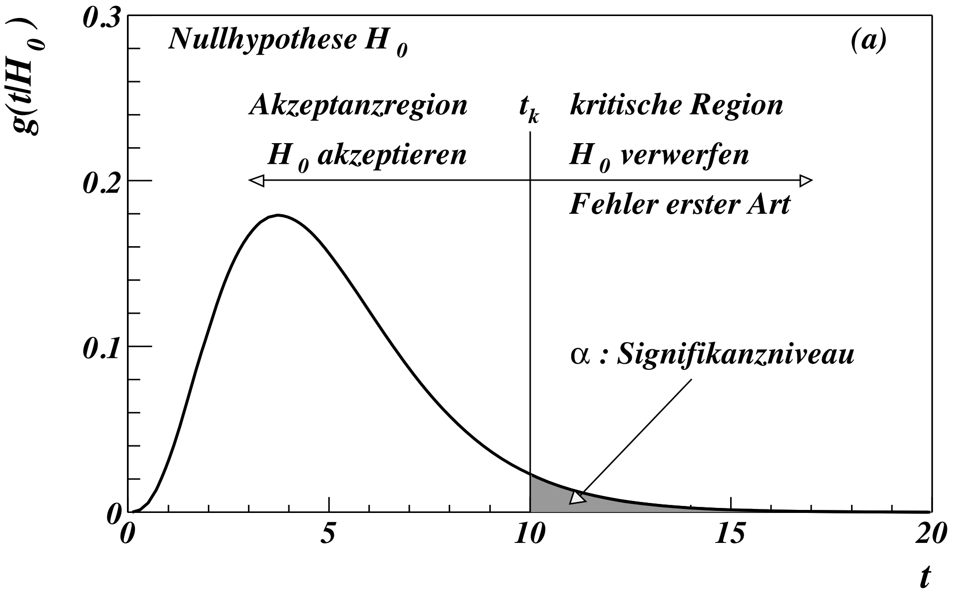
Test statistic t: any function of your data which is used to quantify (dis-)agreement with H_0

$g(t|H_0)$: probability density function PDF for test statistics under null hypothesis H_0

Critical region: range of test statistic for which H_0 is rejected

α: significance (level)
size of test
error of 1st kind.
probability to reject H_0, if H_0 is true

$$\alpha = \int_{t_k}^{\infty} g(t|H_0)dt.$$



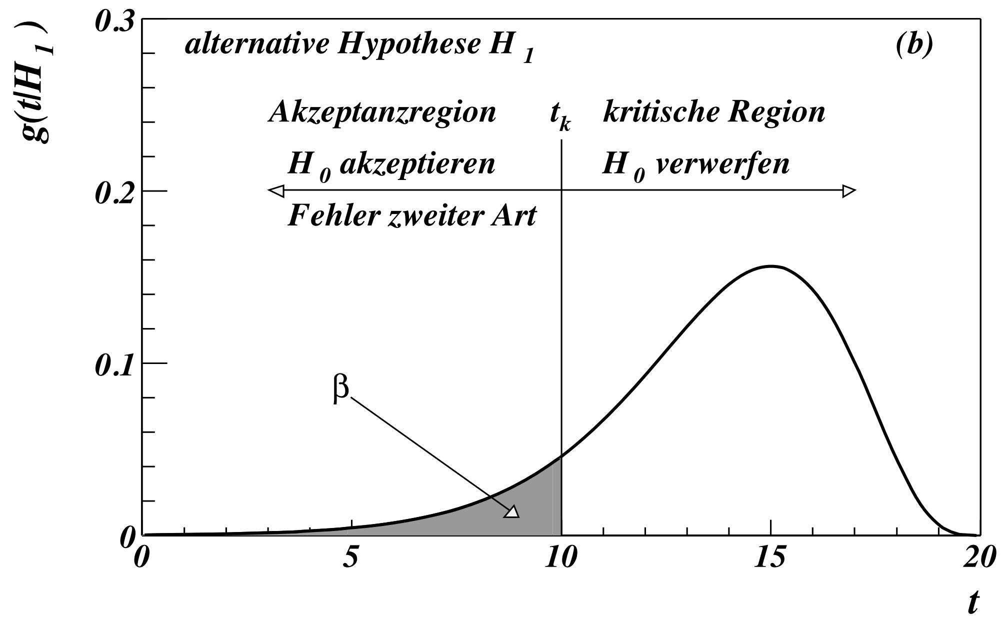
Basics of Hypothesis Tests (2)

In principle: infinity many possibilities to choose critical region for given α
(especially for one sided tests you need an alternative hypothesis to decide what you call inconsistent with null hypothesis)

Alternative hypothesis H_1 : hypothesis which you would like to approve

$g(t|H_1)$: probability density function for test statistics under alternative hypothesis H_1

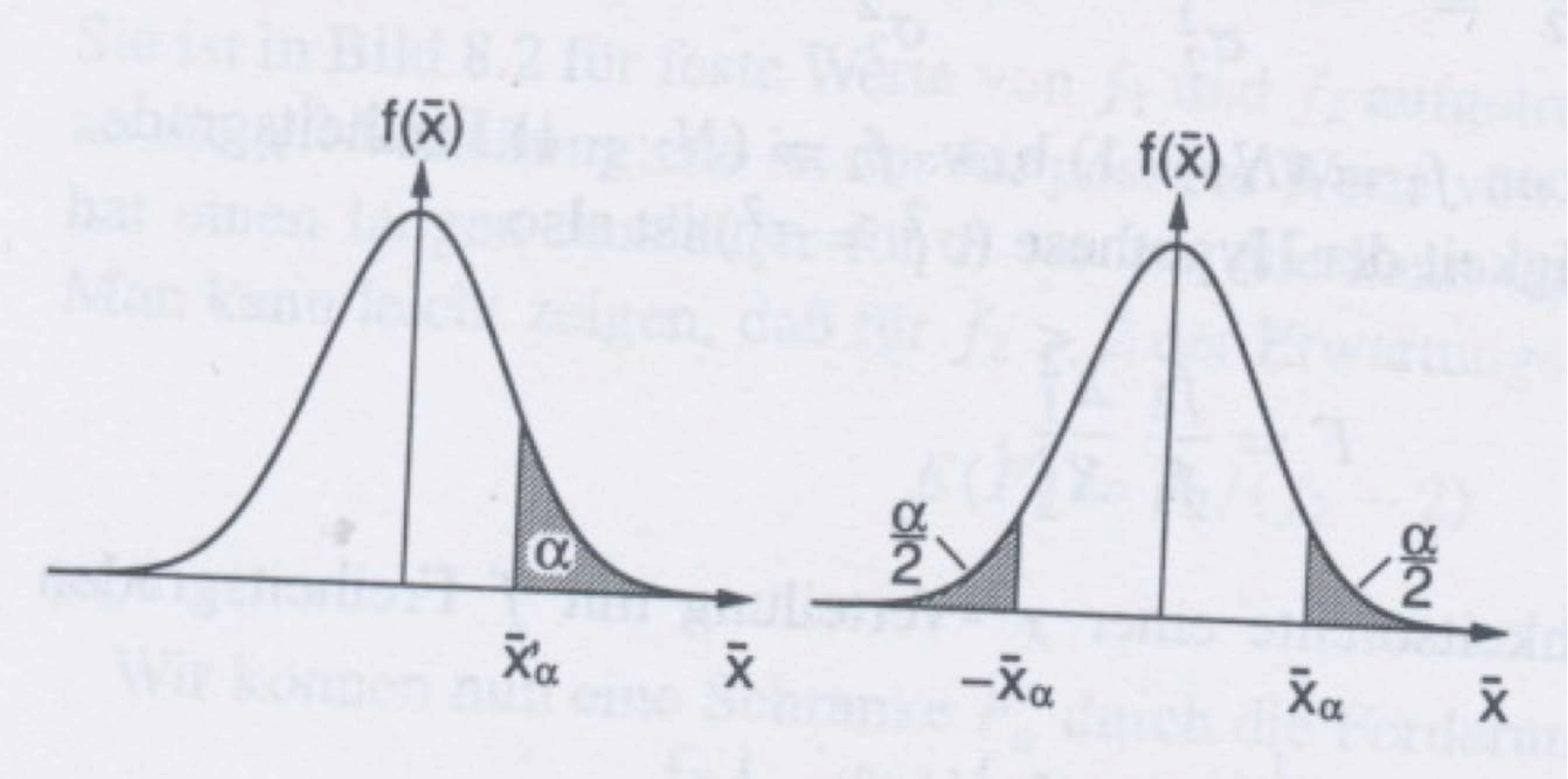
$$\beta = \int_{-\infty}^{t_k} g(t|H_1)dt.$$



β: error of 2[nd] kind
$M=1-\beta$: power

β prob. to reject H_1, if H_1 is true
$1-\beta$ prob to "accept" H_1, if H_1 is true

One- and Twosided Tests

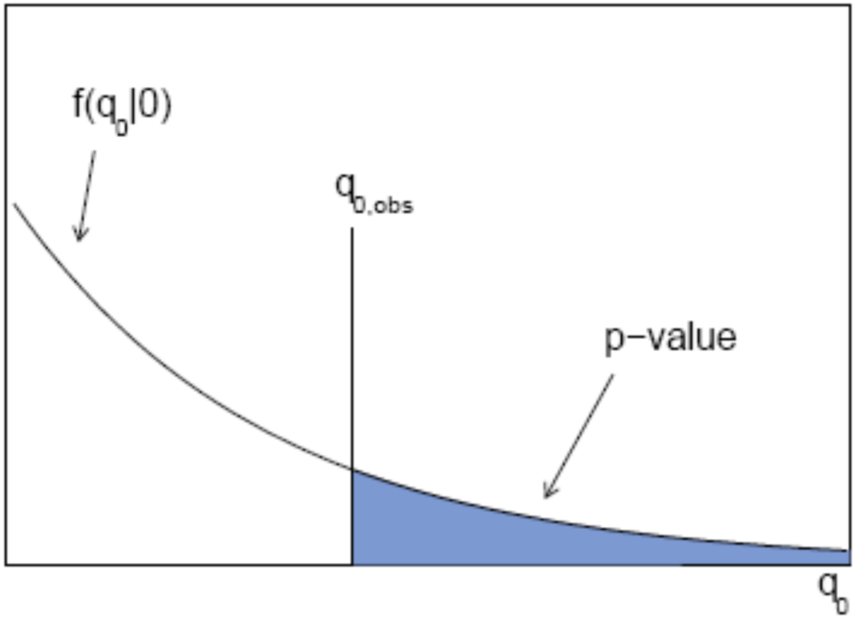


Depending on problem deviation in one or two directions are considered as incompatible with null hypothesis

→ one or two-sided test and critical region
→ if two-sided distribute significance α on both regions (mostly $\alpha/2$)

P-Value

P-value: probability to observe a data set, which is as consistent or less with null hypothesis as the actual observation



Test statistic: q_0
PDF for q_0 under H_0: $f(q_0|0)$
Critical region: large values of q_0
$q_{0,obs}$: observed value in data

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0$$

P-value is random variable (c.f. significance level α fixed before measurement)

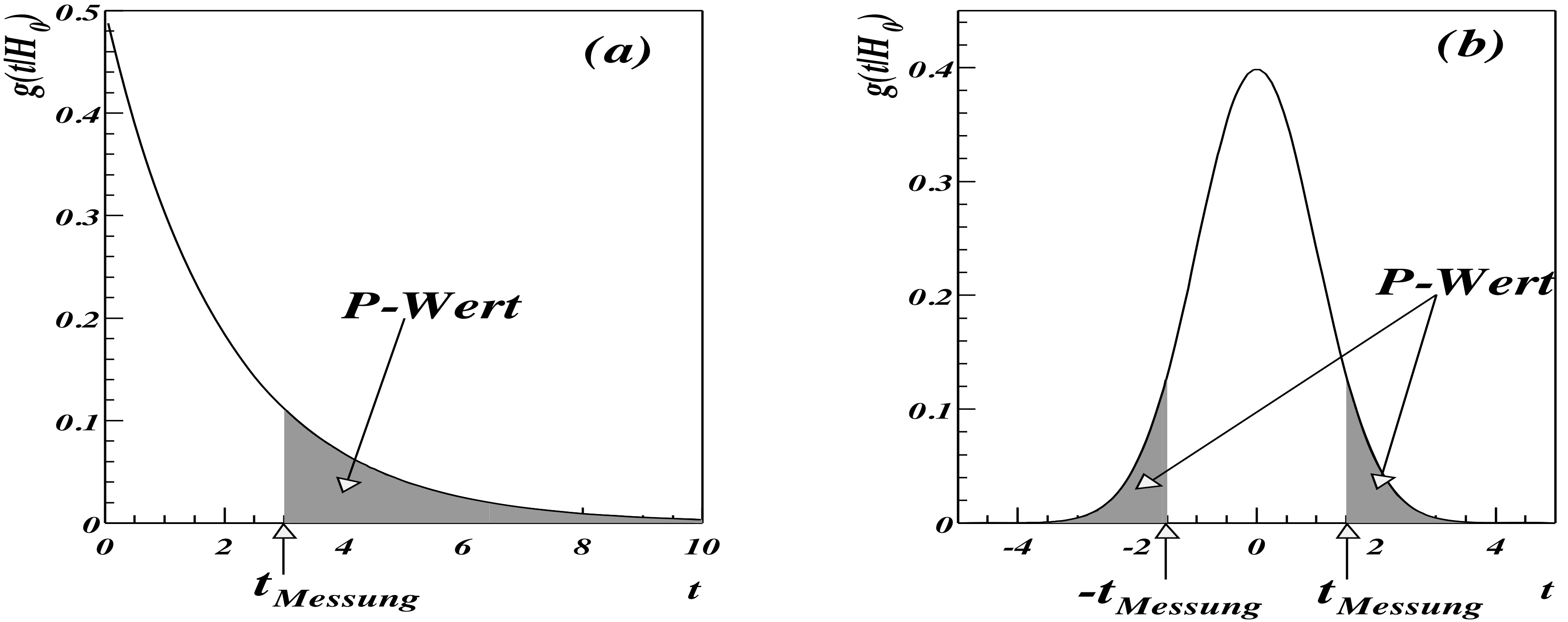
if P-value = significance level α, then $q_{obs.} = q_{critical}$

if P-values less then significance level α then reject null hypothesis

1-P-value = confidence level of the tests

Beware of wrong interpretation: P-value is not probability, that H_0 is wrong
1-P-value is not probability, that H_0 is true

One and Two sided P-Values

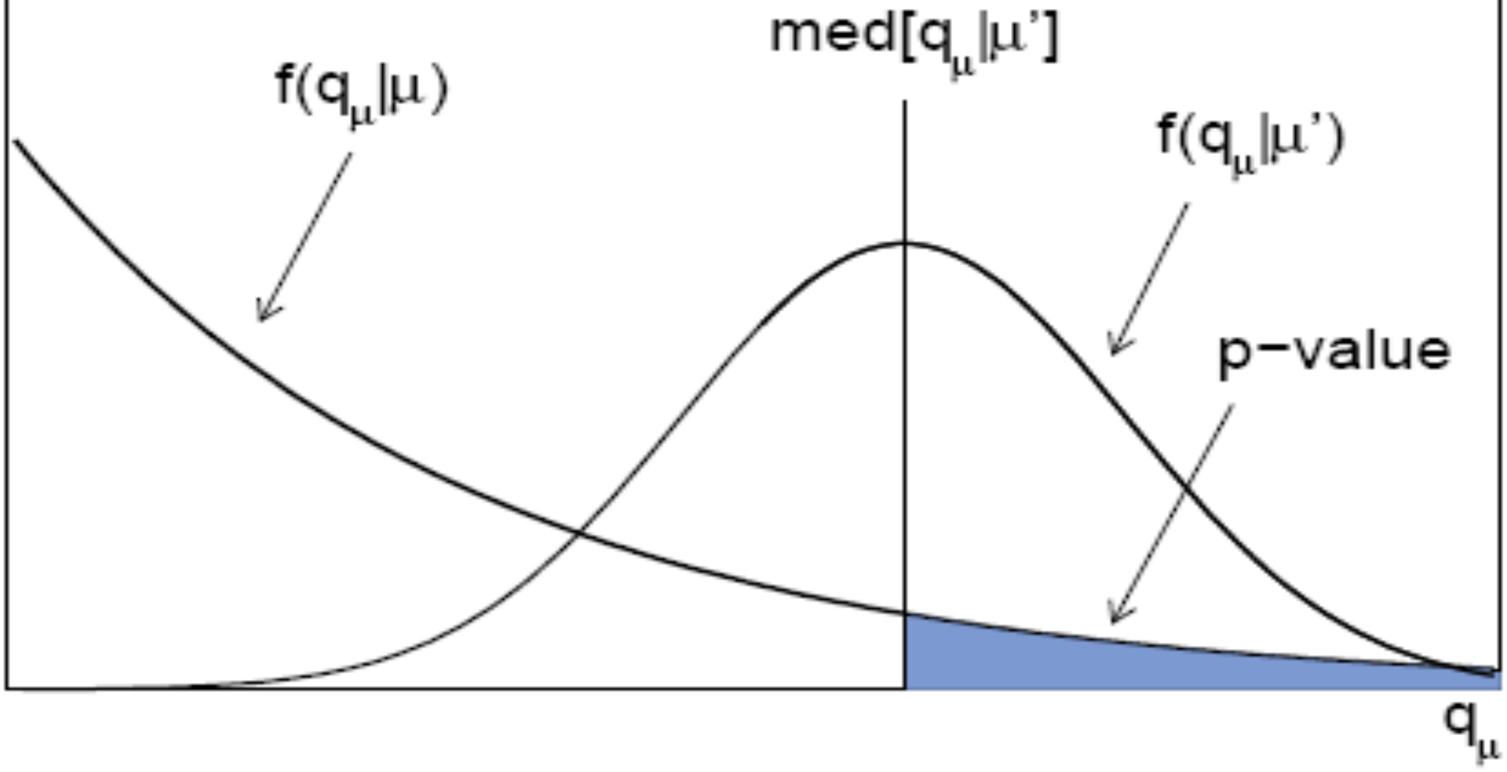


t=0 for perfect agreement between data and H_0

left: one-sided P-vlaue   right two-sided P-value

Expected P-Value / Sensitivity

Often interested in sensitivity of experiment:
evaluate p-value under null hypothesis (μ)
from median value of test statistic under alternative hypothesis (μ')



Example: Test for Mean Value of Gaussian PDF

Null Hypothesis: mean value $\lambda=\lambda_0$

Data set of size n (for illustration =2): $x_1, x_2, \ldots$

Test statistic: maximum likelihood estimate
= arithmetic mean

$$x = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

with PDF given by Gauss
with mean λ_0 und Variance σ^2/n

$$f(x;\lambda_0) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{n}{2\sigma^2}(x-\lambda_0)^2\right)$$

Choice of 4 different critical regions with same significance α

two sided in tails $\quad U_1 : x < \lambda^{\mathrm{I}} \text{ und } x > \lambda^{\mathrm{II}} \quad \text{mit } \int_{-\infty}^{\lambda^{\mathrm{I}}} f(x)\,\mathrm{d}x = \int_{\lambda^{\mathrm{II}}}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}\alpha\,;$

one-sided in upper tail $\quad U_2 : x > \lambda^{\mathrm{III}} \quad \text{mit } \int_{\lambda^{\mathrm{III}}}^{\infty} f(x)\,\mathrm{d}x = \alpha\,;$

one-sided in lower tail $\quad U_3 : x < \lambda^{\mathrm{IV}} \quad \text{mit } \int_{-\infty}^{\lambda^{\mathrm{IV}}} f(x)\,\mathrm{d}x = \alpha\,;$

two-sided in center $\quad U_4 : \lambda^{\mathrm{V}} \leq x < \lambda^{\mathrm{VI}} \quad \text{mit } \int_{\lambda^{\mathrm{V}}}^{\lambda_0} f(x)\,\mathrm{d}x = \int_{\lambda_0}^{\lambda^{\mathrm{VI}}} f(x)\,\mathrm{d}x = \frac{1}{2}\alpha\,.$

An Example: Test for Mean Value of Gaussian PDF

Rows: 4 critical regions

two sided in tails
one-sided in upper tail
one-sided in lower tail
two-sided in center

$$U_1 : x < \lambda^{\mathrm{I}} \text{ und } x > \lambda^{\mathrm{II}} \quad \text{mit } \int_{-\infty}^{\lambda^{\mathrm{I}}} f(x)\,\mathrm{d}x = \int_{\lambda^{\mathrm{II}}}^{\infty} f(x)\,\mathrm{d}x = \tfrac{1}{2}\alpha\,;$$

$$U_2 : x > \lambda^{\mathrm{III}} \quad \text{mit } \int_{\lambda^{\mathrm{III}}}^{\infty} f(x)\,\mathrm{d}x = \alpha\,;$$

$$U_3 : x < \lambda^{\mathrm{IV}} \quad \text{mit } \int_{-\infty}^{\lambda^{\mathrm{IV}}} f(x)\,\mathrm{d}x = \alpha\,;$$

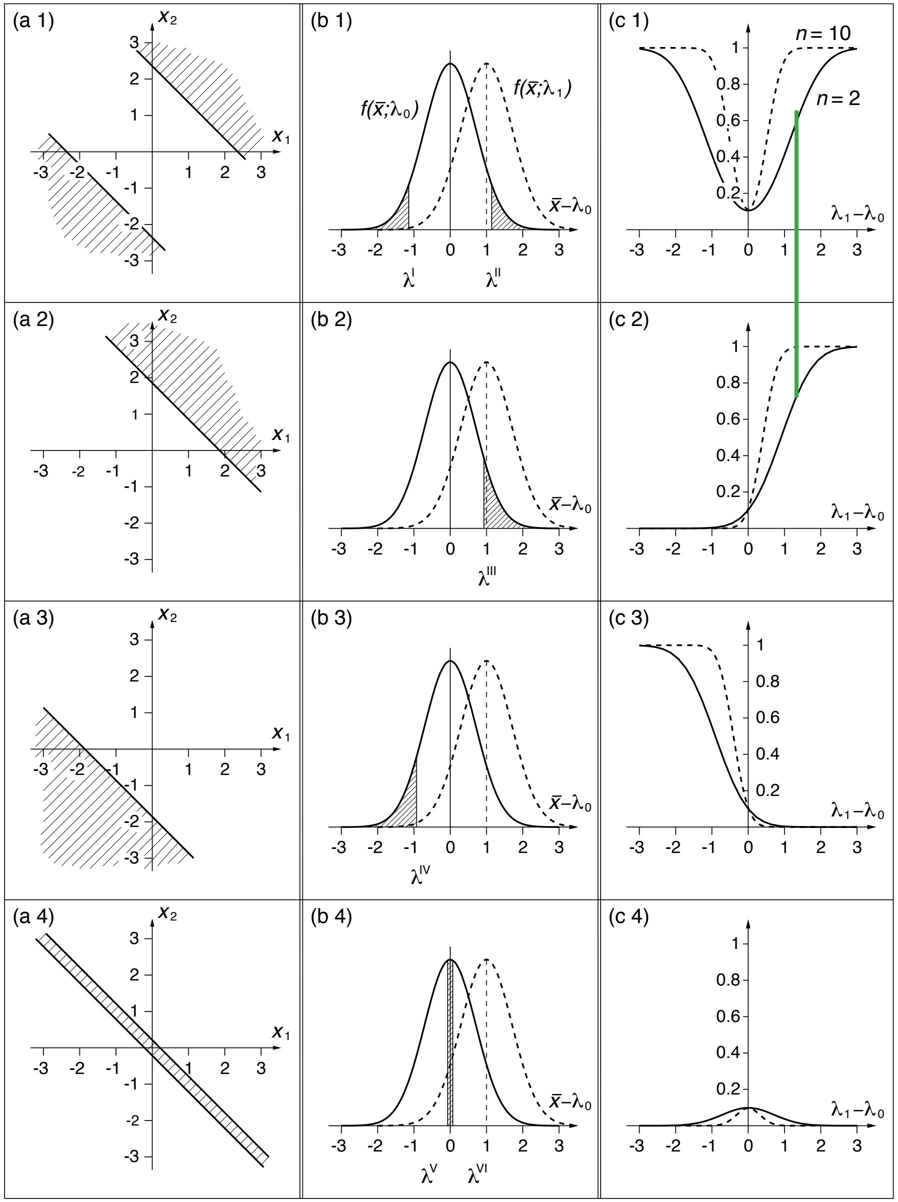
$$U_4 : \lambda^{\mathrm{V}} \leq x < \lambda^{\mathrm{VI}} \quad \text{mit } \int_{\lambda^{\mathrm{V}}}^{\lambda_0} f(x)\,\mathrm{d}x = \int_{\lambda_0}^{\lambda^{\mathrm{VI}}} f(x)\,\mathrm{d}x = \tfrac{1}{2}\alpha\,.$$

Left column: critical region for n=2
in data set space

Middle column: PDF for test statistics
for H_0 and H_1 with critical regions

$$\lambda = \lambda_1 = \lambda_0 + 1$$

Right column: power for n=2 and n=10
depending on λ_1-λ_0



S. Brandt

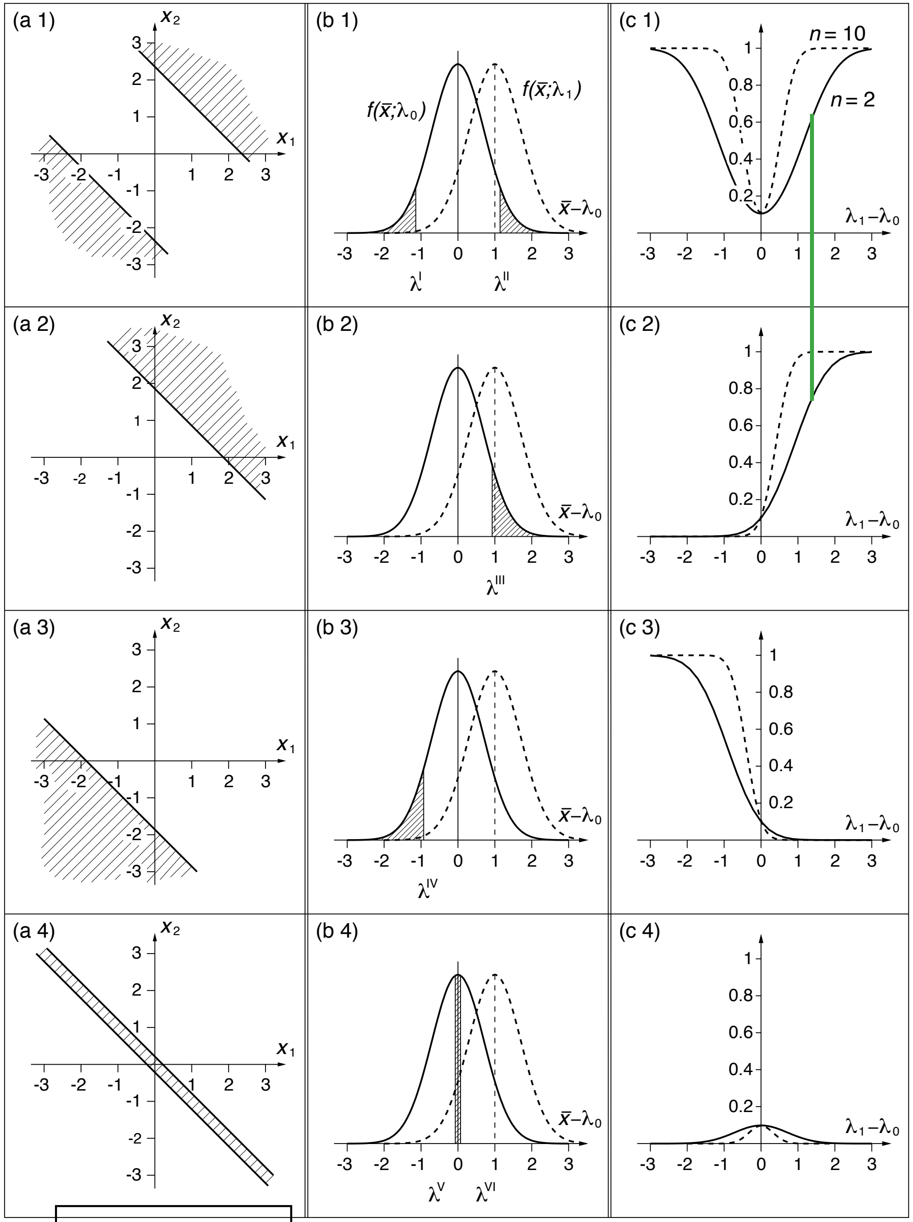
Example: Test for Mean Value of Gauss PDF

U_1 power ≥ significance for all λ
two sided test from ratio
of profiled likelihoods
for H_1: $\lambda = \lambda_1 \neq \lambda_0$

U_2: larger power for $\lambda_1 > \lambda_0$
one sided test from NPL
for H_1: $\lambda = \lambda_1 > \lambda_0$

U_3: larger power for $\lambda_1 < \lambda_0$
one sided test from NPL
for H_1: $\lambda = \lambda_1 < \lambda_0$

U_4: no useful test
maximal power for $\lambda_1 = \lambda_0$



S. Brandt

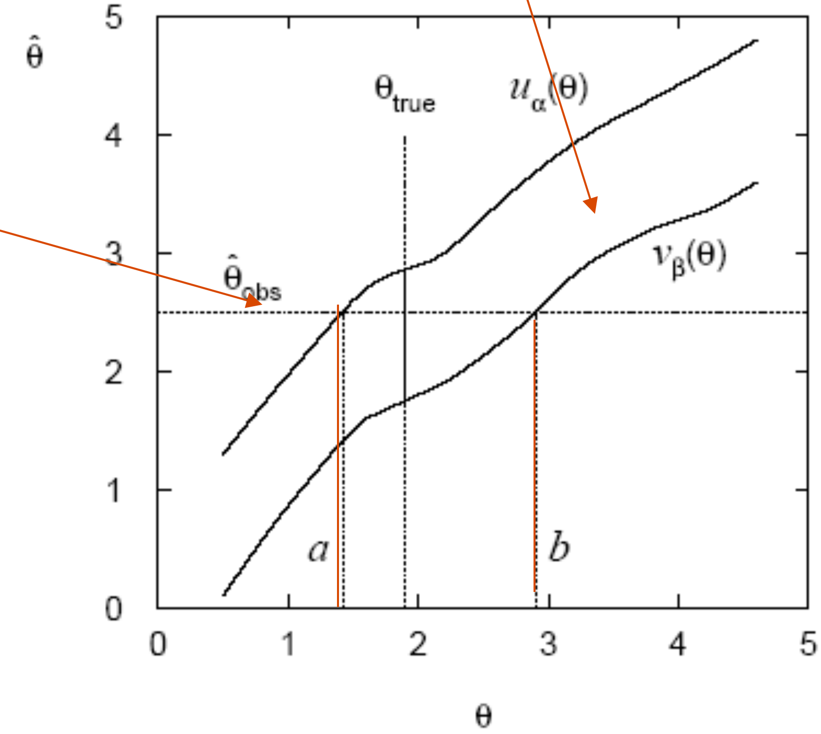
Classical Frequentist Intervals

Region btw. $u_\alpha(\theta)$ and $v_\beta(\theta)$ is the confidence belt

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

Boundaries of confidence interval given by intersect of observed value with confidence belt → [a,b]

$$a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta})$$
$$b(\hat{\theta}) \equiv l_\beta^{-1}(\hat{\theta}).$$



For all possible true values θ holds:

$$\hat{\theta} \geq u_\alpha(\theta) \Leftrightarrow a(\hat{\theta}) \geq \theta \qquad P(a(\hat{\theta}) \geq \theta) = \alpha$$
$$\hat{\theta} \leq l_\beta(\theta) \Leftrightarrow b(\hat{\theta}) \leq \theta, \qquad P(b(\hat{\theta}) \leq \theta) = \beta$$

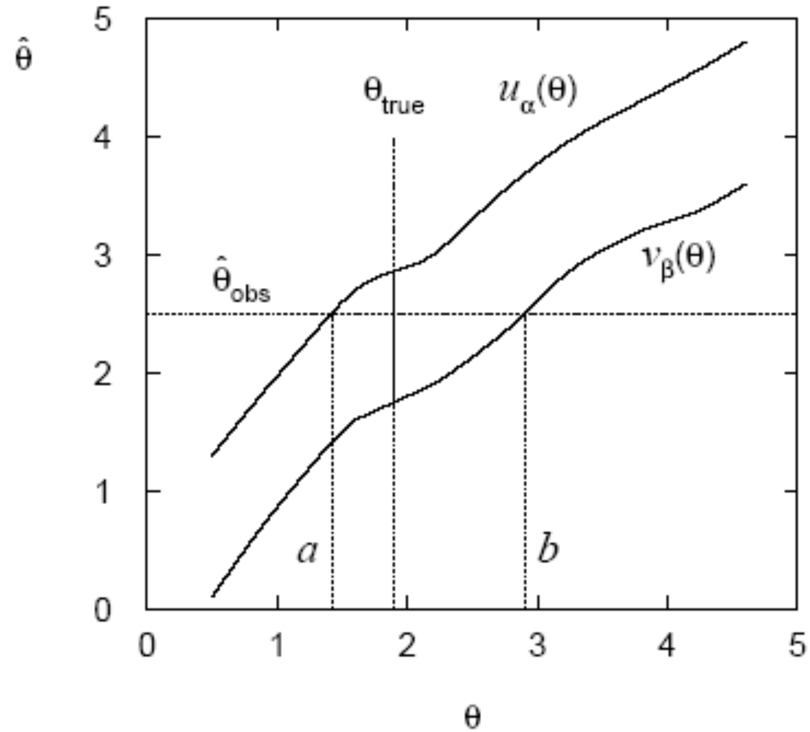
$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

Correct coverage by construction

CI from Inversion of Hypothesis Test

The Confidence belt is the acceptance region of all possible hypothesis tests.

CI for a parameter θ :
find all true hypothetical values θ which are not rejected in a test of size 1-CL given the observed value θ_{obs}



An upper limit b for θ is the smallest value for which holds $p_\theta \geq \gamma$.

In practical life: for given sizes / tail probabilites α and β find largest a and smallest b, fulfilling the equations:

$$\alpha = \int_{\hat{\theta}_{obs}}^{\infty} g(\hat{\theta}; a) d\hat{\theta} = 1 - G(\hat{\theta}_{obs}; a)$$

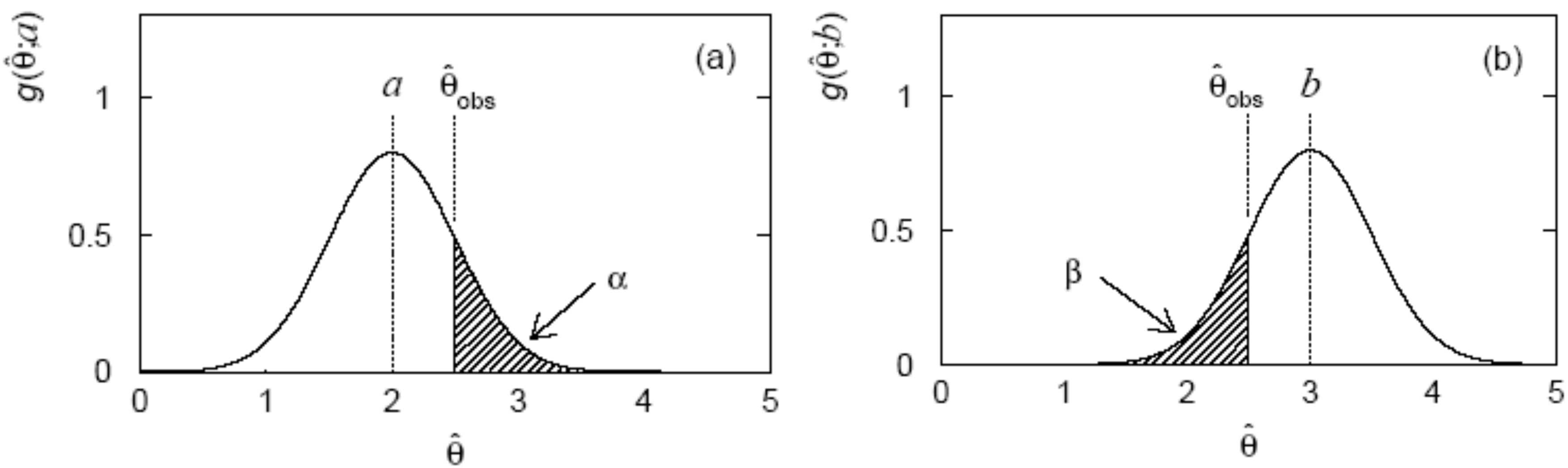
$$\beta = \int_{-\infty}^{\hat{\theta}_{obs}} g(\hat{\theta}; b) d\hat{\theta} = G(\hat{\theta}_{obs}; b).$$

Determination of CI

The recipe to find [*a*, *b*] reduces to solve

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

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



→ *a* is max. hypothetical value of θ for which $P(\hat{\theta} > \hat{\theta}_{\text{obs}}) = \alpha\,.$

→ *b* is min. hypothetical value of θ for which $P(\hat{\theta} < \hat{\theta}_{\text{obs}}) = \beta\,.$

CI for Estimator in Gaussian PDF

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

Very simple if variance known and constant:

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

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

Solved by:

$$a \;=\; \hat\theta_{obs} - \sigma_{\hat\theta}\Phi^{-1}(1-\alpha)$$

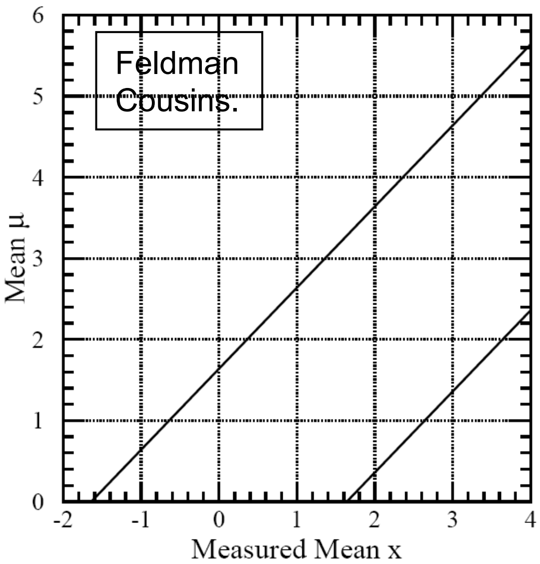
$$b \;=\; \hat\theta_{obs} + \sigma_{\hat\theta}\Phi^{-1}(1-\beta).$$

For $\alpha=\beta= 0.16$

1-σ Intervall

$$[\theta - \sigma_{\hat\theta}, \theta + \sigma_{\hat\theta}]$$

Confidence belt for σ=1 at 90 % CI



CI =
$\mu \pm 1{,}64\sigma$

FIG. 3. Star
Gaussian, in uni

Two sided | 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
4	$1-6.3\times10^{-5}$		
5	$1-5.7\times10^{-7}$		
6	$1-2.0\times10^{-9}$		

CI for Estimator in Gaussian PDF

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

Very simple if variance known and constant:

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

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

Solved by:

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

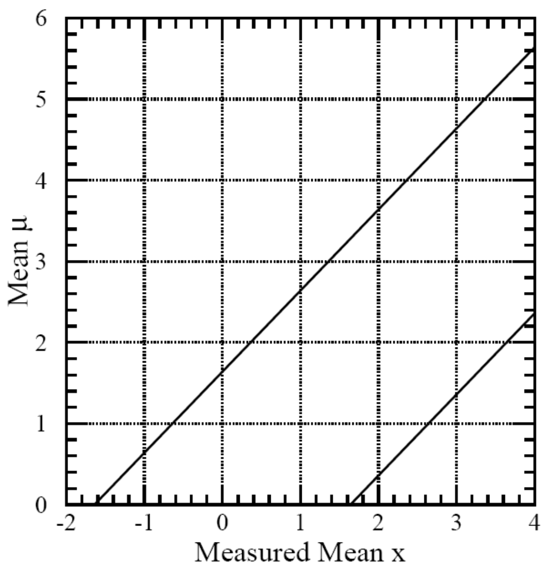
$$b = \hat{\theta}_{obs} + \sigma_{\hat{\theta}}\Phi^{-1}(1-\beta).$$

For $\alpha=\beta= 0.16$

1-σ Intervall

$$[\theta - \sigma_{\hat{\theta}}, \theta + \sigma_{\hat{\theta}}]$$

Confidence belt for $\sigma=1$ at 90 % CI



Two sided | 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
0.999	3.29		
0.9999	3.89		

CI at Physical Boundary

Gaussian estimator with known variance   allowed range: true value $\theta \geq 0$.

Classical Neyman construction yields upper limit:

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

example: observation = -2 variance = 1 ; CL = 95%

→ b= -2 + 1.645 = -0.355 CI „empty“ / completely in unphysical region

Frequentist: no problem. If true value is „0“, 5% of all CI should not contain „0“

Bayesian: not satisfactory. Worked for years, spent many Euros to get this answer.

Option 0: increase CL until upper limit > 0

CL = 99% → b = -2 + 2.36 = 0.326 b << resolution=1 → arbitrary

even worse: adjust CL for best limit CL = 97.725% → b= 10^{-5}

this option is not to be used!

CI at Physical Boundary: Solutions

Option 1: replace measurement by boundary value if measurement in unphysical region

- upper limit (CL ≥ 68%) > resolution
- for measurement above border identical to classical CI
- coverage 100% for measurement in unphysical region
  (equivalent to Power Constrained Limit with minimal power = 50%)

Option 2: Bayesian limit

$$P(\mu;x) = \frac{L(x;\mu)\pi(\mu)}{\int_{-\infty}^{+\infty} L(x;\mu)\pi(\mu)d\mu}$$

$$CL = 1-\alpha = \int_{-\infty}^{\mu_{up}} P(\mu;x)d\mu$$

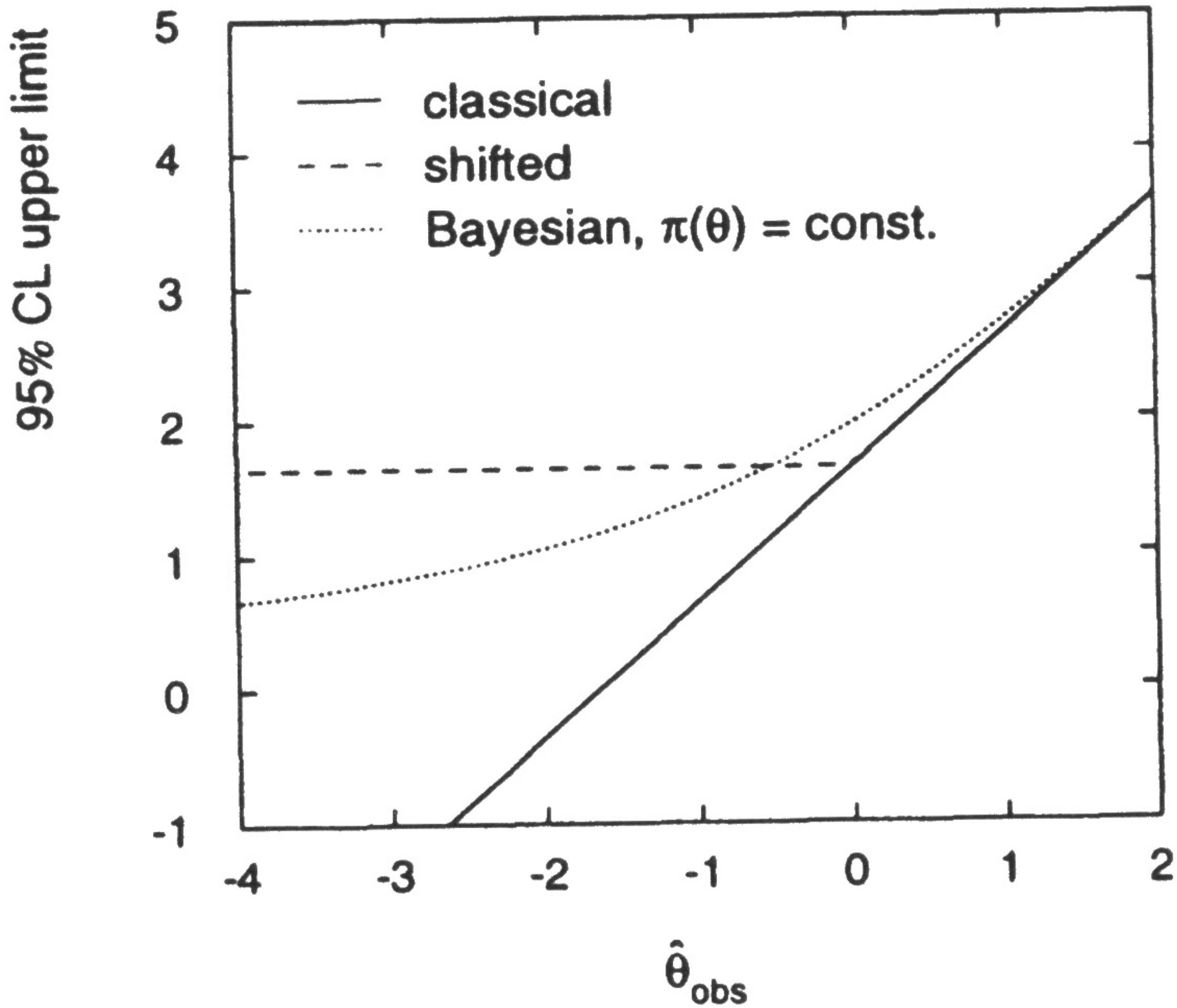
$$CL = 1-\alpha = \frac{\int_{-\infty}^{\mu_{up}} L(x;\mu)\pi(\mu)d\mu}{\int_{-\infty}^{+\infty} L(x;\mu)\pi(\mu)d\mu}$$

Implement physical boundary via $\pi(\mu)$: $\pi(\mu) = 0$ in forbidden region
mostly: $\pi(\mu)$ = const else

Integrate posterior-PDF $P(\mu \mid x)$ to get correct credibility

Coverage larger than quoted CL, but not goal of Bayesian method

Frequentist, Shifted and Bayesian Limit



upper limit for mean of Gauss PDF with Variance 1 at 95% CL

Bayesian Upper Limit for Gauss PDF

Condition for upper limit

$$CL = 1 - \alpha = \frac{\int_{-\infty}^{\mu_{up}} L(x;\mu)\pi(\mu)d\mu}{\int_{-\infty}^{+\infty} L(x;\mu)\pi(\mu)d\mu}$$

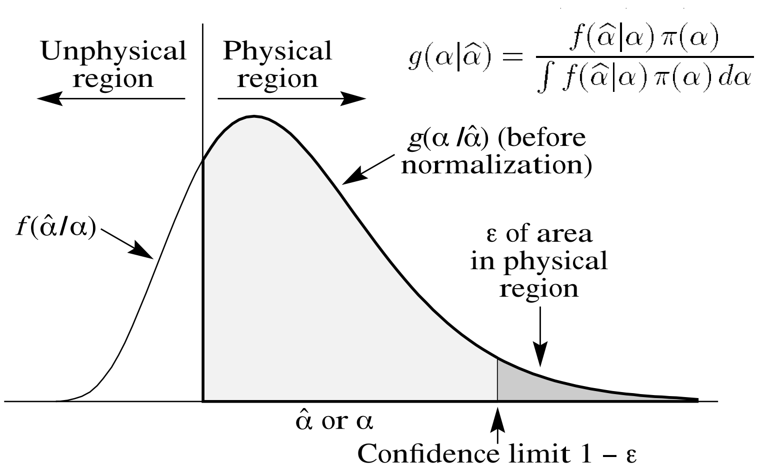
Likelihood function

$$L(x;\mu) = \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}}$$

A-priori probability

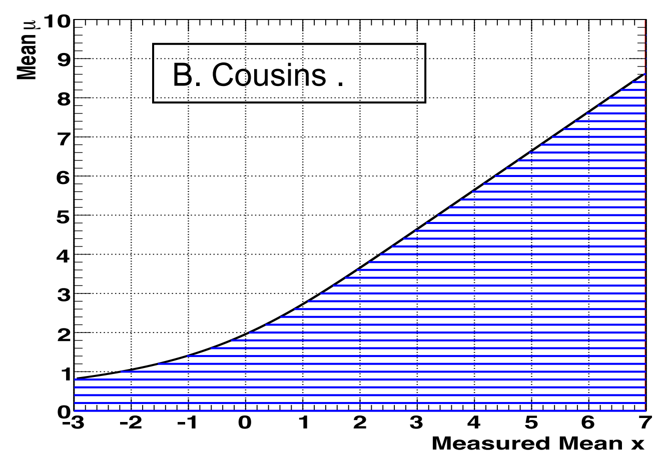
$$\begin{aligned} \pi(\mu) &= 0 \quad \text{for } \mu < 0 \\ &= const. \text{ for } \mu \geq 0 \end{aligned}$$

Yields ratio of two integrasl over Gauss PDF starting at physical boundary



$$CL = 1 - \alpha = \frac{\int_{0}^{\mu_{up}} \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}} d\mu}{\int_{0}^{+\infty} \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}} d\mu}$$

Bayesian upper limit at 95% CL

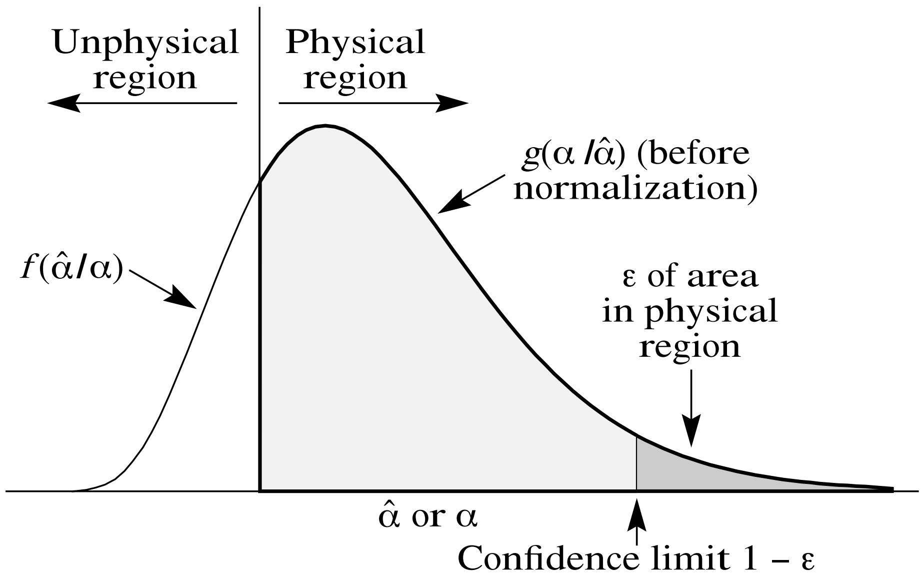


Upper limit always > 0

Coverage greater CL

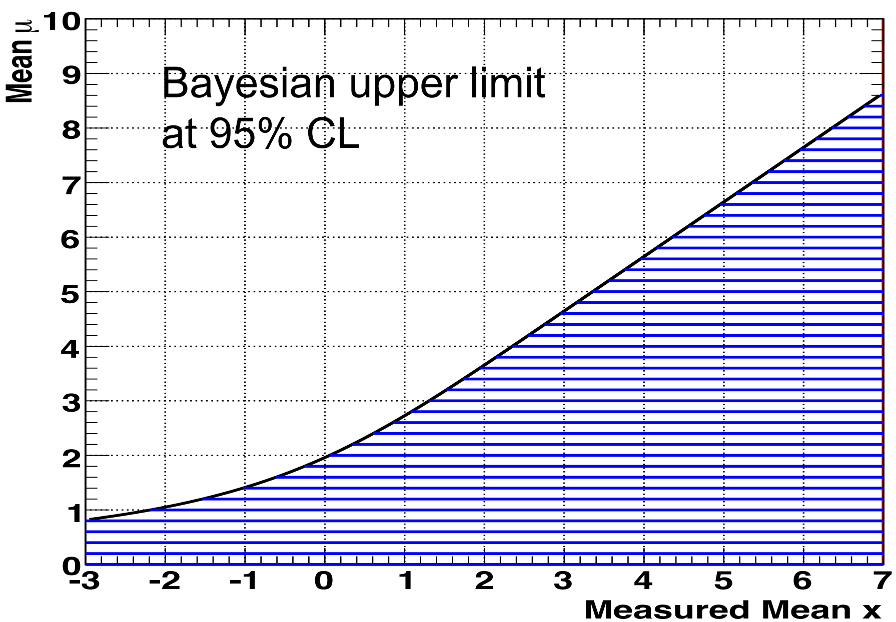
For large measured x approaching classical limit of x+1.64 (σ=1)

Bayesian Upper Limit for Gauss PDF



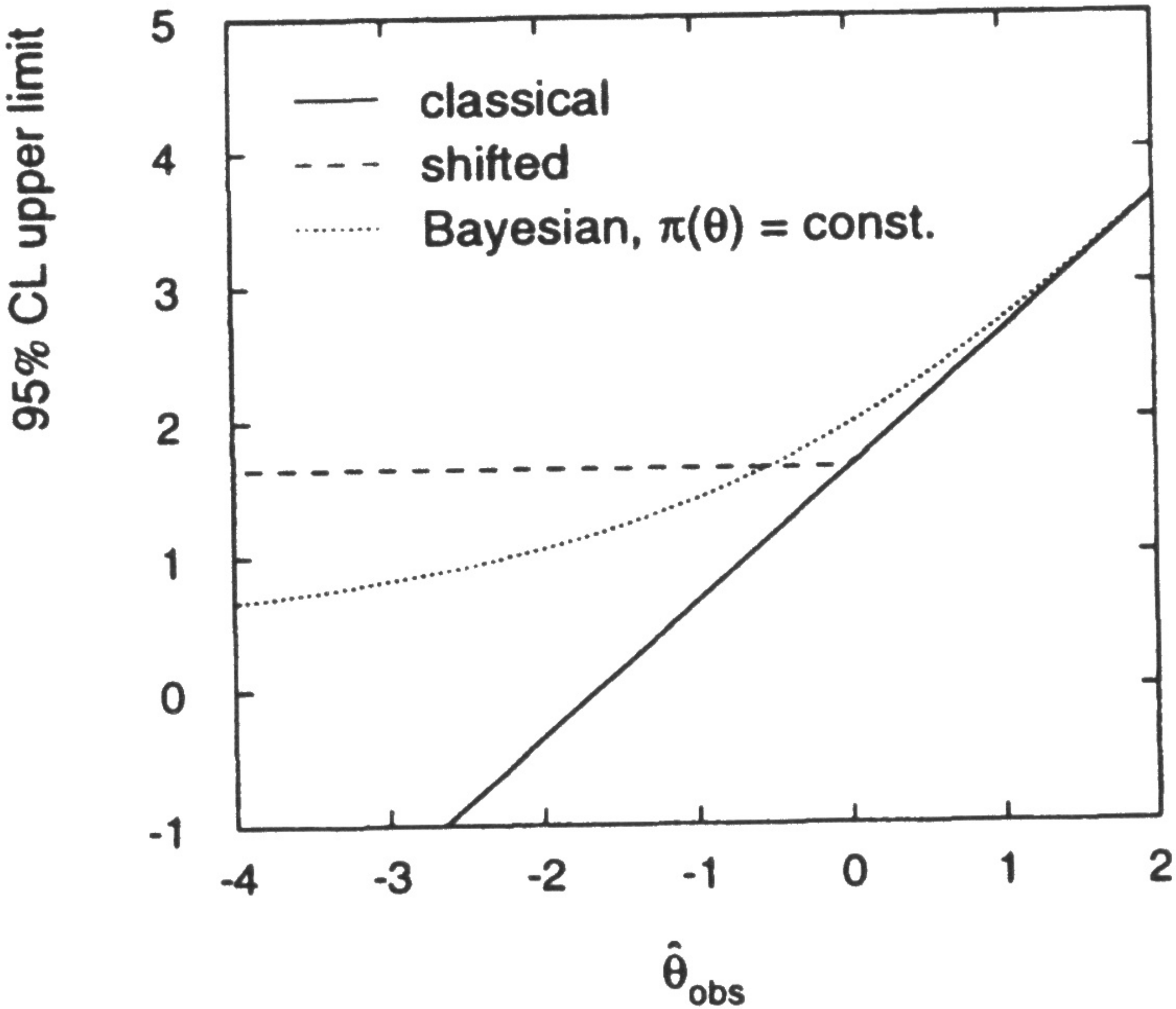
$$g(\alpha|\widehat{\alpha}) = \frac{f(\widehat{\alpha}|\alpha)\,\pi(\alpha)}{\int f(\widehat{\alpha}|\alpha)\,\pi(\alpha)\,d\alpha}$$

Yields ratio of two integrals
over Gauss PDF
starting at physical boundary



Upper limit always > 0
Coverage greater CL
For large measured x approaching
classical limit of x+1.64 (σ=1)

Frequentist, Shifted and Bayesian Limit



upper limit for mean of Gauss PDF with Variance 1 at 95% CL

The "problem" with the classical Freq. Method

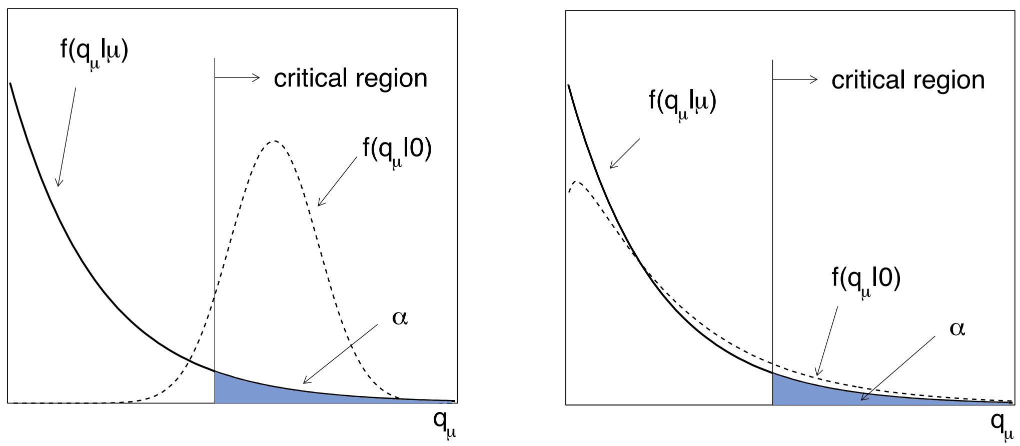
$$p_\mu \;=\; P(\,\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}\,|\,\text{signal+background}) \;=\; \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})\,d\tilde{q}_\mu$$

Pure frequentist would stop and say: „signal + background" hypothesis is excluded with a confidence level CL_{S+B} of 1- p_μ

„Problem": Spurious exclusion of hypothesis (signal) with no sensitivity (s<<b)

large s

power M= 1-ß large w.r.t. significance level α



s<<b

power M = 1-ß ~ significance level α

signal+BG-like ← → BG only like,

By construction: probability to reject μ if μ is true is α

for s<<b probability to reject very small μ if μ=0 is true ~ α + epsilon

→ probability to exclude hypotheses with zero signal (due to downwards fluctuation) ~ α „spurious exclusion w/o sensitivity"

CL_S for Continuous Random Variable

$$CL_S = \frac{\text{P-value}(\mu)}{1-\text{P-value}(\mu=0)} = \frac{\text{P-value}(\mu)}{\text{Power}(\mu=0\,\text{vs.}\,\mu)} = \frac{P(x \le x_{obs};\mu)}{P(x \le x_{obs};\mu=0)}$$

A hypothesis is called excluded at confidence level CL if $CL_S \le 1 - CL$

Motivation for this "ad hoc" correction of P-value (A. Read 1997) later in lecture

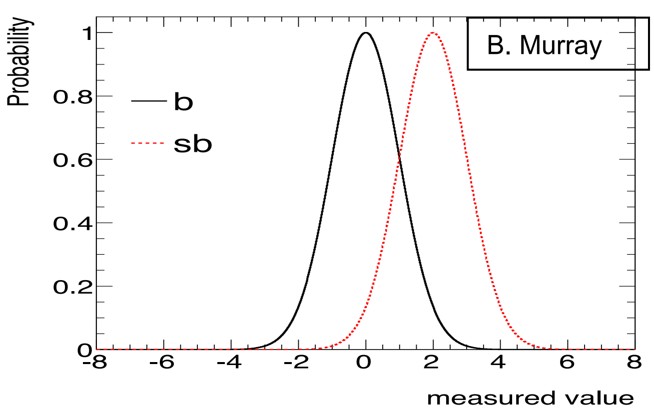
Gaussian example: small (large) value of x inconsistent with μ (μ=0) hypothesis

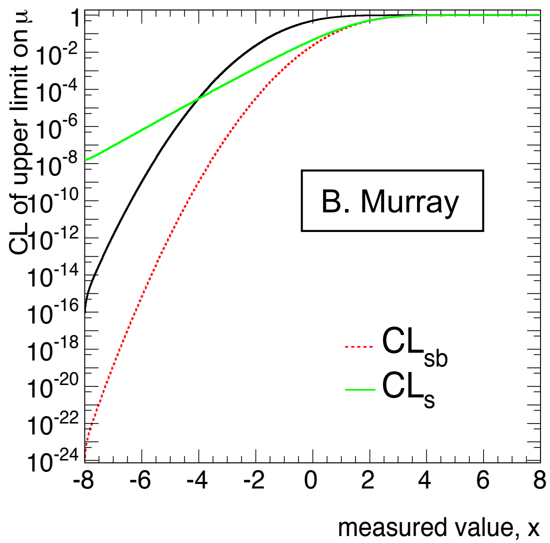
$$CL_S = \frac{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}} dx}{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-0)^2}{\sqrt{2\pi\sigma^2}}} dx}$$

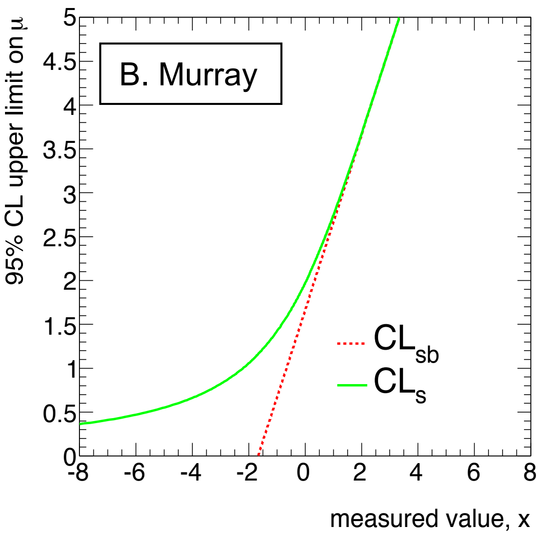
Numerically identical to Bayesian limit

CL_{SB} = P-value(μ)
black = 1-P-value(0)

CL_{SB} = classical limit
CL_S = CL_S limit

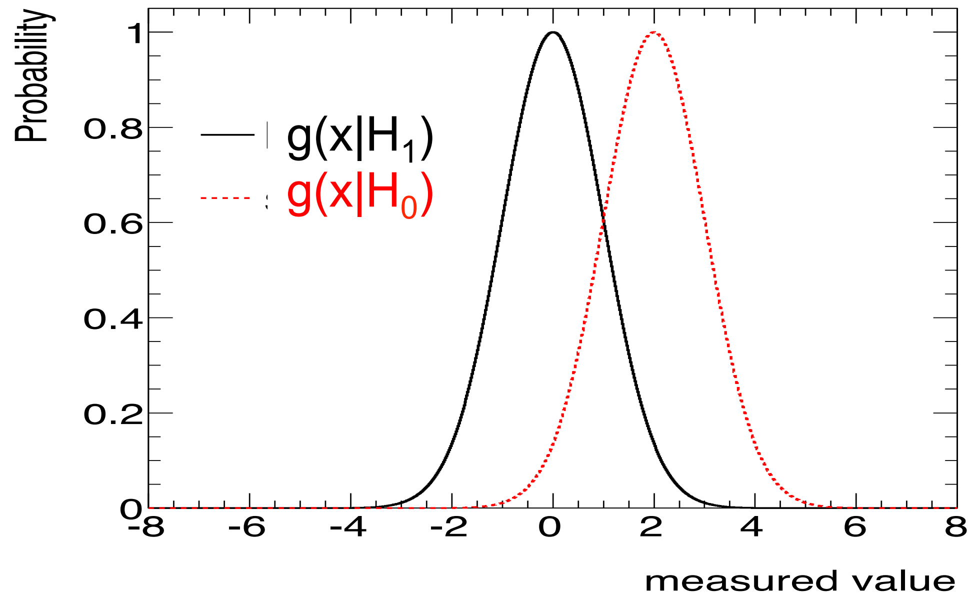


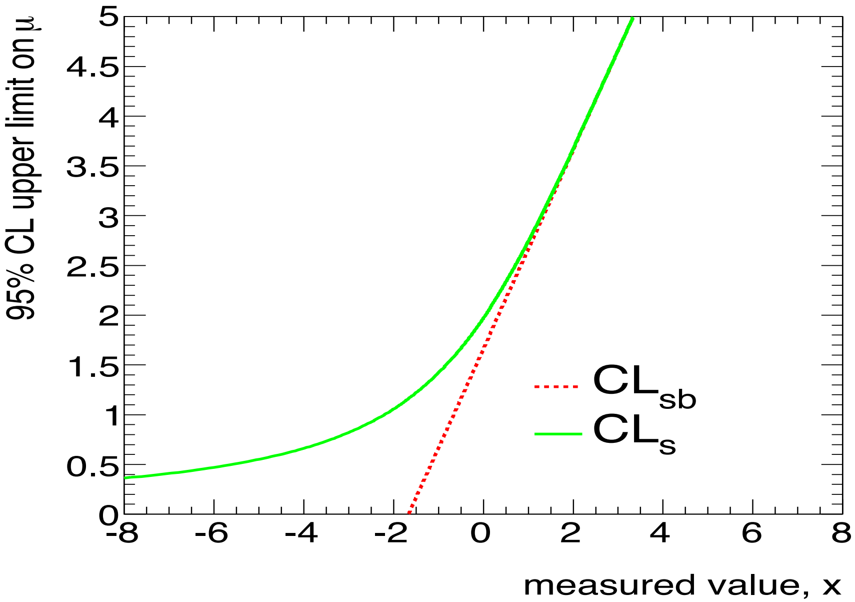




CL_S for Mean of Gaussian PDF (V[x] = 1)

$$CL_S = \frac{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}} dx}{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-0)^2}{\sqrt{2\pi\sigma^2}}} dx}$$

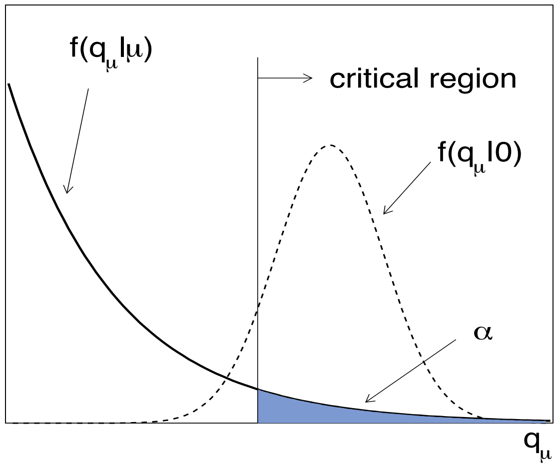




CL_{SB} = classical limit

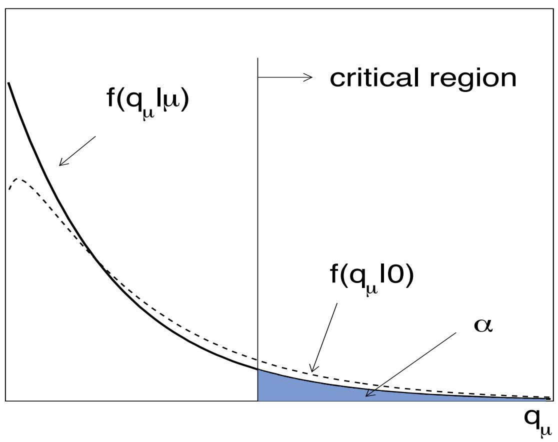
CL_S = CL_S limit

Constraint Limits (PCL) (Cowan et al. 2010)



Upper limit from inversion of hypothesis test
All values $\mu \geq \mu_{up}$ are called excluded

First normal condition for exclusion of a value of μ:
measurement x is in critical region (ω_μ) for a test of μ
or p-value for x is smaller than size of test α=1-CL



Supplemented by second condition:
sufficient sensitivity for discrimination of μ
from alternative hypothesis μ'=0
or power M=1-β of testing μ' vs $\mu \geq$ minimal value

Power M defined with critical region or via p-value w.r.t. μ

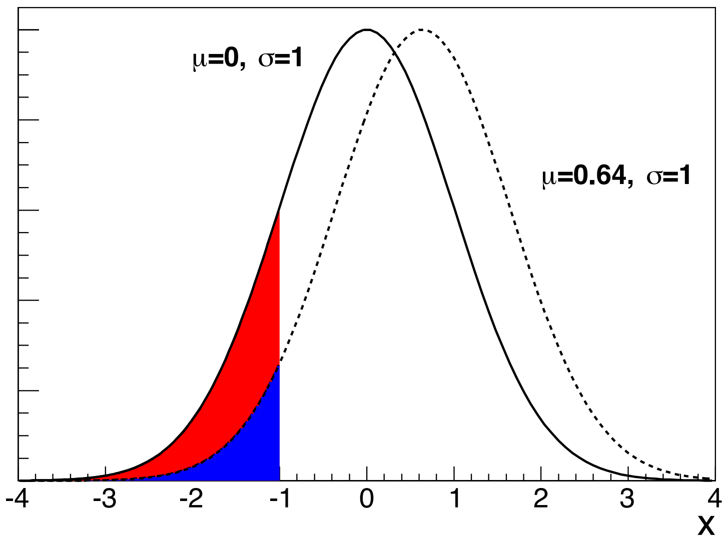
$$M_{\mu'}(\mu) = P(\mathbf{x} \in w_\mu|\mu')$$

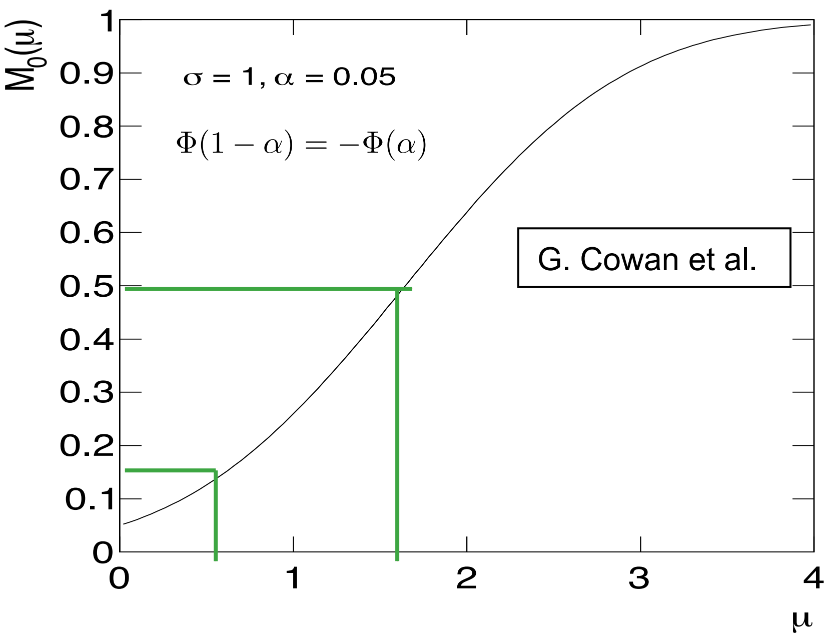
$$M_{\mu'}(\mu) = P(p_\mu < \alpha|\mu')$$

Procedure: determine "usual" upper limit μ_{up}
Find minimal μ value which has minimal power M_{min} μ_{min}
The PCL μ^*_{up} is then given by larger of the two: $\mu^*_{\text{up}} = \max(\mu_{\text{up}}, \mu_{\text{min}})$

For M_{min}=16% μ_{min} = "median expected – 1 σ" under hypothesis μ' = 0

PCL for Gauss-PDF with μ' = 0





x(critical) = - 1

red area: Power = 0.16

μ_{min} = 0.64 vs μ = 0

blue area: significance = 0.05

for μ_{min} = 0.64

$$M_0(\mu) = P\left(\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha)|0\right)$$

$$M_0(\mu) = \Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right)$$

← power of the test for μ w.r.t. μ'=0

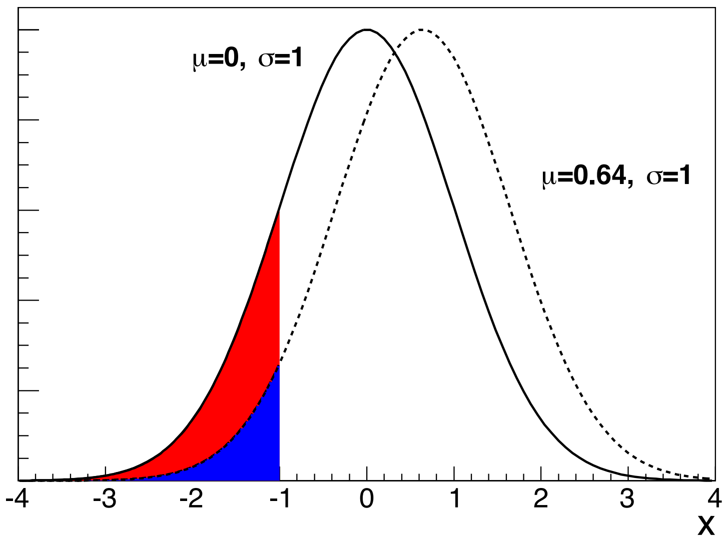
for α = 0.05 and σ = 1

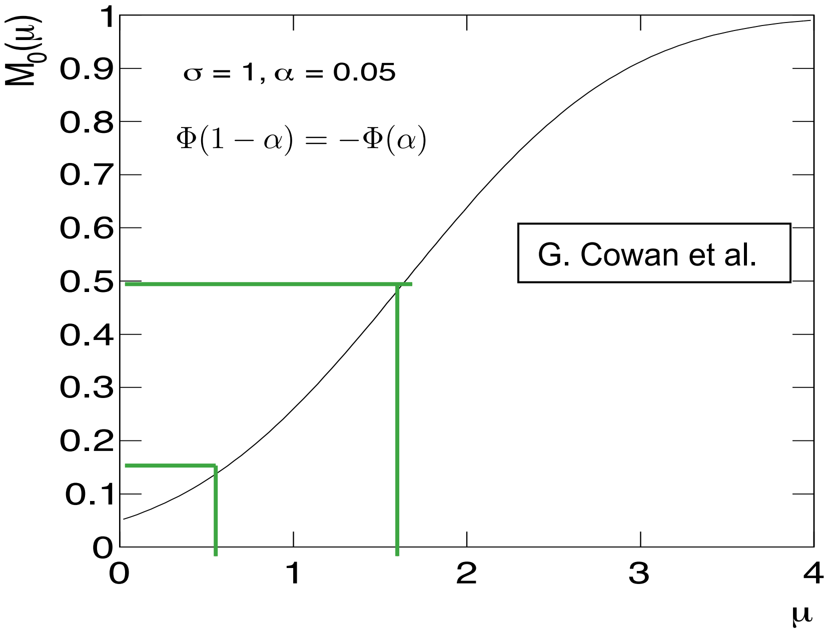
$$M_0(0) = \alpha \qquad M_0(\mu) > \alpha \text{ for all } \mu > 0$$

M_{min} = 16% → μ_{min} = 0.64

M_{min} = 50% → μ_{min} = 1.64

PCL for Gauss-PDF with μ' = 0





Critical region in a test of μ with size α

$$\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha)$$

The "usual" limit is then given by:

$$\mu_{\rm up} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha)$$

The power of the test for μ w.r.t. μ'=0

$$M_0(\mu) = P\left(\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha)|0\right)$$

$$M_0(\mu) = \Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right)$$

←power of the test for μ w.r.t. μ'=0 for α = 0.05 and σ = 1

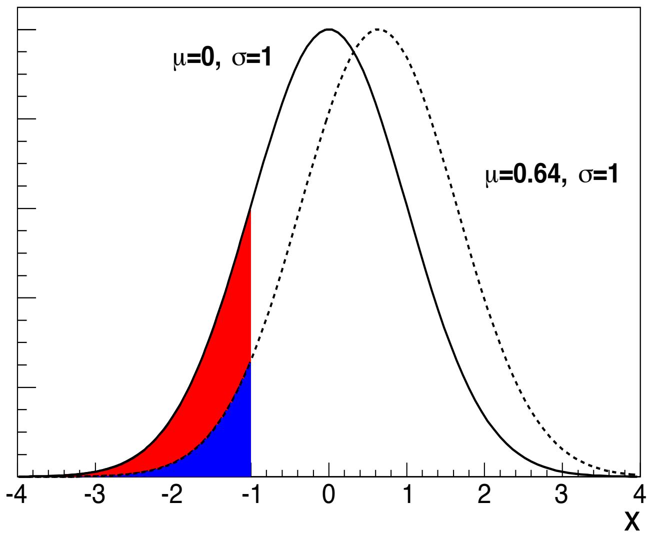
$$M_0(0) = \alpha$$

$$M_0(\mu) > \alpha \text{ for all } \mu > 0$$

M_{min}= 16% → μ_{min} = 0.64

M_{min} = 50% → μ_{min} = 1.64

PCL for Gauss-PDF with $\mu' = 0$

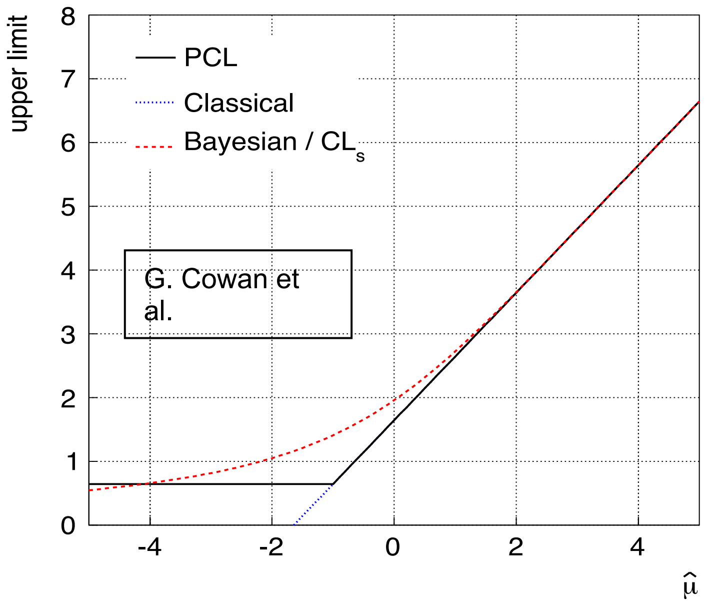


Requiring a minimal power $\Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right) \geq M_{\min}$

Minimal limit is: $\mu_{\min} = \sigma\left(\Phi^{-1}(M_{\min}) + \Phi^{-1}(1-\alpha)\right)$

Unconstrained limit: $\mu_{\mathrm{up}} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha)$

Replace normal limit if : $\hat{\mu} < \sigma\Phi^{-1}(M_{\min})$



PCL given by

$$\mu^*_{\mathrm{up}} = \begin{cases} \sigma\left(\Phi^{-1}(M_{\min}) + \Phi^{-1}(1-\alpha)\right) & \hat{\mu} < \sigma\Phi^{-1}(M_{\min}) \\ \hat{\mu} + \sigma\Phi^{-1}(1-\alpha) & \text{otherwise .} \end{cases}$$

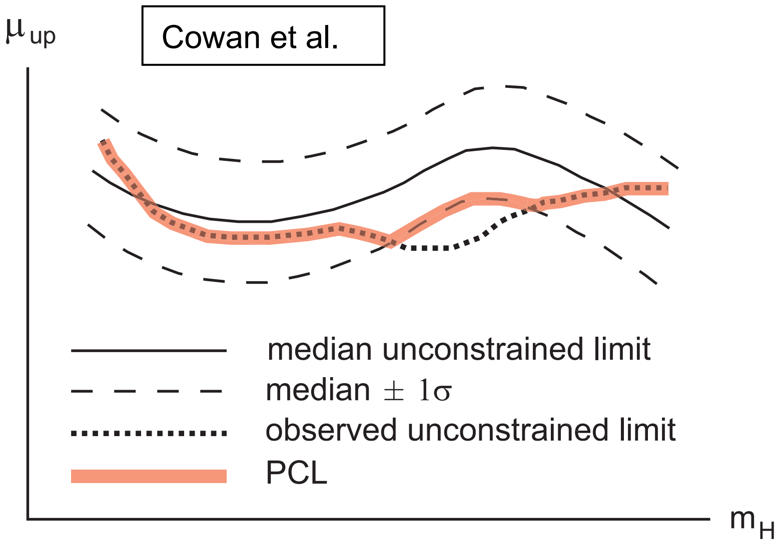
For $\alpha = 0.05$ $M_{\min} = 16\%$ $\sigma = 1$

$\mu_{up} = \mu_{meas} + 1{,}64$ $\mu_{min} = -1+1.64 = 0.64$

$\mu^*_{up} = \max(-1, \mu_{meas}) + 1.64$

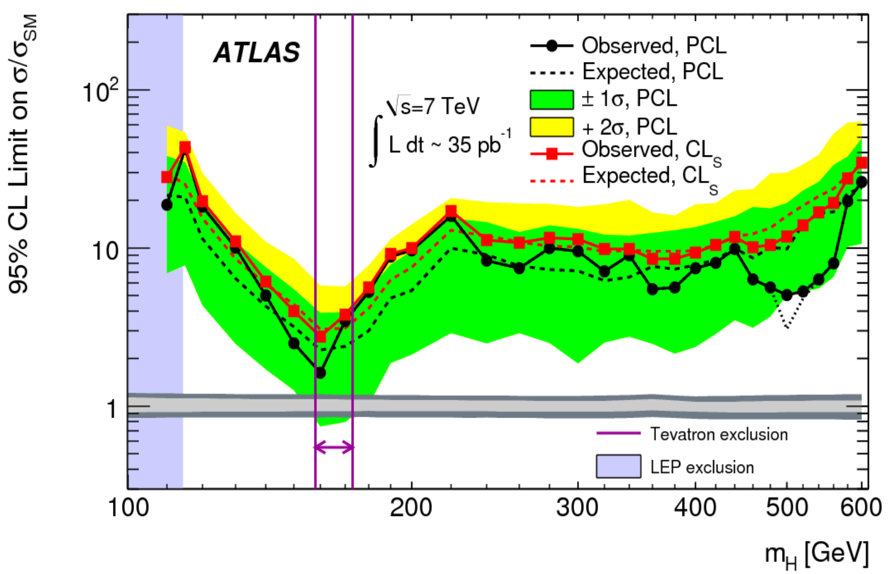
$$\alpha = 0.05 \text{ gives } \Phi^{-1}(1-\alpha) = 1.64$$

Power Constraint Limits at Work



$$\mu_{\rm up}^* = \begin{cases} \sigma\left(\Phi^{-1}(M_{\min}) + \Phi^{-1}(1-\alpha)\right) & \hat{\mu} < \sigma\Phi^{-1}(M_{\min}) \\ \hat{\mu} + \sigma\Phi^{-1}(1-\alpha) & \text{otherwise .} \end{cases}$$

for M_{min}=16%:
replace „observed“ classical limit by expected – 1 σ under H1 hypothesis if less than this value



PCL used in first ATLAS Higgs boson searches from 2010 data at 7 TeV

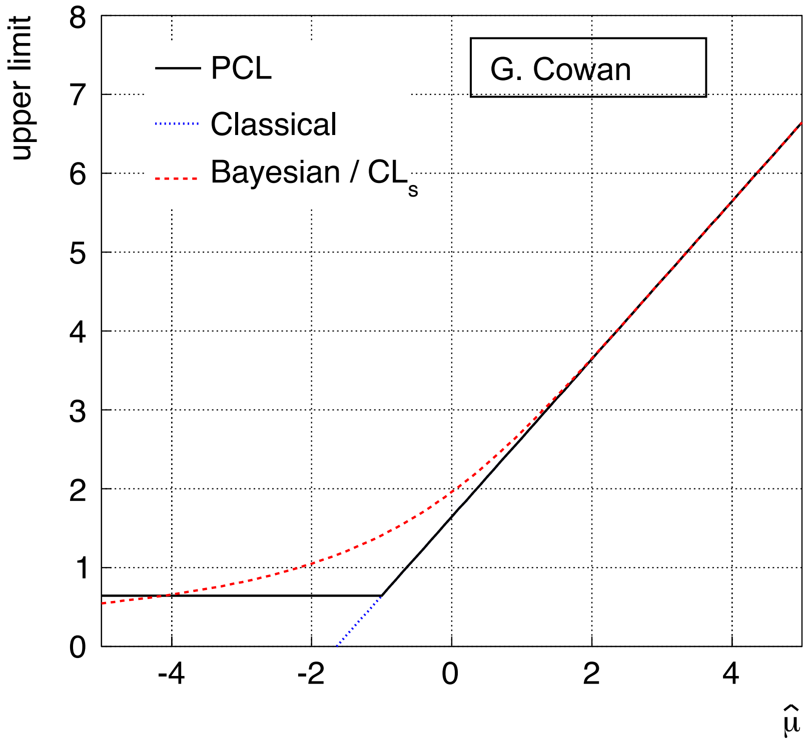
expected limit: median value of μ which will be excluded under BG-only green and yellow bands are 68% (95%) confidence intervals around this

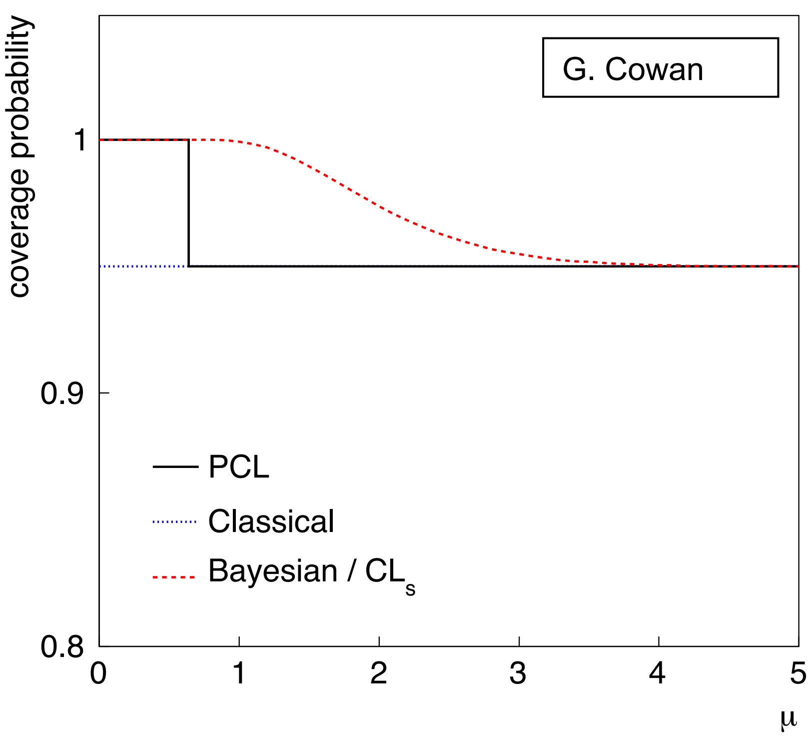
expected CL_S limit worse due to division by 1-p-value(b-only) = 0.5 on average

Comparison of Upper Limits and Their Coverage

Gauss-PDF with variance =1 physical region $\mu \geq 0$ CL=95%

(PCL with M_{min}=16%, equivalent to replace observation by -1 if < -1)





PCL: coverage known either desired one or 100%

CL_S : now preferred at LHC as used for long time and equivalent to Bayesian with flat a-priori probability

Flip-Flop-Problem for Mean of Gauss-PDF

In principle: decide before measurement whether to quote one- or two-sided interval

In praxis: if two-sided CI at XY% CL does not contain 0 then

quote two-sided CI at 68% CL, else upper limit at 95% CL

→ this is the flip-flop problem with too small coverage

One and two-sided CI at 90% CL for variance =1

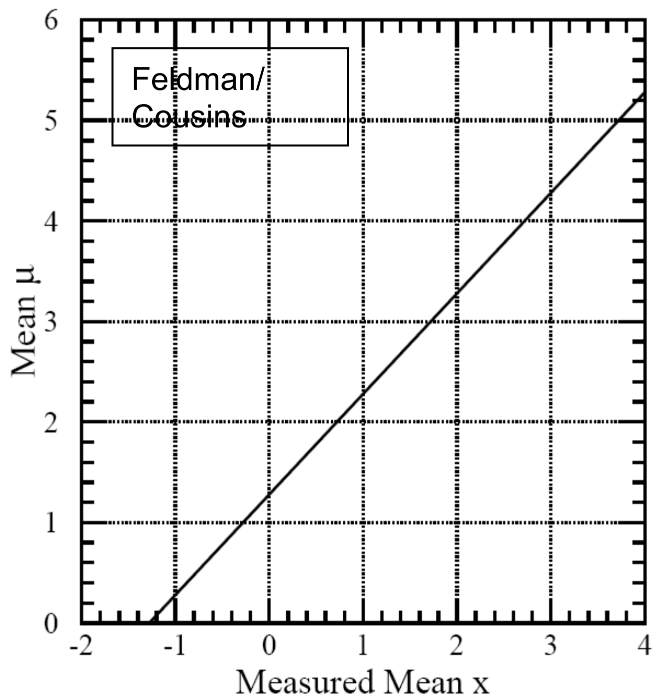
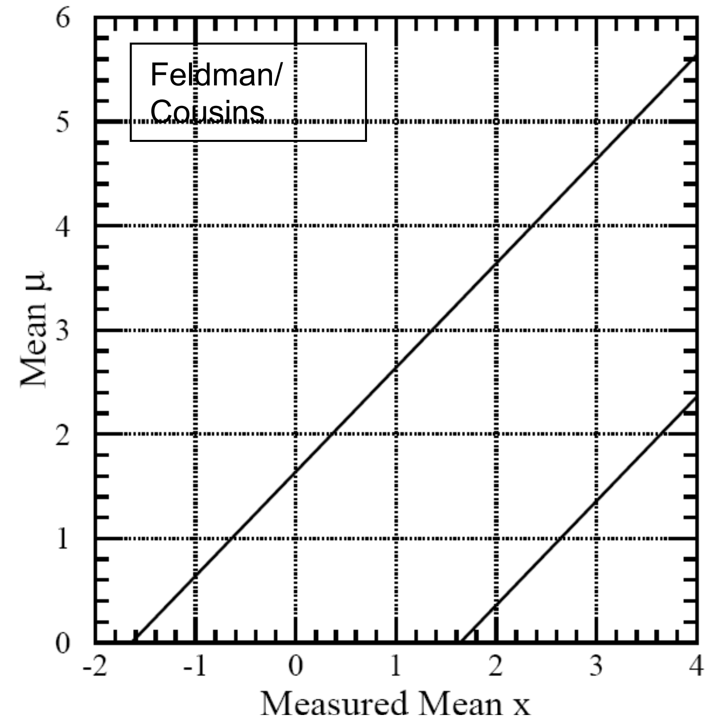


FIG units of

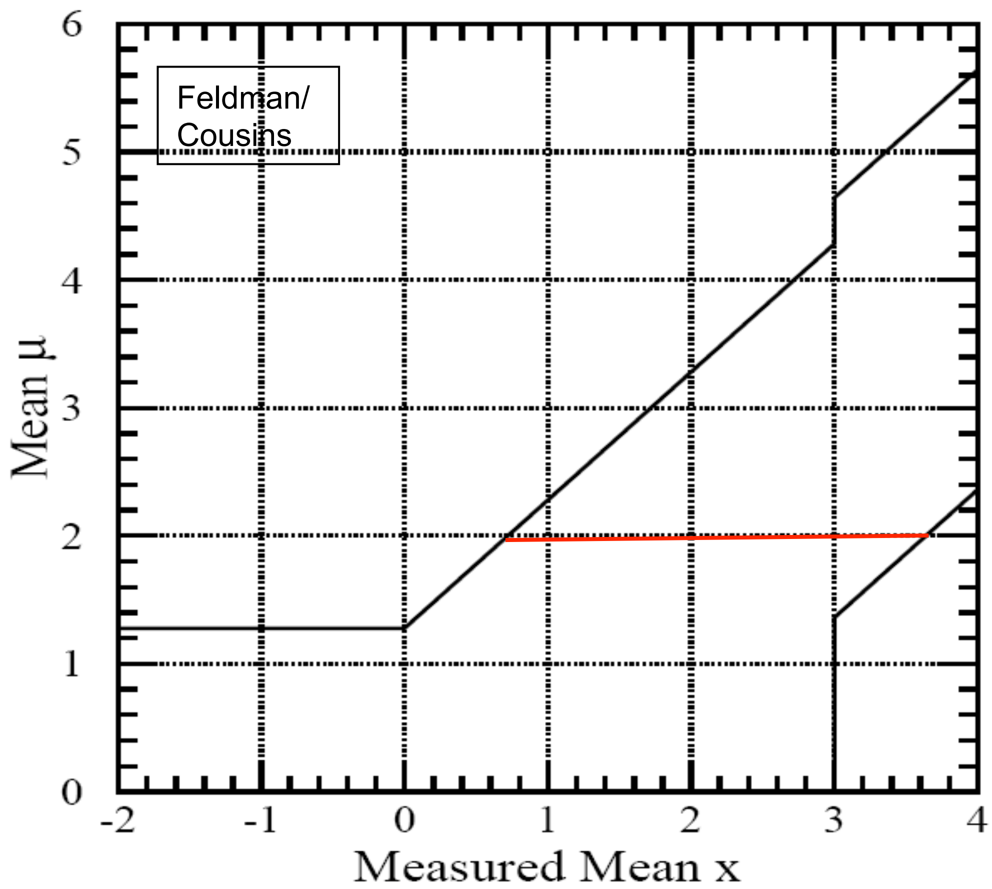
one-sided: measured x + 1.28



two-sided: measured x ± 1.64

ean of a

Flip-Flop-Problem for Mean of Gauss-PDF



Assumption: flip-flop at 3

for $x_{obs} > 3$ two-sided CI

else one-sided CI

Problem:

for $1.36 < \mu < 4.28$

coverage is only 85%

i.e. smaller than quoted value of CL=90%

Solution unified approach / unified confidence intervals

Re-discovered for HEP in 1998 by Feldman and Cousins

Construction of CI using Likelihood Ratio

Ordering principle: include possible measured x values according to decreasing likelihood ratio R(x) in confidence belt

Maximum likelihood estimator for μ given true value constrained to ≥ 0:

$\mu_{best} = x$ for $x \geq 0$

$\mu_{best} = 0$ for $x < 0$

Likelihood for x assuming μ_{best}

$$P(x|\mu_{\rm best}) = \begin{cases} 1/\sqrt{2\pi}, & x \geq 0 \\ \exp(-x^2/2)/\sqrt{2\pi}, & x < 0. \end{cases}$$

Likelihood ratio R(x) defined according to :

$$R(x) = \frac{P(x|\mu)}{P(x|\mu_{\rm best})} = \begin{cases} \exp(-(x-\mu)^2/2), & x \geq 0 \\ \exp(x\mu - \mu^2/2), & x < 0 \end{cases}$$

Determine x_1 and x_2 from

$$\int_{x_1}^{x_2} P(x|\mu)dx = \alpha.$$

With condition

$$R(x_1) = R(x_2)$$

Feldman-Cousins CI for Gauss-PDF

Gauss PDF with variance =1 , physical allowed range $\mu \geq 0$
Confidence belt at 90% CL

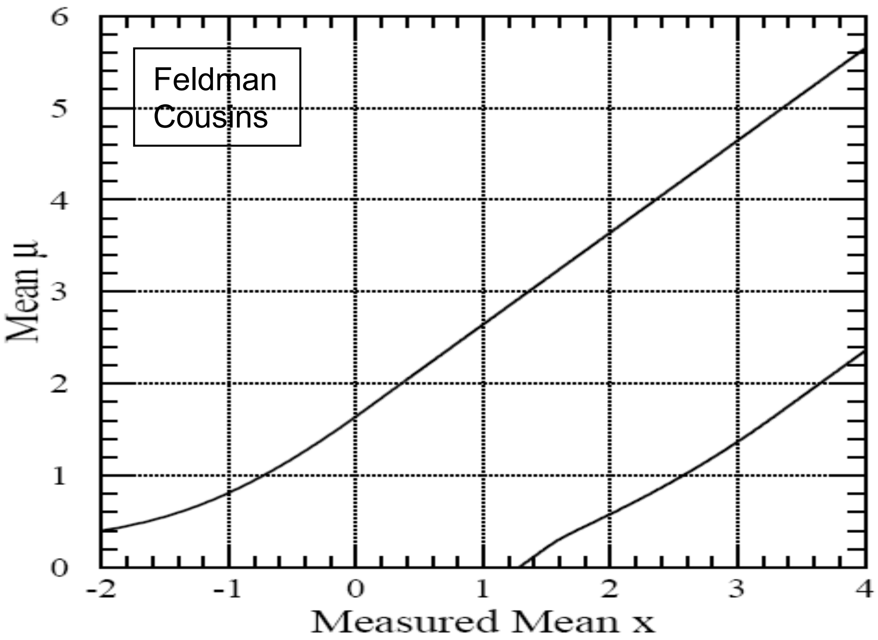


FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

$$R(x) = \frac{P(x|\mu)}{P(x|\mu_{\text{best}})} = \begin{cases} \exp(-(x-\mu)^2/2), & x \geq 0 \\ \exp(x\mu - \mu^2/2), & x < 0 \end{cases}$$

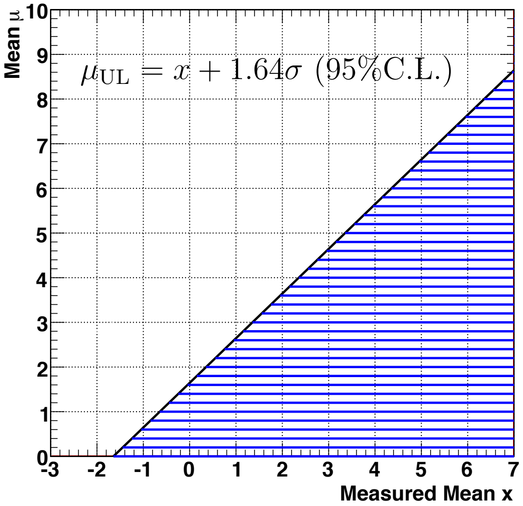
$$\int_{x_1}^{x_2} P(x|\mu)dx = \alpha.$$

$$R(x_1) = R(x_2)$$

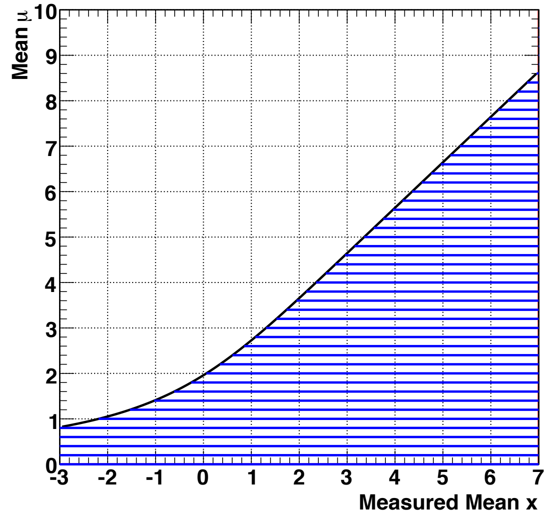
- no empty intervals, automatic transition from one-sided to two-sided CI
- for large measured values of x CI identical to classical (for Gauss-PDF)
- for small measured value of x FC-CI longer than classical CI
  (this is the price one has to pay when avoiding flip-flop-problem)

Comparison of Upper Limits 95% CL for Gauss PDF

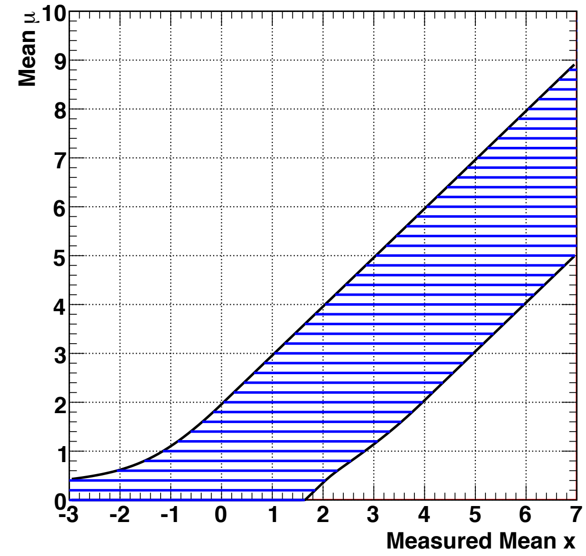
Classical



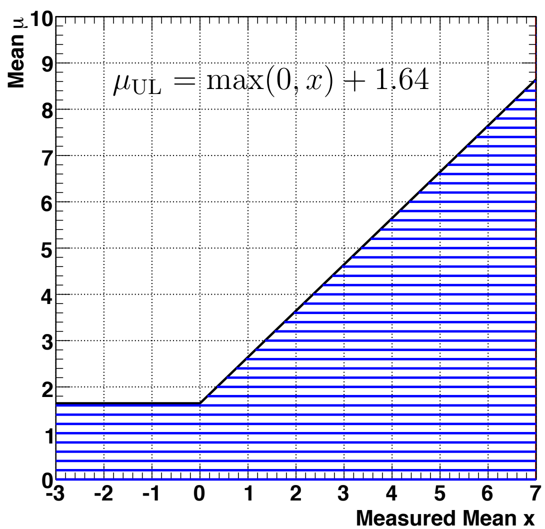
Bayesian / CL_S



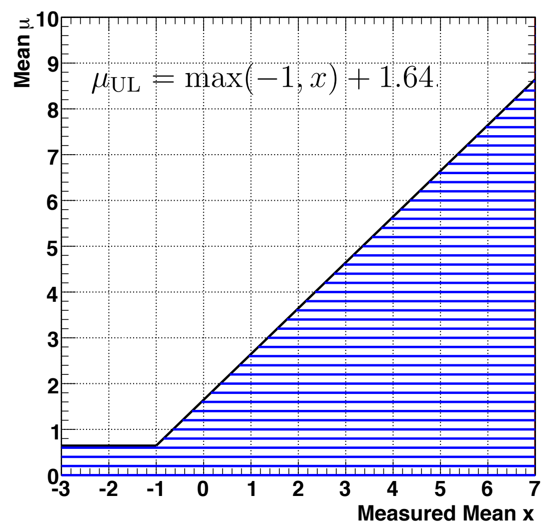
FC unified



PCL 50%



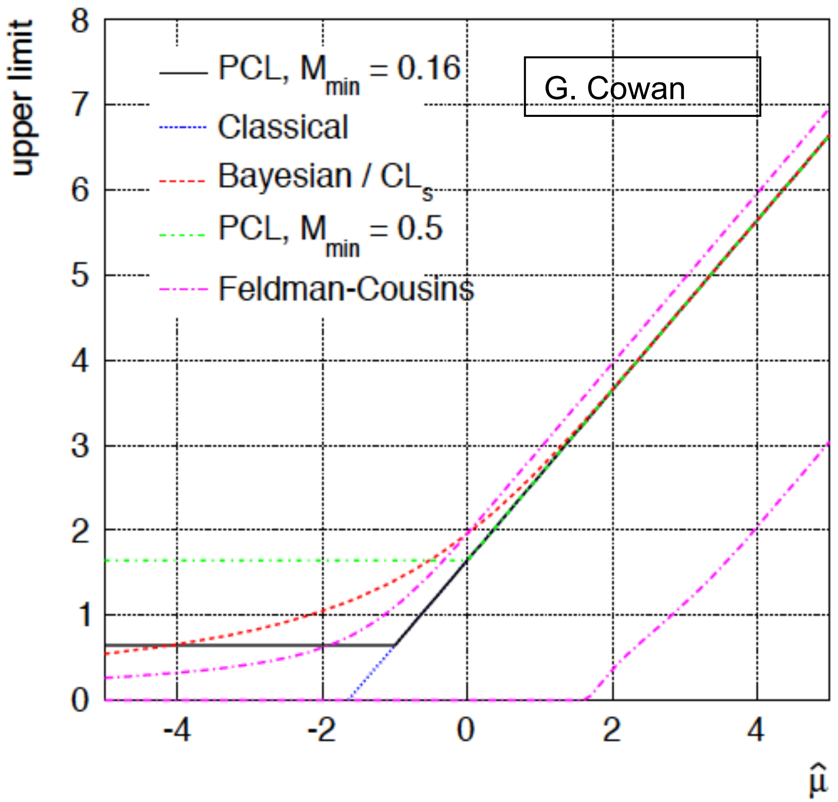
PCL 16 %

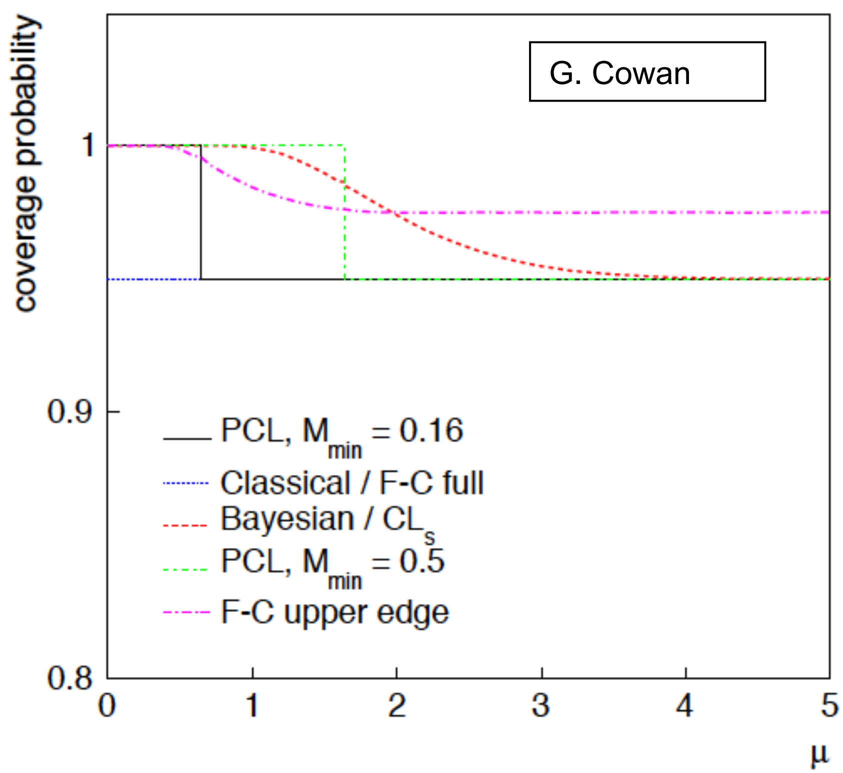


Comparison of Upper Limits and Their Coverage

Gauss PDF with variance =1 , physical region $\mu \geq 0$. upper limit at 95% CL

(PCL with M_{min}=16% (50%), equivalent to replacing observation by -1 if < -1 (0 if < 0))





FC gives smallest upper limits for large negative values

FC/unified approach can be supplemented by power constraint