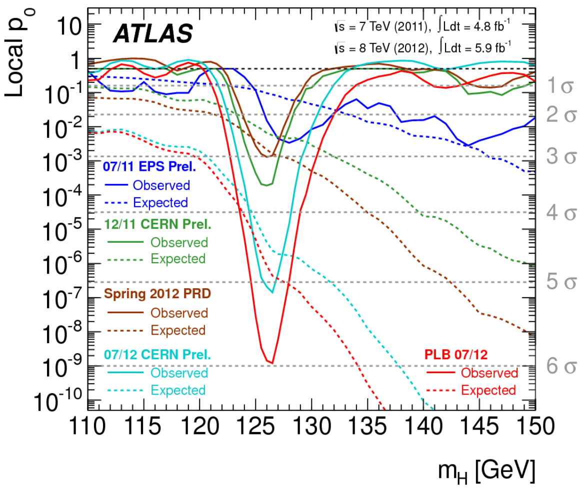
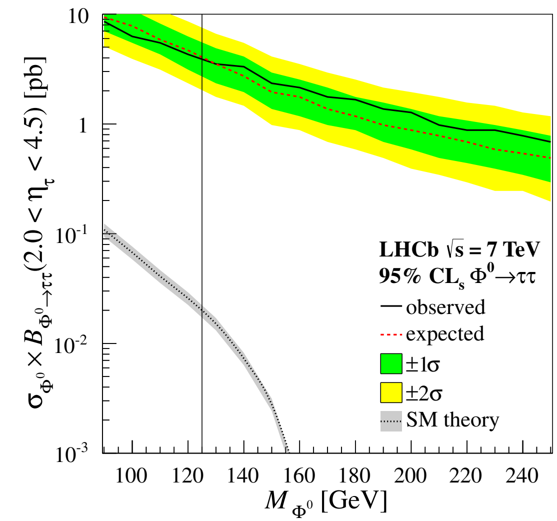
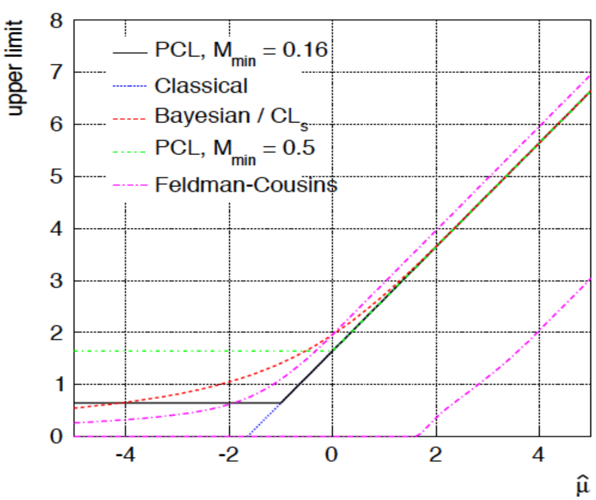
Hypothesis Testing and Confidence Intervals/Limits (Frequentist: Classical, FC, PCL ; Bayesian ; CL_S)

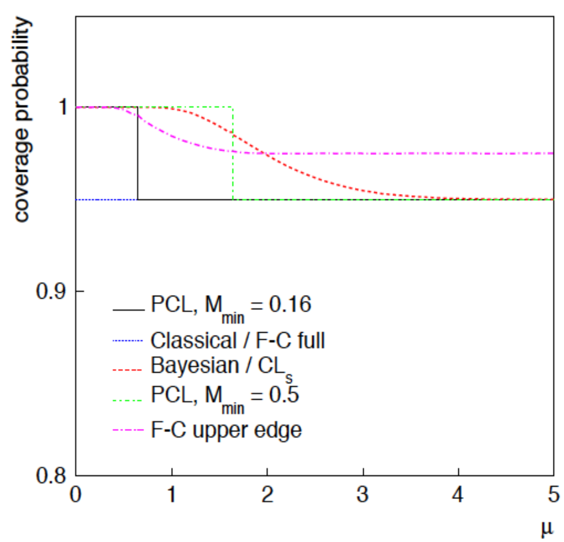


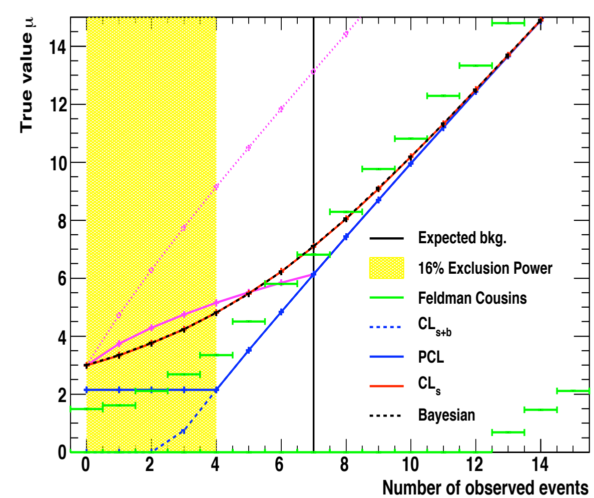
Markus Schumacher

Heidelberg „B-Workshop"
Neckarzimmern
6 March 2014

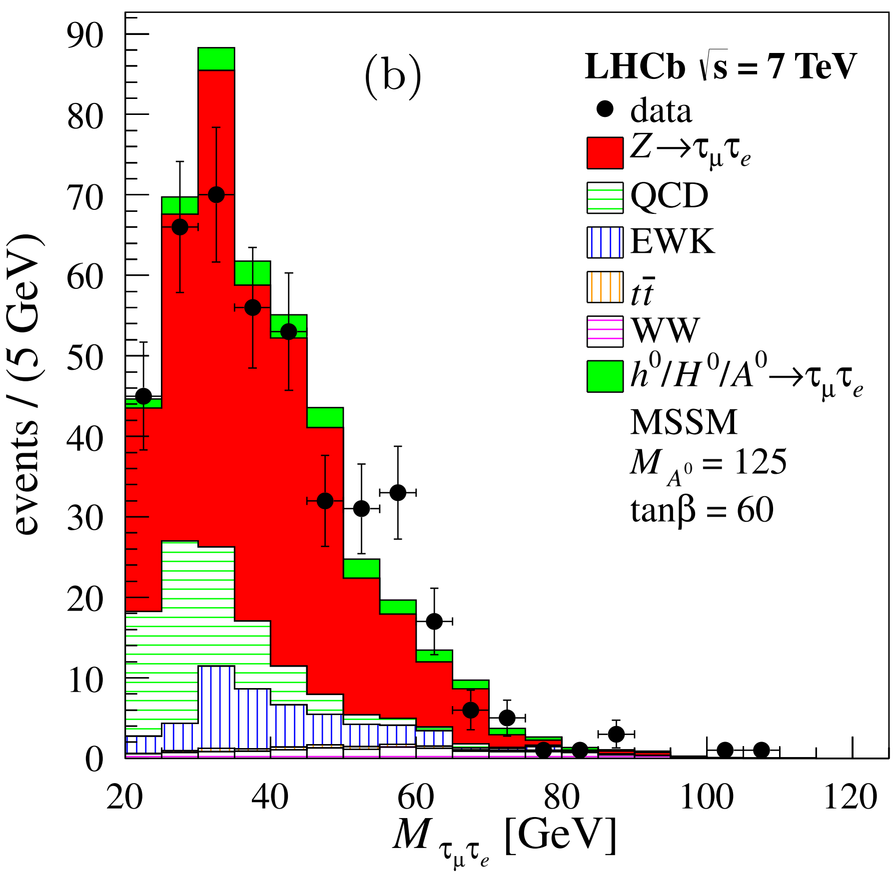








Motivation



Outcome of analysis
- expected background event spectrum
- expected signal event spectrum
  for predefined signal rate *s*
- observed event spectrum

Hypothesis tests:

a) Only Background → Discovery

b) Signal + background → Exclusion
   = Reject one signal model/strength *s*
   (if *s(M)* is a function of mass *M*
   then test different *M* values
   → range of rejected mass hypothesis
   not a confidence interval for *M*)

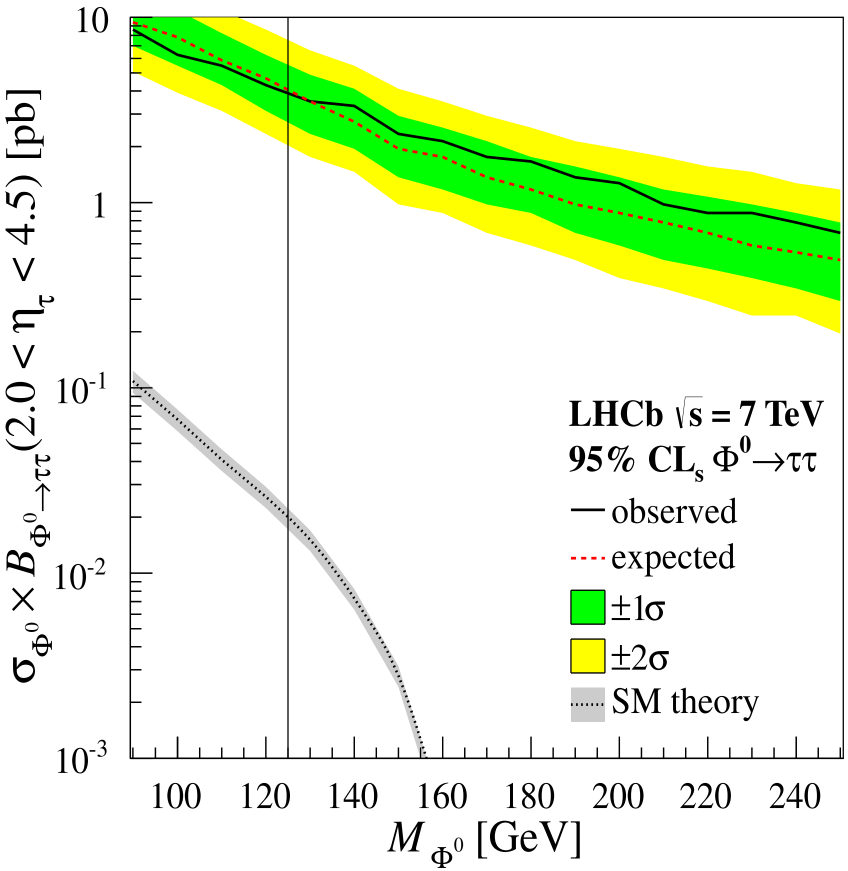
Extension to b):

Confidence interval CI [a,b] = set of signal strength which can not be excluded

One sided CI [-inf, b] → b called "upper limit" on signal strength

Smallest (largest) signal strength, which can (not) be excluded

Motivation



Outcome of analysis
- expected background event spectrum
- expected signal event spectrum
  for predefined signal rate *s*
- observed event spectrum

Hypothesis tests:

a) Only Background → Discovery

b) Signal + background → Exclusion
   = Reject one signal model/strength *s*
   (if *s(M)* is a function of mass *M*
   then test different *M* values
   → range of rejected mass hypothesis
   not a confidence interval for *M*)

Extension to b):

Confidence interval CI [a,b] = set of signal strength which can not be excluded

One sided CI [-inf, b] → b called "upper limit" on signal strength

Smallest (largest) signal strength, which can (not) be excluded

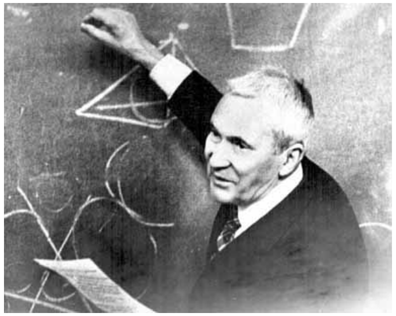
Axiomatic Definition and Conditional Probability

Consider set *S* with subsets *A*, *B*, ...
Assign to each set a number between 0 and 1 with

For all $A \subset S, P(A) \geq 0$

$$P(S) = 1$$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$



Kolmogorov
Axioms (1933)

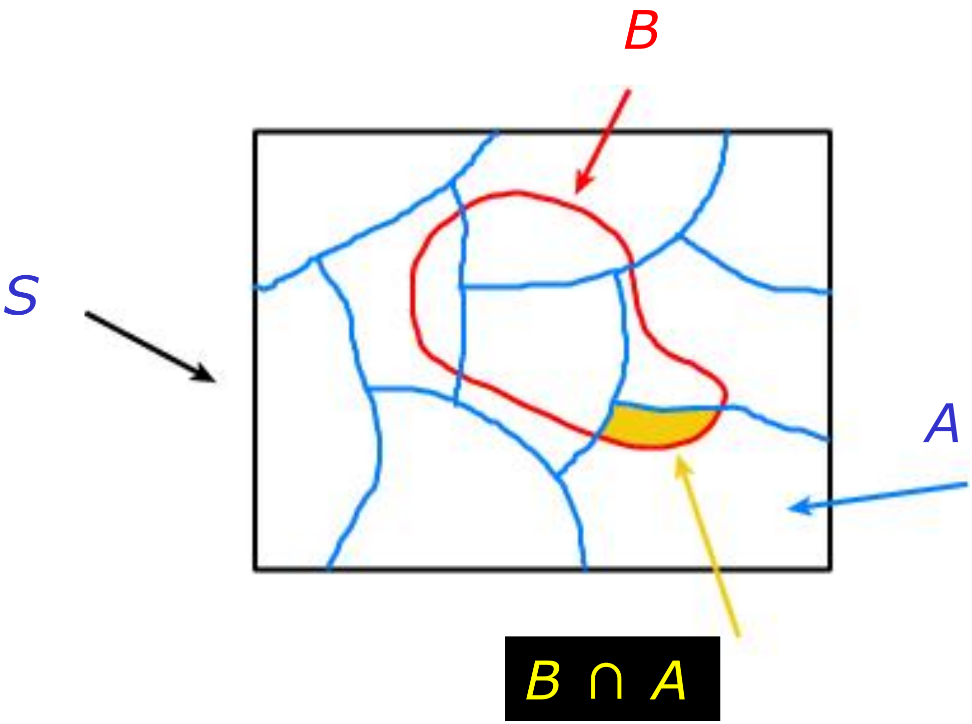
Conditional probability (for *P(B) ≠ 0*)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

If subsets A,B independent:

$$P(A \cap B) = P(A)P(B)$$

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$



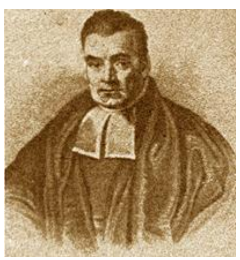
Bayes Theorem

From the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \cap B) = P(B \cap A) \qquad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

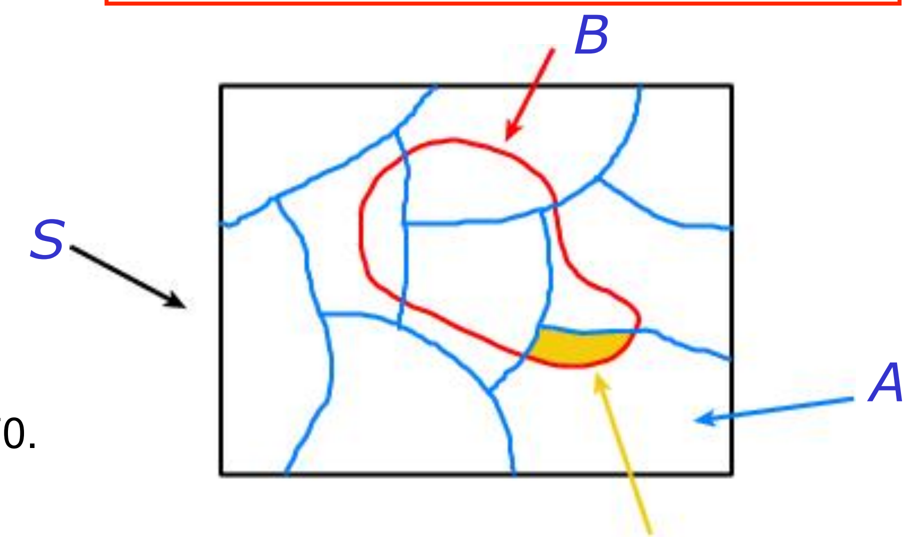
$$P(B) = \sum_i P(B|A_i)P(A_i)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370.



$B \cap A_i$

Axiomatic definition not helpful in real life.
Need: definition of subsets, rule to assign probability values

2 Schools: Frequentists and Bayesians

Bayes Theorems holds and is accepted in both schools

Controversy about: what are the subsets, to which probability values can be assigned

Frequentist and Bayesian

Subsets:

Outcome of (repeatable) experiment — Any hypothesis

Assignment of probabilities:

Relative frequency in limit nr of trials → inf. — Degree of belief in hypothesis

$$P(A) = \lim_{n\to\infty} \frac{\text{times outcome is in } A}{n}$$

$$P(A) = \text{degree of belief that } A \text{ is true}$$

P (SUSY exists)

P (9.81 m/s^2 < g < 9.82m/s^2)

P (rain in Neckarzimmern on 7.3.2014)

Not defined. Either 0 or 1. — No problem. This is the goal.

Bayesian definition: More general (includes Frequentist definition)

Applicable to singular events, "true" values, …

Does not care about repeatability of experiment

Needs a-priori probability in application of Bayes theorem

Bayesian Statistics: General Philosophy

How to use Bayes theorem to update "degree of belief" in light of data

Probability to observe data assuming a hypothesis H (true value of a parameter)
Likelihood function (also used by Frequentists)

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

A-priori probability, i.e. before data taking (not defined in Frequentist school)

Posterior probabilty, i.e. after analysis of the data (not defined in Frequentist school)

Normalisation includes sum/integral over all possible hypothesis/par. values

No general rule for choice of a-priori probability → "subjective"

"Objective" prior = uniform? - not well defined probability for infinite parameter space
- uniform in θ, θ^2 sqrt(θ), ln θ, … ?
→ Jeffrey Prior p(θ) = sqrt(Information (θ))
uniform for mean μ of Gauss pdf
1/sqrt(μ) for Poisson 1/τ for exp(-t/ τ)

Example: Parameter estimation

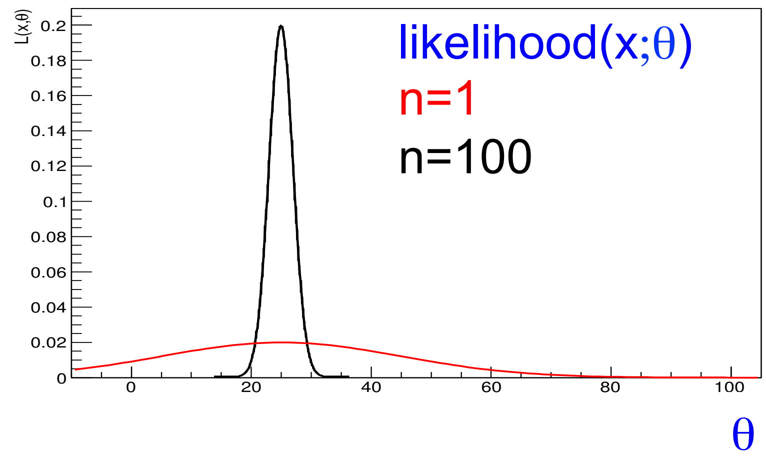
Frequentist: maximise likelihood

Bayesian: maximise posterior probability

Estimation of mean value θ of Gaussian PDF
Resolution σ = 20. Sample mean yields: x =25

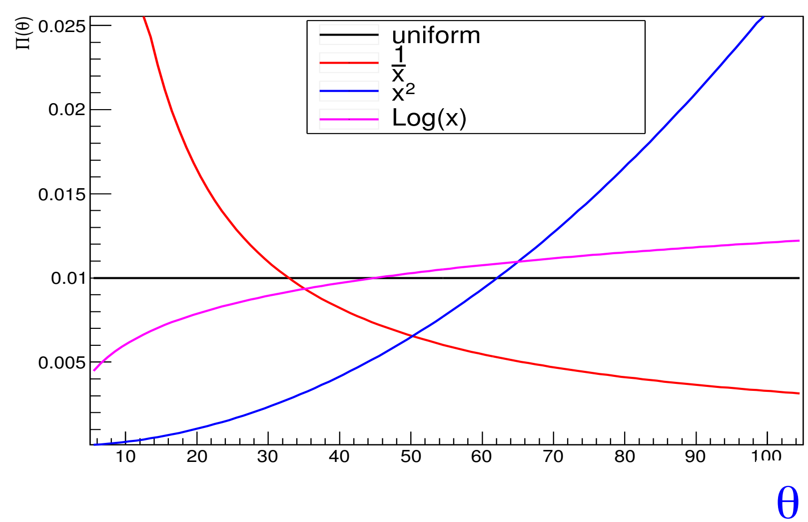
Consider two sample sizes: n= 1 (100)

→ Likelihood functions are
Gaussians with σ/sqrt(n) = 20 (2)



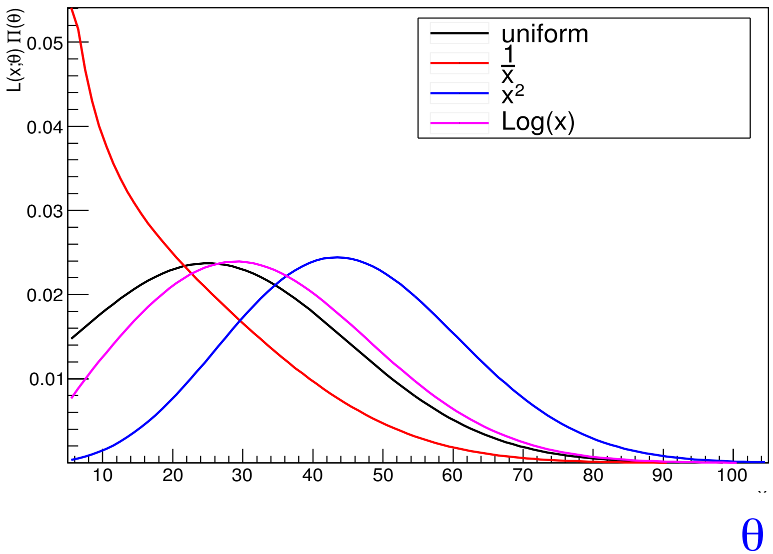
Four different a-priori probabilities
normalised in range 5 to 105

uniform, 1/x, x^2, ln(x)



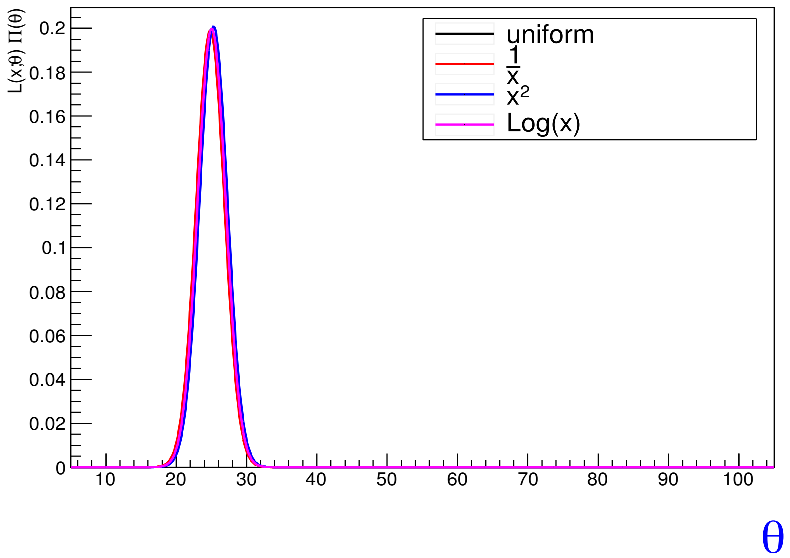
Ex.: Parameter Estimation - Posterior Probabilities

Sample size n = 1
Large spread in posterior prob.



Significant dependence of mode on a-priori probability

Large samples size n = 100
Small spread in posterior prob.



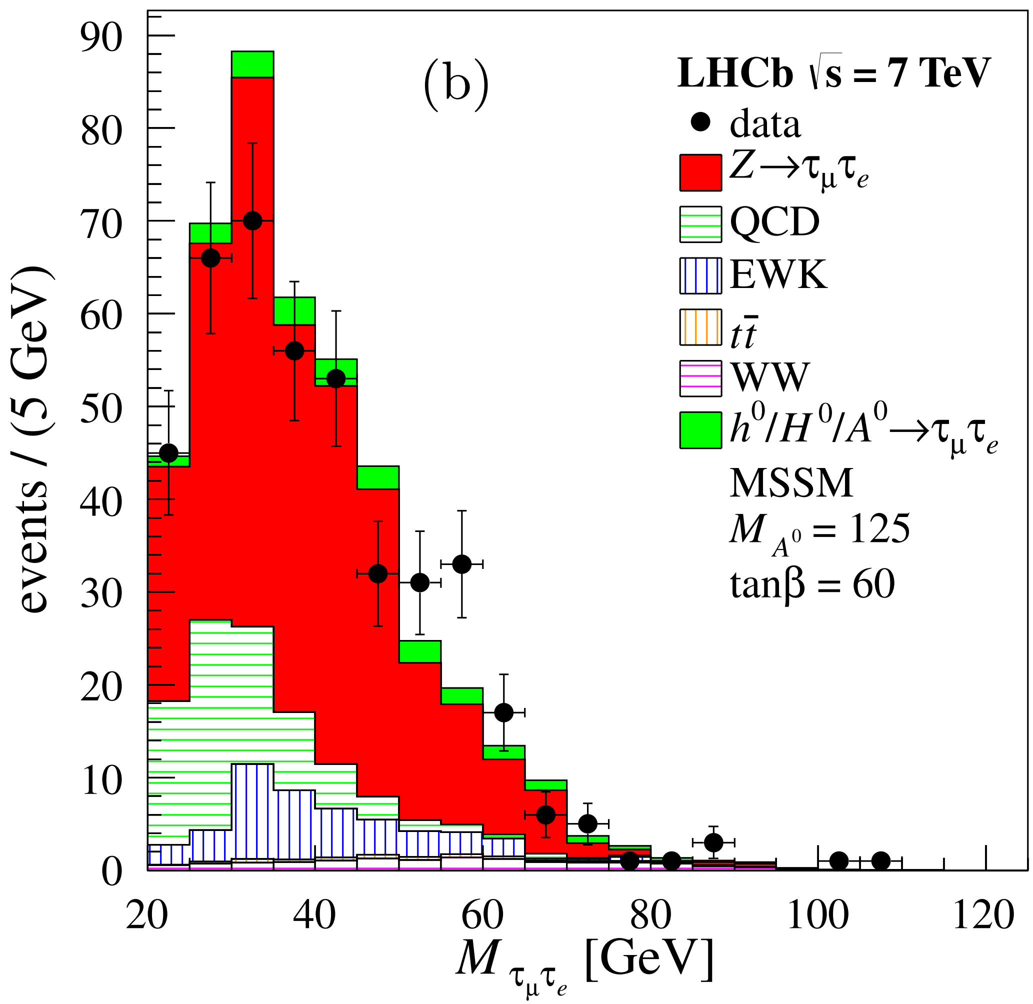
Small dependence of mode on prior probability

For sample size n → infinity Bayesian and Frequentist results identical
Bayesian with uniform a-priority prob. and Frequentist numerical identical

Exception: in special situations e.g. close to a physical boundary

But interpretation is always different in both schools

Hypothesis Testing



Looking for a signal: test "Background only" hypothesis
Excluding a signal: test "Signal + Background" hypothesis

Either decide before measurement or do always both

Hypothesis Testing (Frequentist Technique)

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

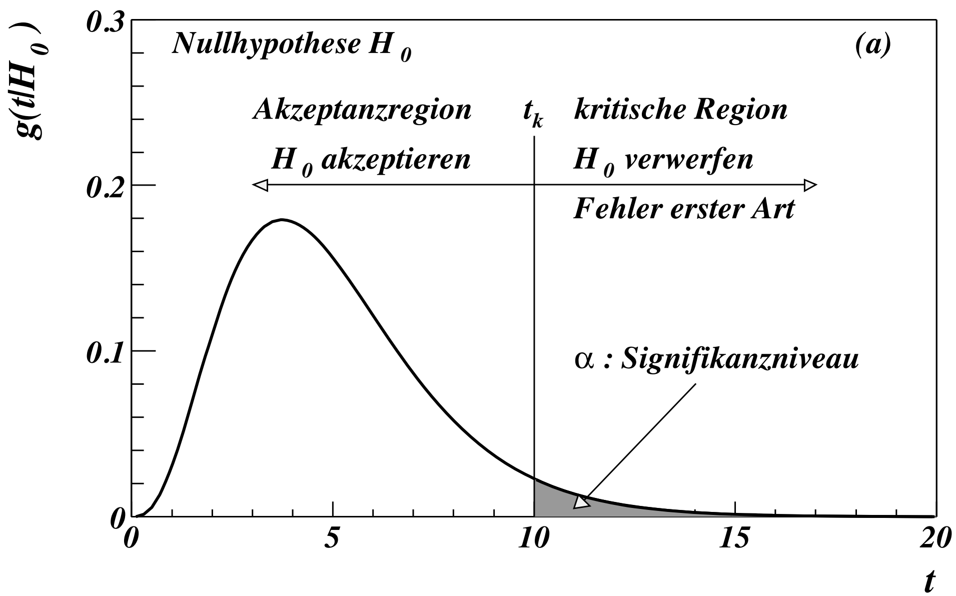
Test statistic t: any function of your data which is used to quantify (dis-)agreement with H_0

g(t|H0): 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

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



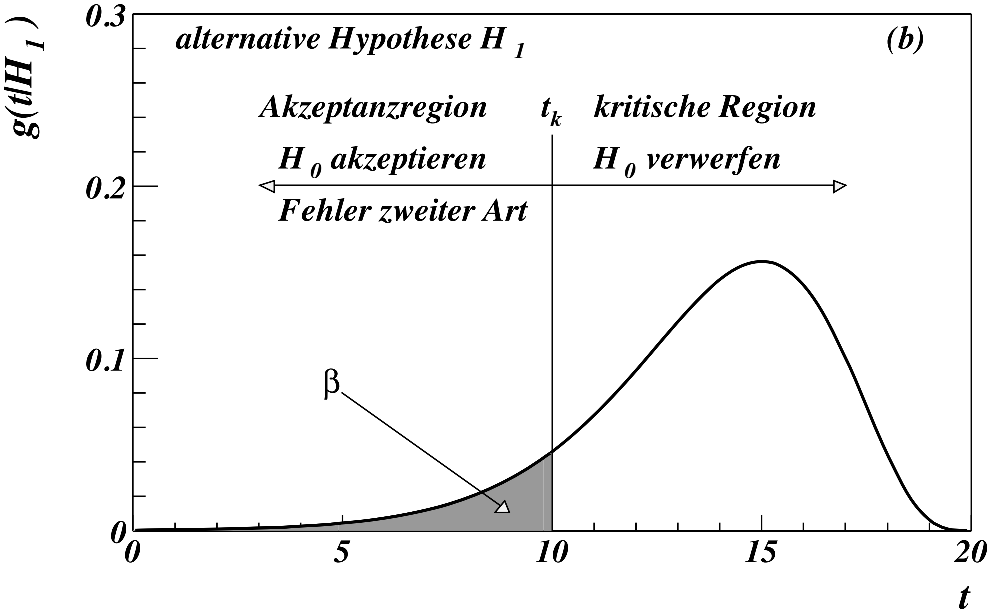
Hypothesis Testing

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 2nd 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

An 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	$U_1 : x < \lambda^{\mathrm{I}} \text{ und } x > \lambda^{\mathrm{II}}$	$\text{mit } \int_{-\infty}^{\lambda^{\mathrm{I}}} f(x)\,\mathrm{d}x = \int_{\lambda^{\mathrm{II}}}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}\alpha\,;$
one-sided in upper tail	$U_2 : x > \lambda^{\mathrm{III}}$	$\text{mit } \int_{\lambda^{\mathrm{III}}}^{\infty} f(x)\,\mathrm{d}x = \alpha\,;$
one-sided in lower tail	$U_3 : x < \lambda^{\mathrm{IV}}$	$\text{mit } \int_{-\infty}^{\lambda^{\mathrm{IV}}} f(x)\,\mathrm{d}x = \alpha\,;$
two-sided in center	$U_4 : \lambda^{\mathrm{V}} \leq x < \lambda^{\mathrm{VI}}$	$\text{mit } \int_{\lambda^{\mathrm{V}}}^{\lambda_0} f(x)\,\mathrm{d}x = \int_{\lambda_0}^{\lambda^{\mathrm{VI}}} f(x)\,\mathrm{d}x = \frac{1}{2}\alpha\,.$

An Example: Test for Mean Value of Gaussian PDF

Rows: 4 critical regions

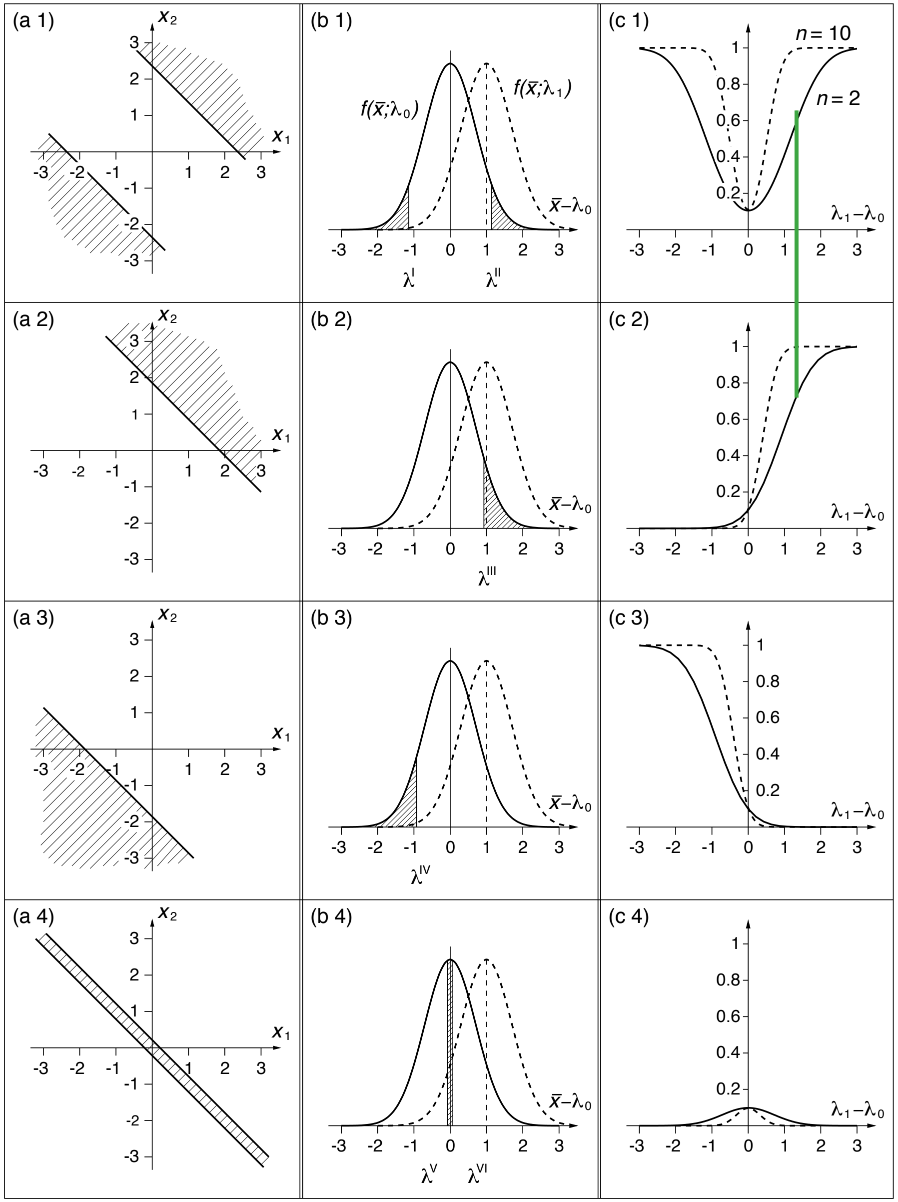
two sided in tails
one-sided in upper tail
one-sided in lower tail
two-sided in center

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

An Example: Test for Mean Value of Gauss PDF

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

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

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

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

U_1 is unbiased test

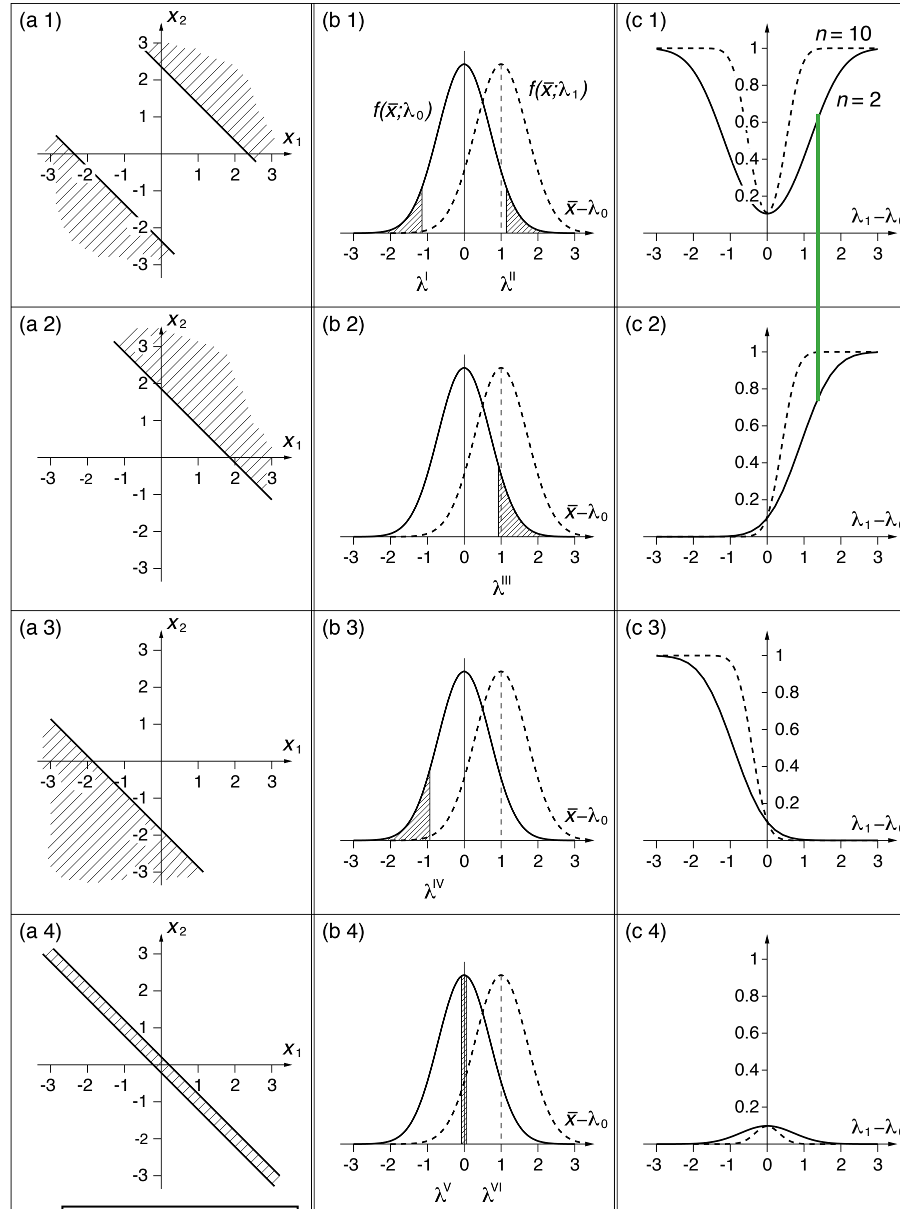
power ≥ significance for all λ

U_2: larger power for $\lambda_1 > \lambda_0$

U_3: larger power for $\lambda_1 < \lambda_0$

U_4: no useful test

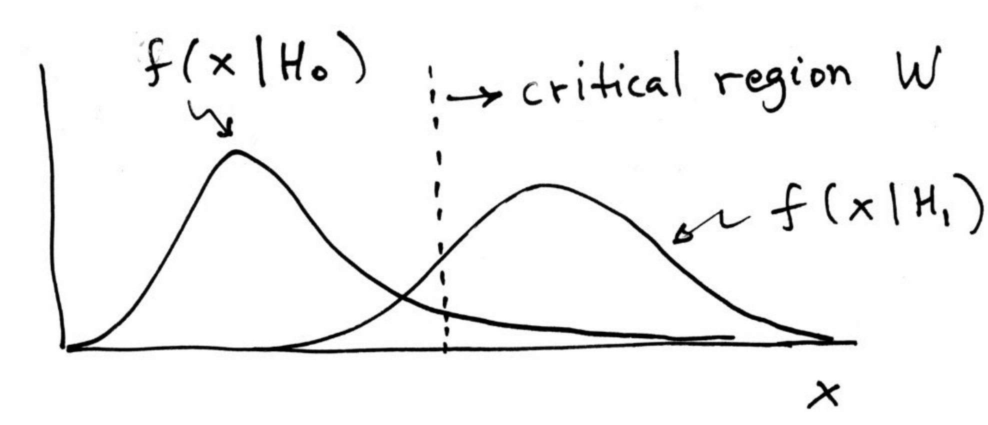
maximal power for $\lambda_1 = \lambda_0$



S. Brandt

Best Test and Neyman-Pearson-Lemma NPL

Best test: for given significance level α, maximize power M=1-β



Questions: Which test statistic t?
Which choice of critical region?

Simple hypothesis: completely fixed, no free parameters to be determined from data

Neyman-Person-Lemma: a test of a simple null hypothesis H_0 w.r.t. to the simple alternative hypothesis H_1 is a best test, if the critical region is chosen such that inside it holds:

$$\frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)} > c$$

P = probability to observe sample x
(≤ c outside critical region)
c is a constant depending on α

Equivalent statement: the optimal test statistics is given by the likelihood ratio
(or any monotonic function 1/t(, t/(1+t), ln t)

$$t(\mathbf{x}) = \frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)}$$

Challenge in praxis: determination of PDFs for t under different hypothesis

Example for Neyman-Pearson-Test

Test for the mean value λ in Gauss PDF with known variance σ^2
Sample of size n: $x^1, x^2, \ldots x^n$

Likelihood for data set under hypotheses H_0: $\lambda = \lambda_0$ und H_1: $\lambda = \lambda_1$

$$f(X|H_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(\mathsf{x}^{(j)} - \lambda_0)^2\right] \qquad f(X|H_1) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(\mathsf{x}^{(j)} - \lambda_1)^2\right]$$

The NPL test statistics is given by

$$Q = \frac{f(X|H_0)}{f(X|H_1)} = \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{N}(\mathsf{x}^{(j)} - \lambda_0)^2 - \sum_{j=1}^{N}(\mathsf{x}^{(j)} - \lambda_1)^2\right\}\right]$$
$$= \exp\left[-\frac{1}{2\sigma^2}\left\{N(\lambda_0^2 - \lambda_1^2) - 2(\lambda_0 - \lambda_1)\sum_{j=1}^{N}\mathsf{x}^{(j)}\right\}\right]$$

With non-negative constant k

$$\exp\left[-\frac{N}{2\sigma^2}(\lambda_0^2 - \lambda_1^2)\right] = k \geq 0$$

the NPL condition for Q is given by :

$$k \exp\left[\frac{\lambda_0 - \lambda_1}{\sigma^2}\sum_{j=1}^{N}\mathsf{x}^{(j)}\right] \begin{cases} \leq c, & X \in S_c \\ \geq c, & X \notin S_c \end{cases}$$

Example for Neyman-Pearson-Test

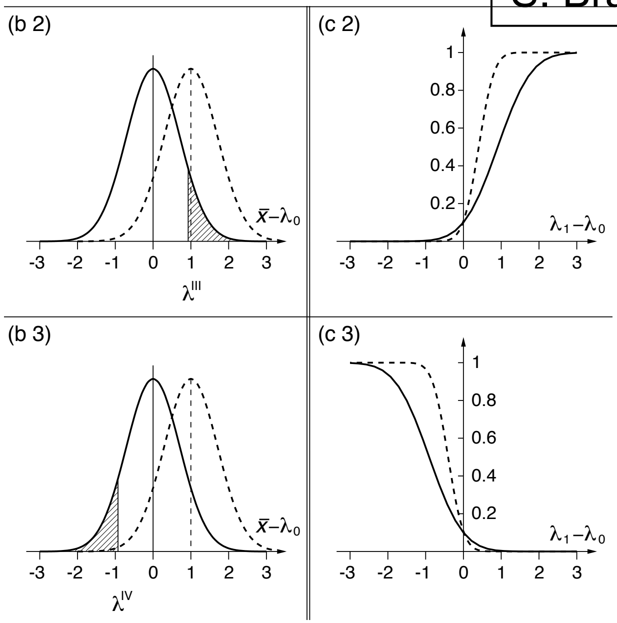
The critical region S_c according to the NPL is given by:

$$k \exp\left[\frac{\lambda_0-\lambda_1}{\sigma^2}\sum_{j=1}^{N}\mathbf{x}^{(j)}\right]\begin{cases}\le c, & X \in S_c\\ \ge c, & X \notin S_c\end{cases} \quad \text{or} \quad (\lambda_0-\lambda_1)\mathbf{x}\begin{cases}\le c', & X \in S_c\\ \ge c', & X \notin S_c\end{cases}$$

$$\mathbf{x} = \frac{1}{n}(\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_n)$$

c are c' constants depending on choice of significance level α

S. Brandt



NPL: arithmetic mean *X* is optimal test statistic

critical region in one tail of distribution

→ one-sided test

a) $\lambda_1<\lambda_0$: left sided b) $\lambda_1>\lambda_0$: right sided

$$\mathbf{x}\begin{cases}\le c'', & X \in S_c\\ \ge c'', & X \notin S_c\end{cases} \qquad \mathbf{x} \ge c''' \quad X \in S_c$$

There is no uniform most powerful test, due to change in sign of $\lambda_1-\lambda_0$

Profile Likelihood Ratio

NPL strictly valid only for simple hypothesis

Now: parameter (μ) is only fixed under H_0 but not under H_1 (composite hyp.)

Proposed test statistic t
profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

numerator: μ fixed to value under H_0

denominator: find ML estimate for μ

$0 \leq \lambda(\mu) \leq 1$

In praxis as optimal as NPL. Allows easy treatment of nuisance parameters Θ

Wilks theorem (for this special case):

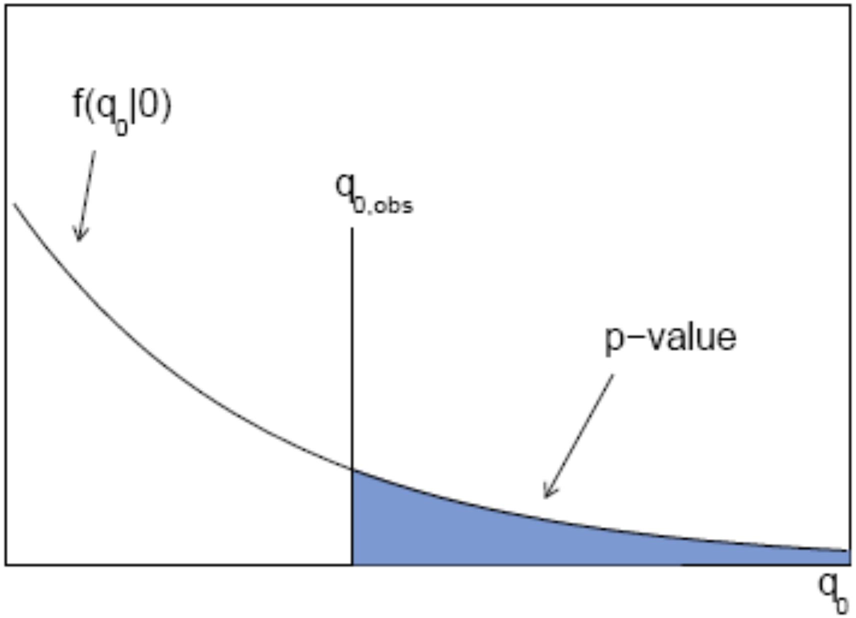
PDF for $-2\ln\lambda$ is given by a Chi^2-PDF with number of degrees of freedom DOF equal to the number of parameters fixed under H_0 in the limit of large sample size. Here: consider only the case of 1 parameter fixed under H_0

Applied to Gaussian test of $\lambda = \lambda_0$ versus $\lambda \neq \lambda_0$ yields:

- test statistic is monotic function of arithmetic mean, two sided test recommended
- PDF for $-2\ln(\lambda)$ is exactly Chi^2–PDF with 1 DOF also for limited sample size

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)

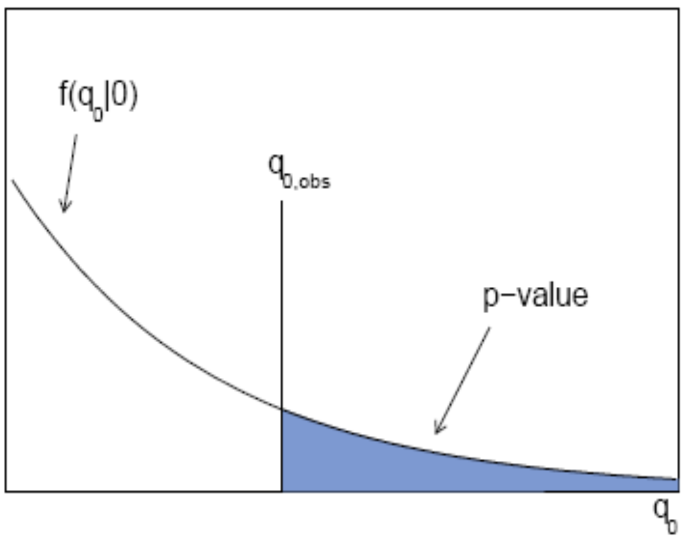
if P-value = significance level α, then $q_{obs.} = q_{critical}$

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

P-Value and Significance



If P-Value < predefined value α
then reject null hypothesis

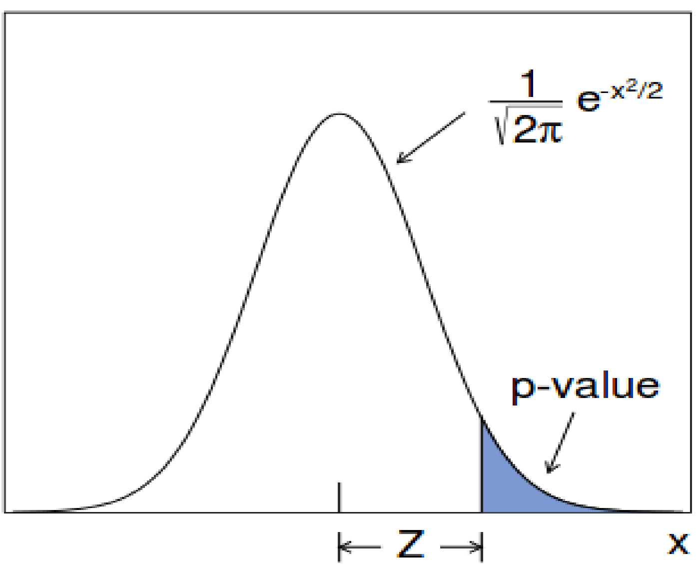
Convention:

for discovery require p-value $< 2.87 \times 10^{-7}$

for exclusion require p-value < 0.05

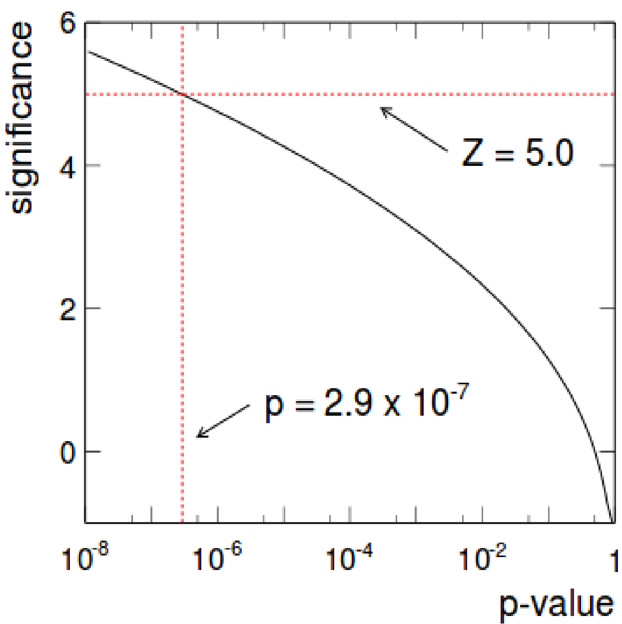
p-value translated to significance Z via Standard Gauss PDF

Significance of 5 (1.64) corresponds to $P = 2.87 \times 10^{-7}$ (0.05)



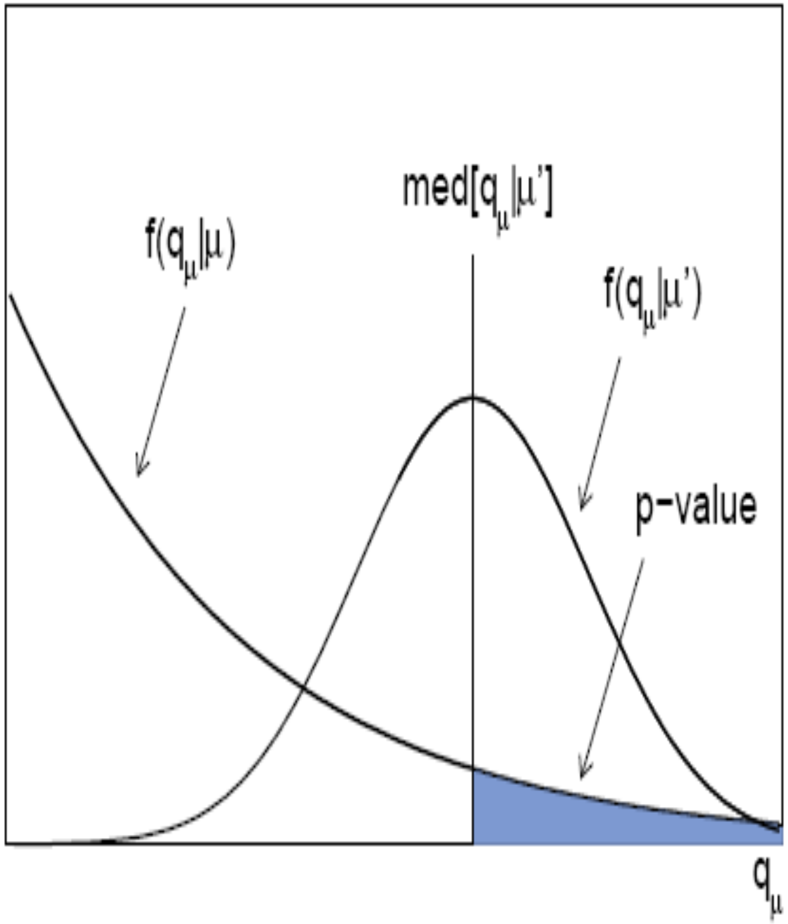
$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1 - p_0)$$



Expected Sensitivity

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



a) „Discovery“: reject H_0= backgr.-only hyp.
- determine median of q_μ under alternative sig+background hypothesis
- determine p-value for median q_μ under null hypothesis = background-only

→ expected significance

b) „Exclusion“: reject H_0= signal+backgr hyp.
- determine median of q_μ under alternative background-only hypothesis
- determine p-value for median q_μ under null hypothesis = signal+background

→ expected exclusion

Different Choices For Hypothesis Test

LLR test statistics

Higgs Days at Santander 2011 (A. Read)

	Test statistic	Test statistic	Nuisance parameters	Pseudo-experiments
LEP	$-2\ln\frac{L(\mu,\tilde{\theta})}{L(0,\tilde{\theta})}$	Simple LR	Fixed by MC	Nuisance parameters randomized about MC
Tevatron	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\hat{\theta}})}$	Ratio of profiled likelihoods	Extracted from priors	Nuisance parameters randomized from priors
LHC	$-2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$	Profile likelihood ratio	Profiled (fit to data)	New nuisance parameters fitted for each pseudo-exp.

LHC sampling of test statistic is frequentist, LEP and Tevatron Bayes-frequentist hybrid. CL_s can be used together with any of these – must be specified! No longer sufficient to write e.g. “the CL_s method was used”.

Decisions to take: which test statistics?
how to deal with systematic uncertainties?
how to determine PDF for test statistic ?
how to handle results close to physical boundary?

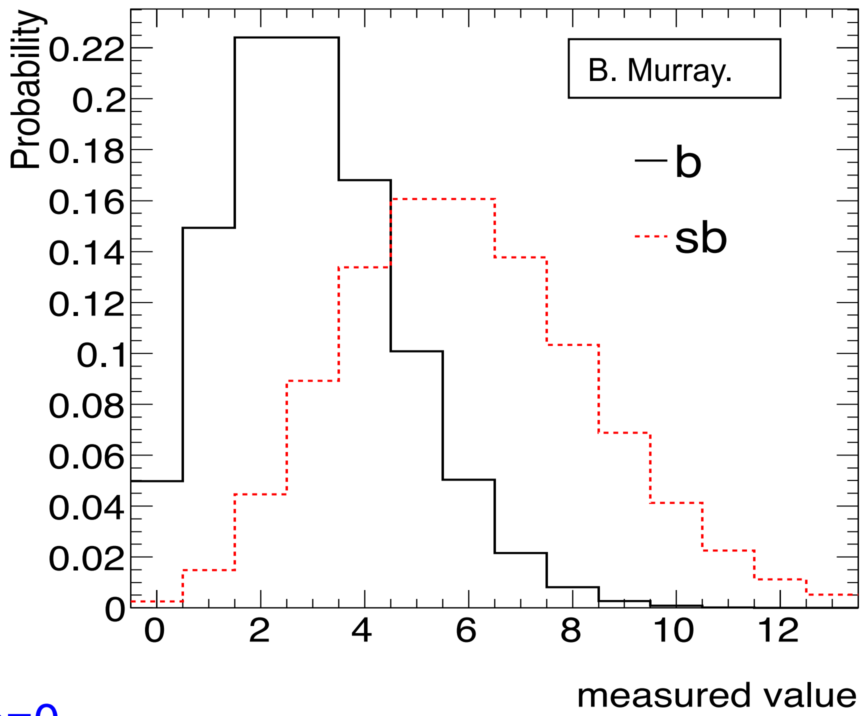
Poisson-PDF: Simple Test Statistic $t = n_{observed}$

Expected background rate b
Expected signal rate s

Test statistics*:* observed events n

with known PDF for "b" and "s+b"

$$P(n; s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$



Test "background only" hypothesis → s=0

One sided test as n<b not considered as hint for existence of signal

$$p_0 = P(n \geq n_{\text{obs}} | s = 0, b) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b}$$

P-value and test statistic independent on signal strength

NPL Test for Counting Experiment

The Likelihood to observe n given H_0 (s=0,b) is:

$$L_b = \frac{b^n}{n!} e^{-b}$$

The Likelihood to observe n given H_1 (s,b) is:

$$L_{s+b} = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

→ Neyman-Pearson-Lemma: best test given by

$$\frac{L_{s+b}}{L_b}$$

or monotonic function

$$\ln \frac{L_{s+b}}{L_b} = n \ln \left(1 + \frac{s}{b}\right) - s$$

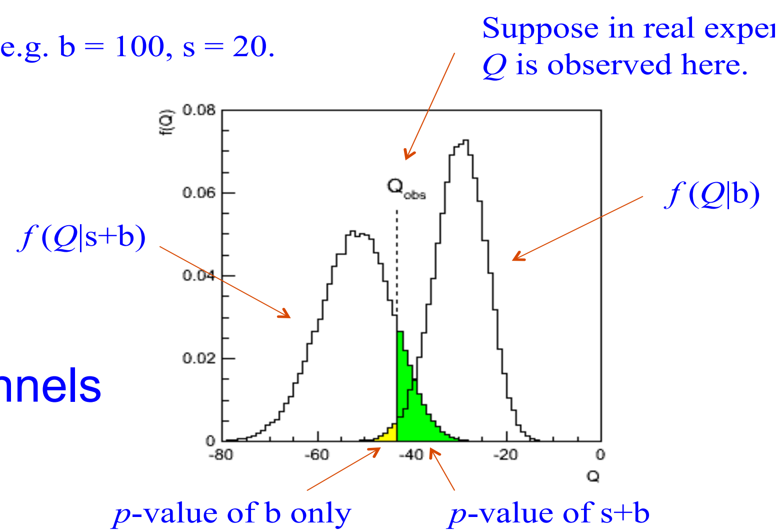
Likelihood ratio is monotonic function of n.
PDF for optimal test statistic is also Poisson distribution

→ Counting rate n is optimal test statistic

Often used:

$$Q = -2 \ln \frac{L_{s+b}}{L_b}$$

Take e.g. b = 100, s = 20.



Optimal use of distributions/ combination of channels
→ product of likelihoods per bin/channel
or sum of ln lik. per channel/bin

Profile Likelihood Ratio for Composite H_1

So far: signal rate fixed (known) under alternative hypothesis
Now: find best number of signal events under H_1 via maximum likelihood fit
i.e. H_1 is composite hypothesis with signal count as free parameter

Likelihood function
$$L(n; s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

Test statistic:
$$\lambda(s) = \frac{L(s)}{L(\hat{s})}$$

λ in [0, 1]:
1 good agreement with H_0

Enumerator: likelihood for H_0 (s fixed, for discovery s=0)
Denominator: likelihood for H_1 (s estimated from data)

Maximum likelihood estimate for signal counts:
$$\hat{s} = n - b$$

Test statistics for discovery (s=0 in enumerator):

$$\ln \lambda(0) = n \ln(b) - b - n \ln n + n$$

ln λ in [0, -infinity]:
0 good agreement with H_0

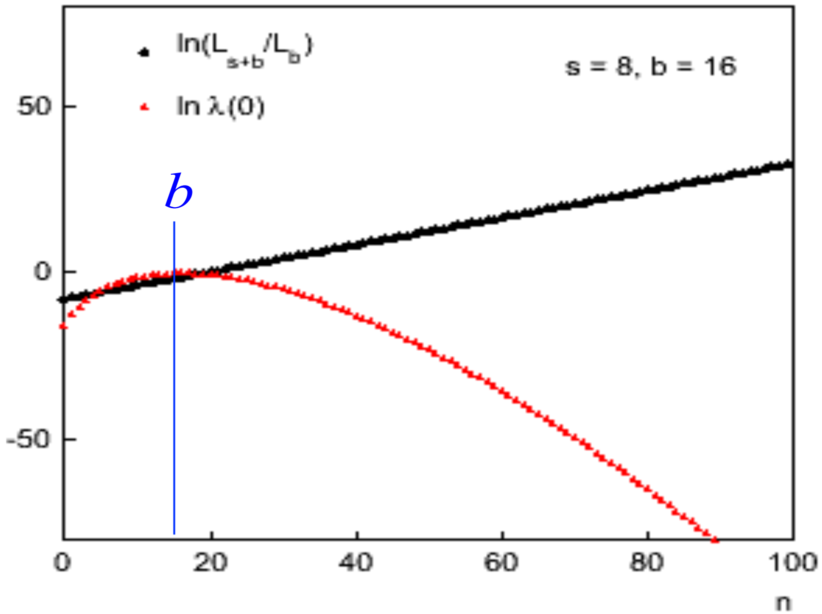
Comparison of the Test Statistic for Discovery

From Neyman-Pearson-Lemma
(simple hypothesis):

$$\ln \frac{L_{s+b}}{L_b} = n \ln\left(1 + \frac{s}{b}\right) - s$$

From profile likelihood ratio
(composite alternative hypothesis H_1)

$$\ln \lambda(0) = n \ln(b) - b - n \ln n + n$$



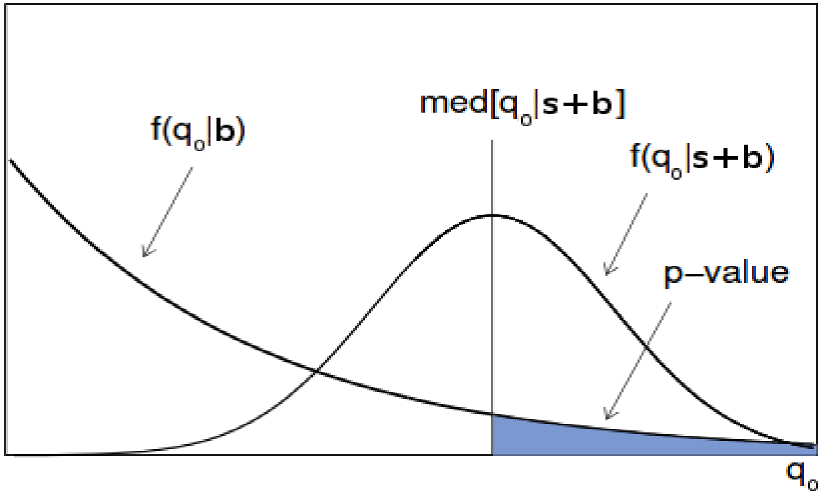
If we consider a deviation from background only hypothesis only for n>b (e.g. set ln λ(0) = 0 for n<b)

then both are monotonic and as optimal as using n
(for counting experiment neglecting systematic uncertainties)

ln λ(0)   preferred for multiple channels / distributions
add values of ln λ (0) for each/bin channel
PDF   for -2 ln λ(0) for „b only" given by Wilks theorem

Evaluation of p-Values and Signifcances Z

$$q_0 = \begin{cases} 2\left(n \ln \frac{n}{b} + b - n\right) & \hat{s} \geq 0 \\ 0 & \hat{s} < 0 \end{cases}$$



Calculation of p-values require
- PDF under null hypothesis for observed Z
- in addition PDF under alternative hypothesis for sensitivity

PDF know in large n limit due to theorems of Wilks and Wald (advantage w.r.t to t_{NPL})

Under null hypothesis s=0 theorem of Wilks gives:

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}$$

sqrt(q_0) follows standard Gauss for $q_0 \geq 0$ p-values ≤ 0.5. Significance given by

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$$

Under alternative hypothesis μ' (theorem of Wald, non-central Chi²-PDF)

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

Quality of Approximation for Counting Exp

For sensitivity: median of Poisson not known analytically
replace median by expectation value or n by s+b (called Asimov Data)

Gauss approximations:

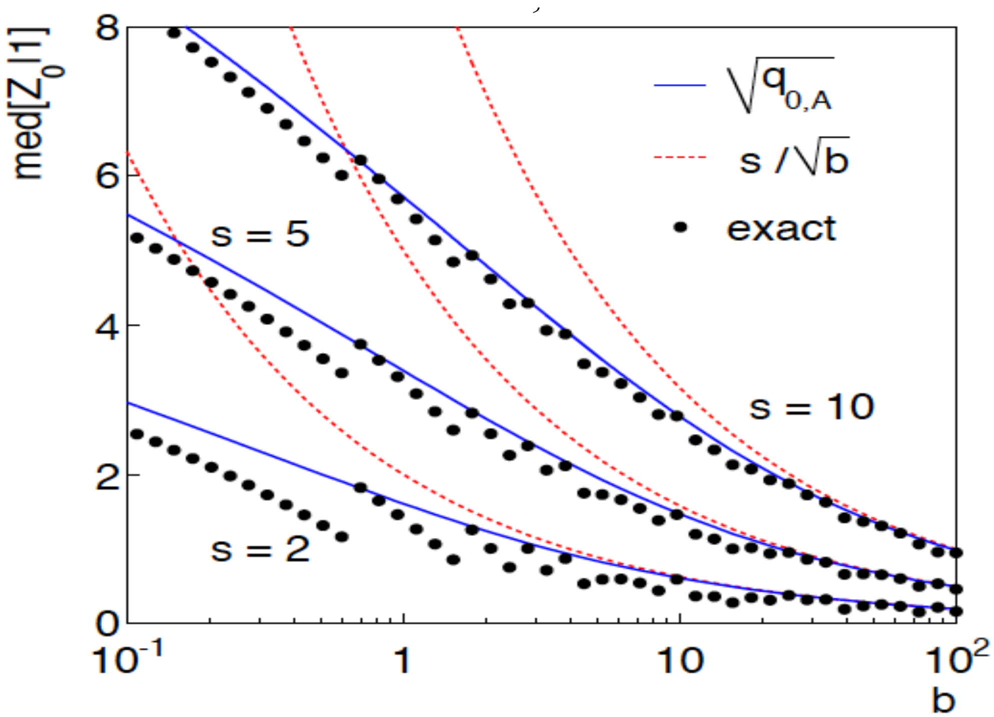
$$Z = \frac{n-b}{\sqrt{b}}$$

$$\mathrm{med}[Z|s+b] = \frac{s}{\sqrt{b}}$$

Wilks approximation
+ Asimov data:

$$Z = \sqrt{2\left(n\ln\frac{n}{b} + b - n\right)}$$

$$Z_A = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right) - s\right)}.$$



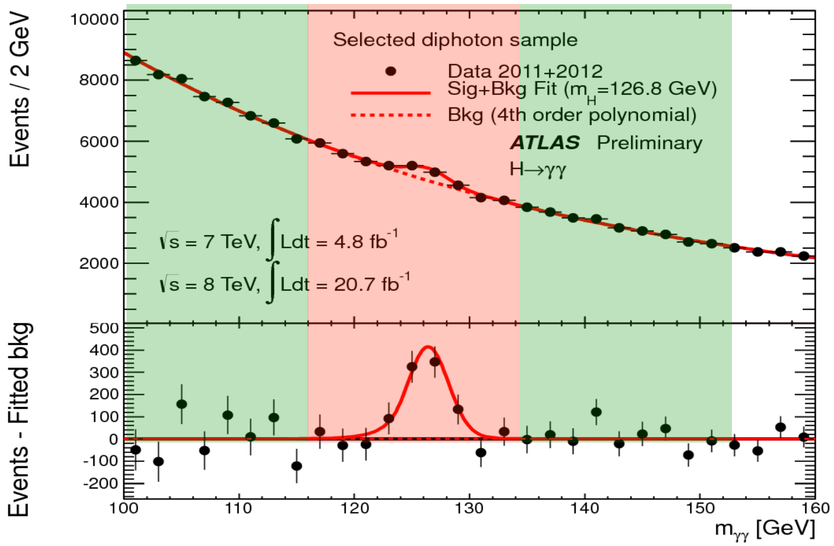
Exact values from toy MC show „jumps“ due to discrete nature of Poisson PDF

Wilks + Asimov Approximation good for large ranges of s and b

s/sqrt(b) only good for s<<b
and b not to small

Profile Likelihood Ratio with b from Control Region

Control region (CR) for background
with expectation τb (b in SR)
Transfer factor τ known (MC, τ>>1)



Observation yields n (SR) and m (CR)

$$n \sim \text{Poisson}(s+b)$$ $$m \sim \text{Poisson}(\tau b)$$

Common likelihood function: $$L(s,b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-(\tau b)}$$

Test statistic = profile likelihood ratio (nuisance parameter b)

$$\lambda(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})}$$

numerator: conditional ML estimate for b given s under H_0

denumerator: unconditional ML estimate for s and b

Advantage: -2 ln λ(s) distributed according to Chi^2-PDF with N_{DOF}=1

Profile likelihood Ratio with b from Control Region

Unconditional maximum likelihood estimates:

$$\hat{s} = n - m/\tau \qquad \hat{b} = m/\tau$$

Conditional maximum likelihood ML estimate for b assuming s:

$$\hat{\hat{b}}(s) = \frac{n + m - (1+\tau)s + \sqrt{(n+m-(1+\tau)s)^2 + 4(1+\tau)sm}}{2(1+\tau)}$$

Conditional ML estimate for b assuming s=0: $\hat{\hat{b}}(0) = \frac{n+m}{1+\tau}$

Plugging this in yields $\lambda(s=0)$ and – 2n $\lambda(s=0) = q_0$

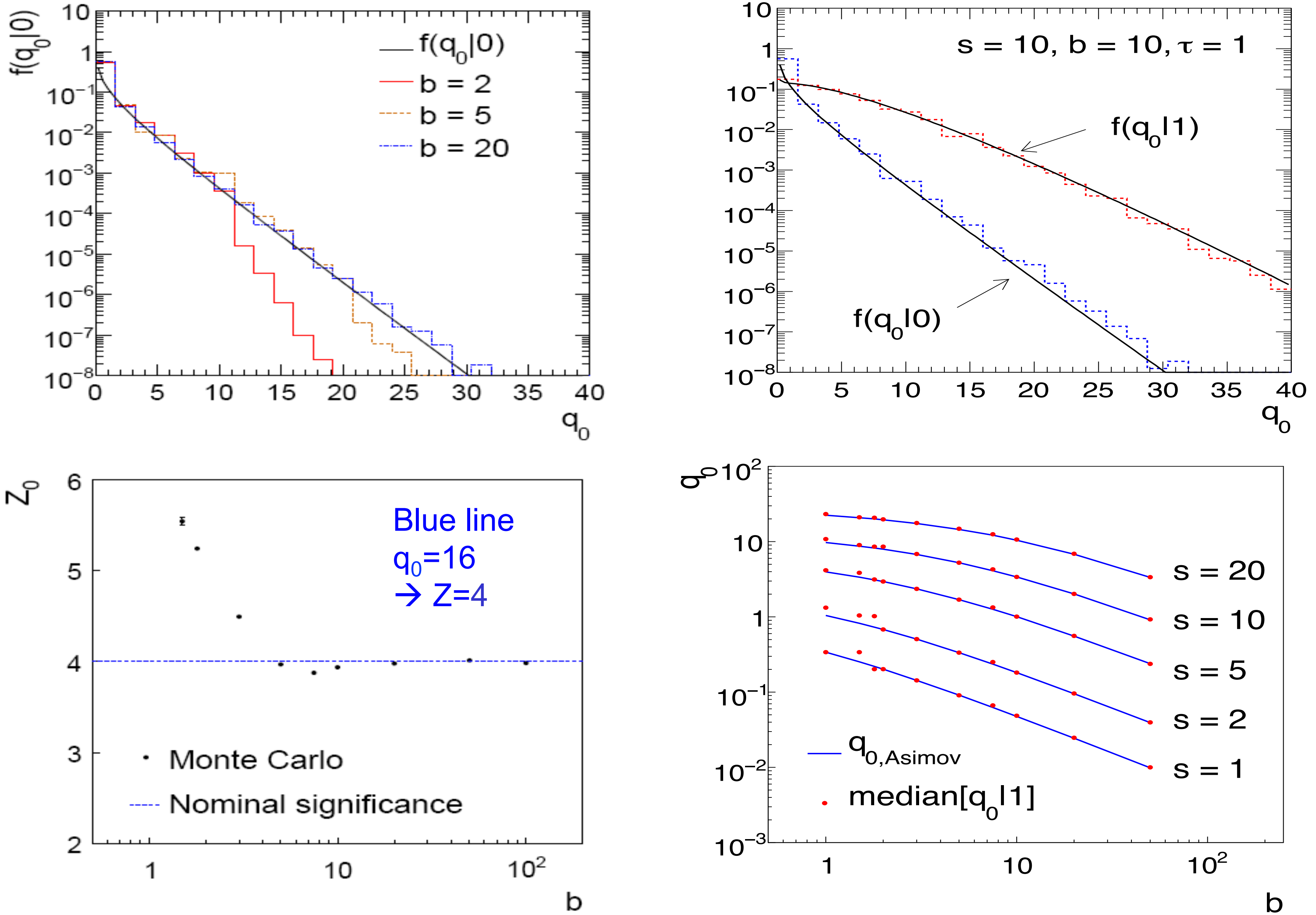
Using Wilks approximation and Asimov data set ($n=s+b$, $m=\tau b$) yields:

$$Z = \left[-2\left(n \ln\left[\frac{n+m}{(1+\tau)n}\right] + m \ln\left[\frac{\tau(n+m)}{(1+\tau)m}\right]\right)\right]^{1/2}$$

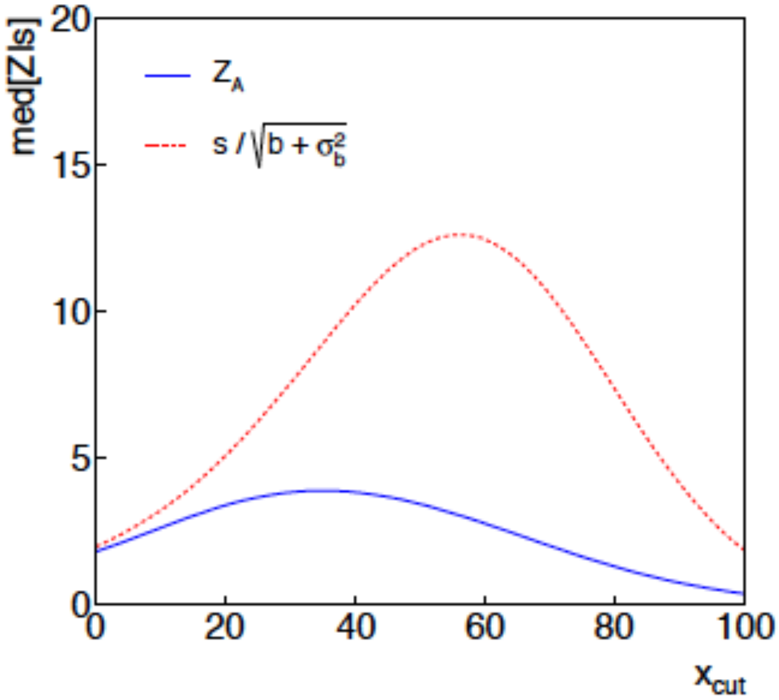
Similar more lengthy expression for (expected) exclusion significance

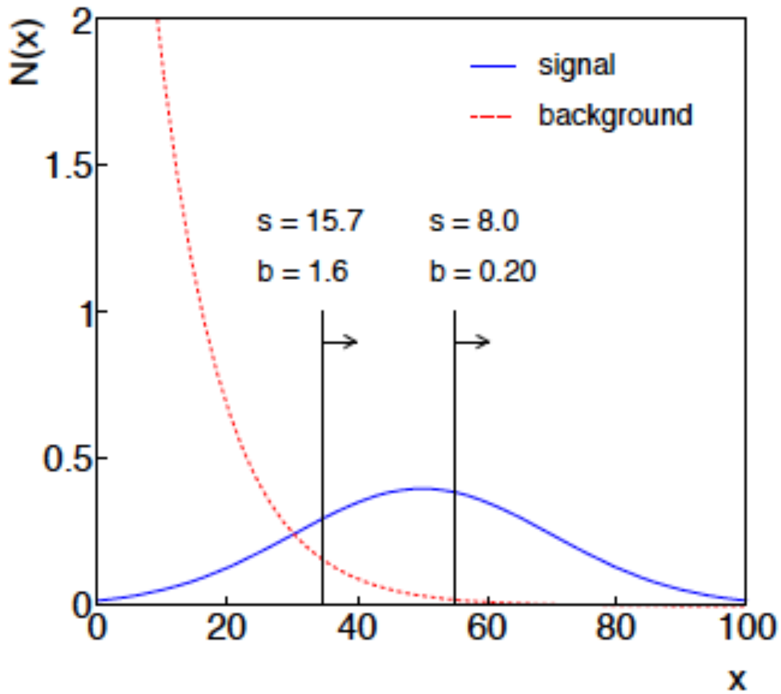
$$Z_A = \left[-2\left((s+b)\ln\left[\frac{s+(1+\tau)b}{(1+\tau)(s+b)}\right] + \tau b \ln\left[1 + \frac{s}{(1+\tau)b}\right]\right)\right]^{1/2}$$

Quality of Approximation for $\tau = 1$



Optimising the Sensitivity of a Selection





Different optimal work point in selection for small background yields

Translation of τ to relative background uncertainty σ_b/b

$$V[\hat{b}] \equiv \sigma_b^2 = \frac{b}{\tau}$$

$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

In the limit of small s/b and small σ_b/b this reduces to

$$Z_A = \frac{s}{\sqrt{b+\sigma_b^2}}\left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b)\right)$$

Using Distributions / Combination of Channels

Consider each bin in final discriminant and channels as independent counting exp.

Correlation among bins: a) common signal strength $\mu = \sigma/\sigma_{standard}$
b) sys. uncertainties described by nuisance parameters θ

$(\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_{tot})$

Expectation in bin of SR and CR: $E[n_i] = \mu s_i + b_i \quad E[m_i] = u_i(\boldsymbol{\theta})$

Observation n bins of SR and m bins of CR: $\mathbf{n} = (n_1, \ldots, n_N) \quad \mathbf{m} = (m_1, \ldots, m_M)$

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

μ fixed assuming H_0

$\hat{\hat{\boldsymbol{\theta}}}$ conditional ML estimate assuming H_0

$\hat{\mu}, \hat{\boldsymbol{\theta}}$ unconditional ML estimate

Again: PDFs for -2 ln $\lambda(\mu)$ known due to Wilks‘ and Wald’s theorems

Profiled Likelihood Test Statistic for Discovery

H_0: only background → $\mu=0$ H_1: signal and background,
μ parametrises strength w.r.t. "standard predicton"

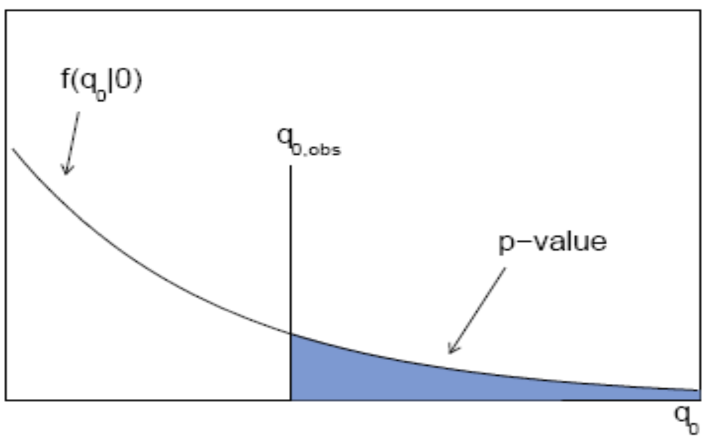
Test statistic q_0:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

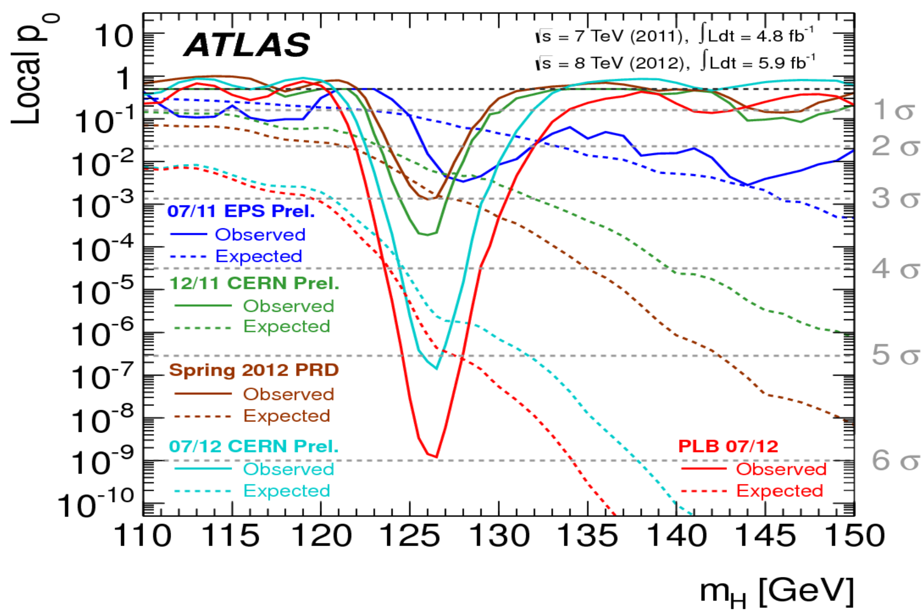
$\lambda(0)$ btw. 0: H_1 like and 1:H_0 like

→ q_0 between 0 and infinifity
0: H_0 like   >> 0 H_1-like

One sided test, only positive signal strength considered as deviation from H_0



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



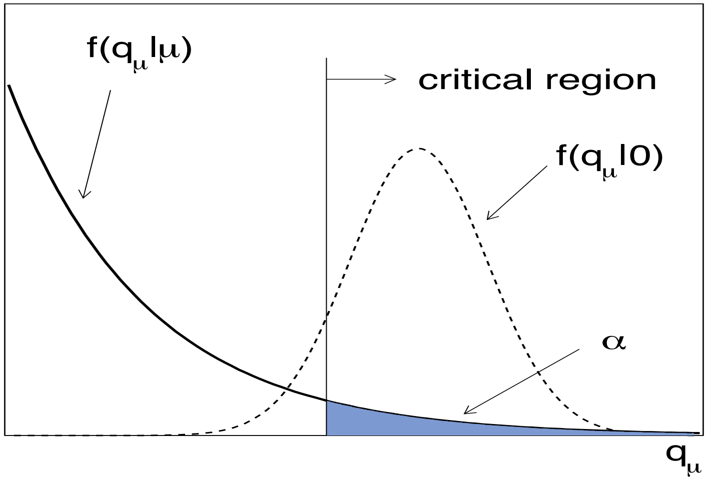
Profiled Likelihood Test Statistic for Exclusion

H_0: signal+background → μ=1 H_1: background only
μ parametrises strength w.r.t. "standard prediction"

Test statistic q_μ:

$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0, \\ \dfrac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 . \end{cases} \qquad \tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

For negative signal strength set it to 0 and determine then nuisance pars.
One sided test, only signal strength < μ considered as inconsistent with H_0



different test statistic then for discovery

here values ~0 are signal+background like observations

$\mu^{95\%\mathrm{CL}}$ decrease tested μ until
P-value = significance level α

The „Problem“ with the Pure Frequentist Method

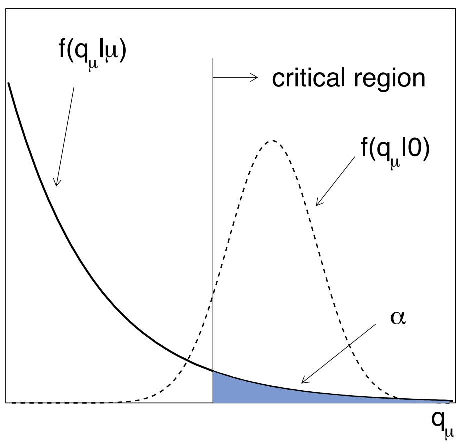
$$p_\mu \;=\; P(\,\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}\,|\,\text{signal+background}) \;=\; \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})\,d\tilde{q}_\mu$$

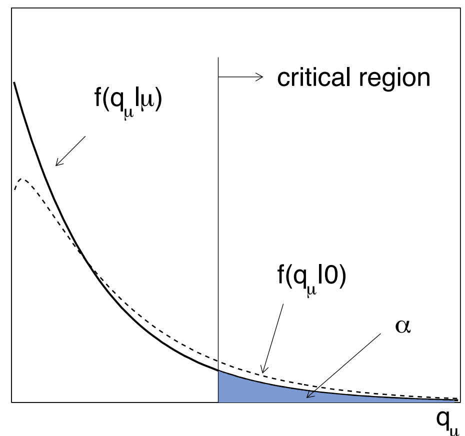
Pure frequentist would stop and say: „signal + background“ hypothesis is excluded with a confidence level CL_{S+B} of 1- p_μ

„Problem“: Spurious exclusion of signals with no sensitivity (s<<b)

large s

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





s<<b

powerM = 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

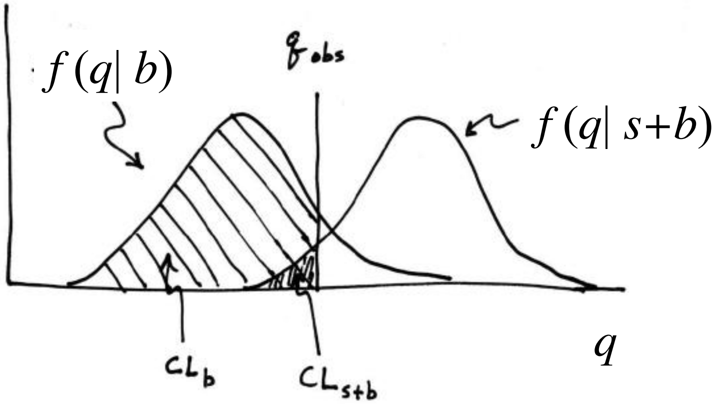
→ probability to exclude hypotheses with zero signal (due to downwards fluctuation) ~ α „spurious exclusion w/o sensitivity“

„Solutions“ to the „Problem“

1) CL_S technique (A. Read, T. Junk)
ad hoc correction to “normal” p-value(s+b) p_μ

$$p_\mu \rightarrow \frac{p_\mu}{1-p_b} = CL_S$$

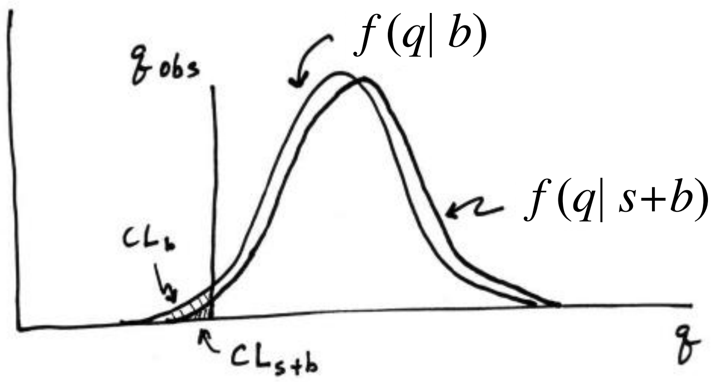
$$CL_S = \frac{CL_{S+B}}{CL_B} = \frac{Prob(t \leq t_{obs}|s+b)}{Prob(t \leq t_{obs}|b)}$$



hypothesis rejected at CL= 1-α if

$$CL_S \leq \alpha$$

(true error of
1st kind smaller)

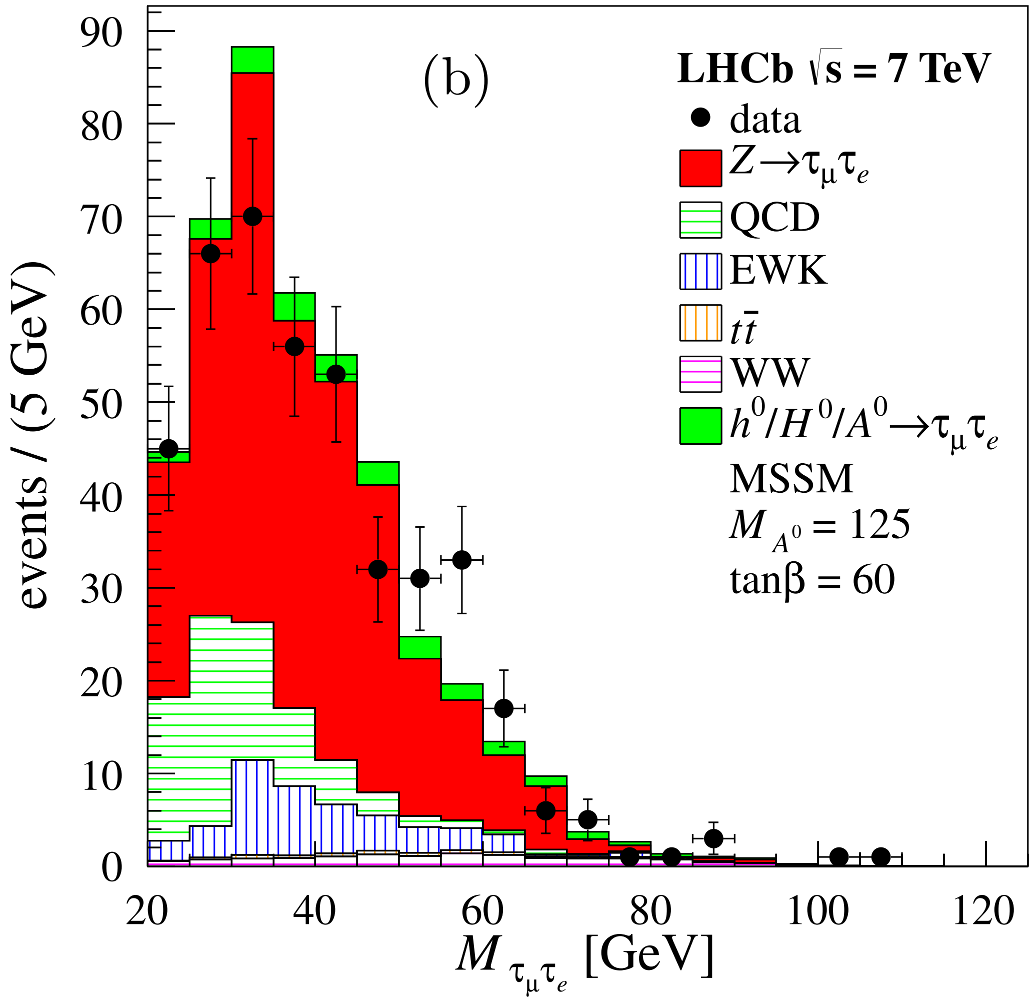


2) Power constrained tests: two requirements for rejection

a) $p_\mu < 5\%$ b) minimal power in test against H_1 „b-only“
$M = 1\text{-}ß = 1 - p_b > 16\%$; $> 50\%$, ...

and in the context of upper limit setting:
3) Bayesian Limits or 4) Feldman-Cousins Limits

Limit Setting / Confidence Intervals



Give a confidence interval for signal strength at xy% confidence level

two-sided: $[s-\Delta s_1, s+\Delta s_2]$ one sided = upper limit s

Either decide before measurement or do always both or use FC unified approach

Confidence Intervals CI

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
- "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
- as usual: needs to assume an a-priori probability

Classical Frequentist Intervals

Neyman construction for equal tailed CI at CL = 1 – α –β = 1- γ α=β=γ/2

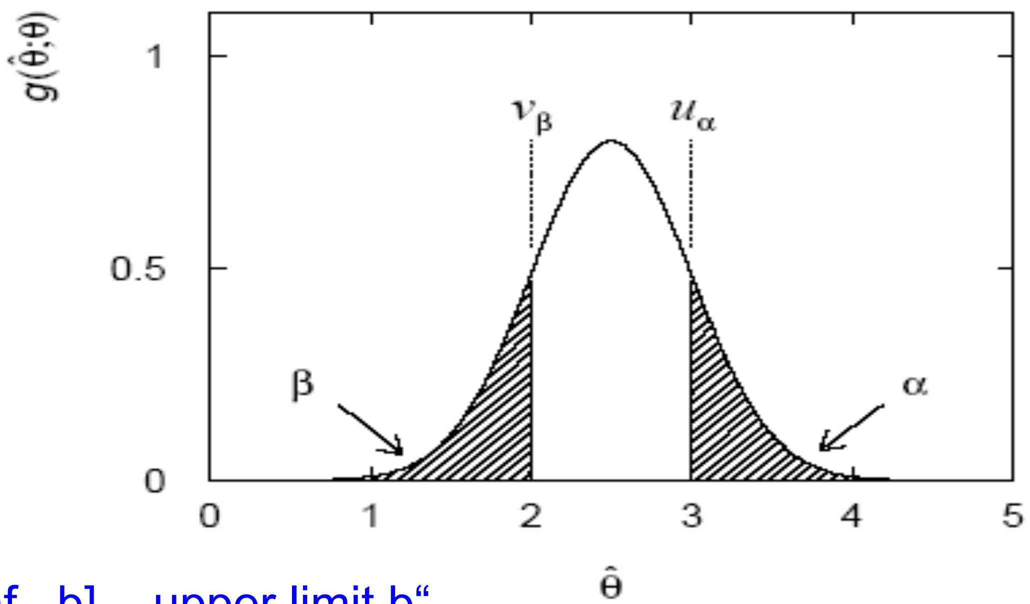
Consider: estimate $\widehat{\theta}$ for parameter θ and measured value $\widehat{\theta}_{\mathrm{obs}}$.

Need PDF for estimate for all possible true values θ $g(\widehat{\theta};\theta)$.

Specify tail probabilities e.g. $\alpha = \beta$ = 0.025 (0.16) and determine functions $u_\alpha(\theta)$ und $v_\beta(\theta)$ with:

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

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



for α = 0 , $u_\alpha(\theta)$ → inf →]- inf., b] „upper limit b“

for β = 0 , $v_\beta(\theta)$ → - inf → [a, + inf] „lower limit a“

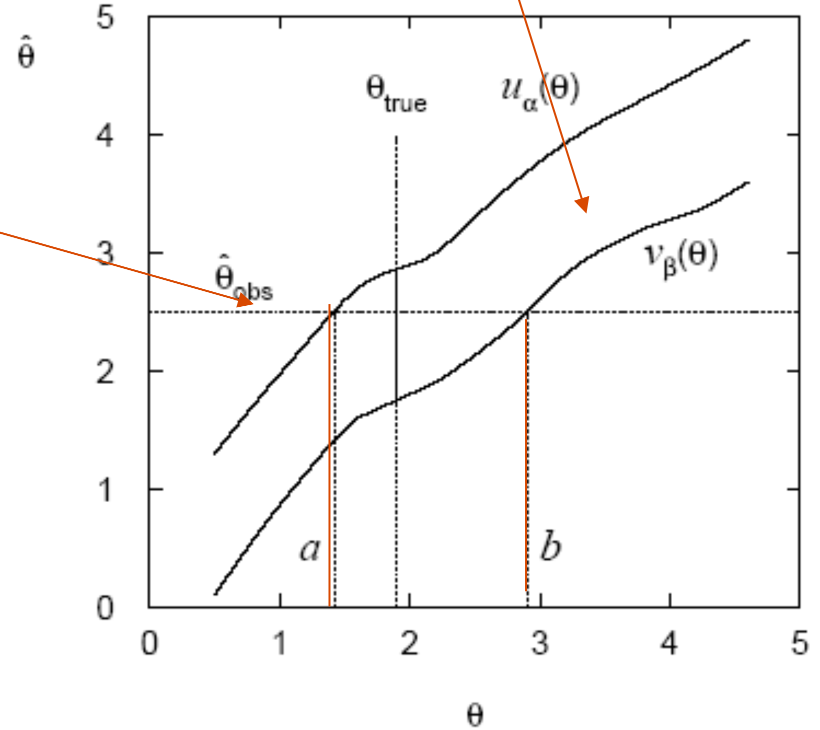
Classical Frequentist Intervals

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:

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

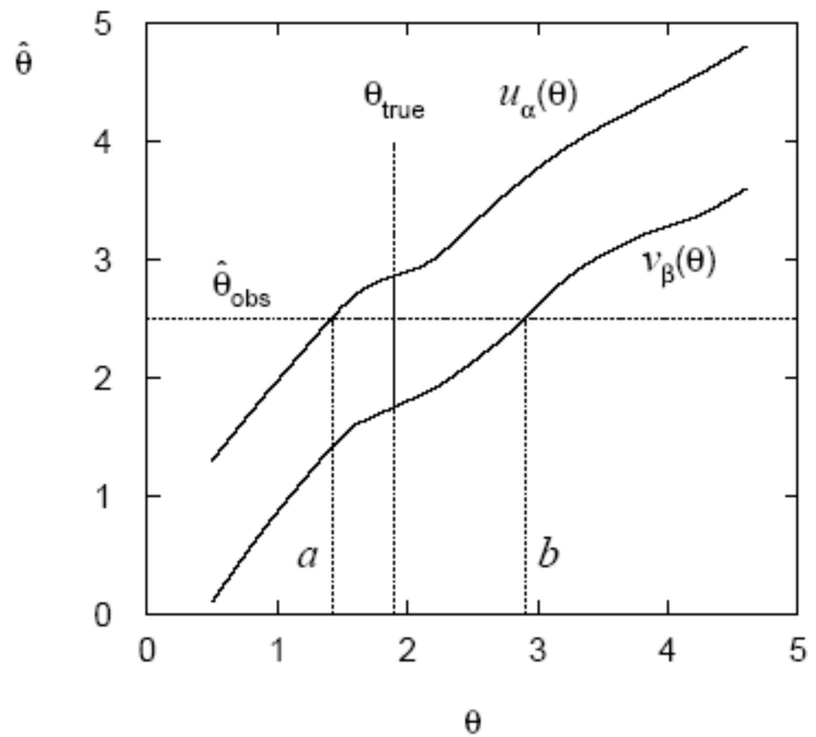
$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta.$$

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 values 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:

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

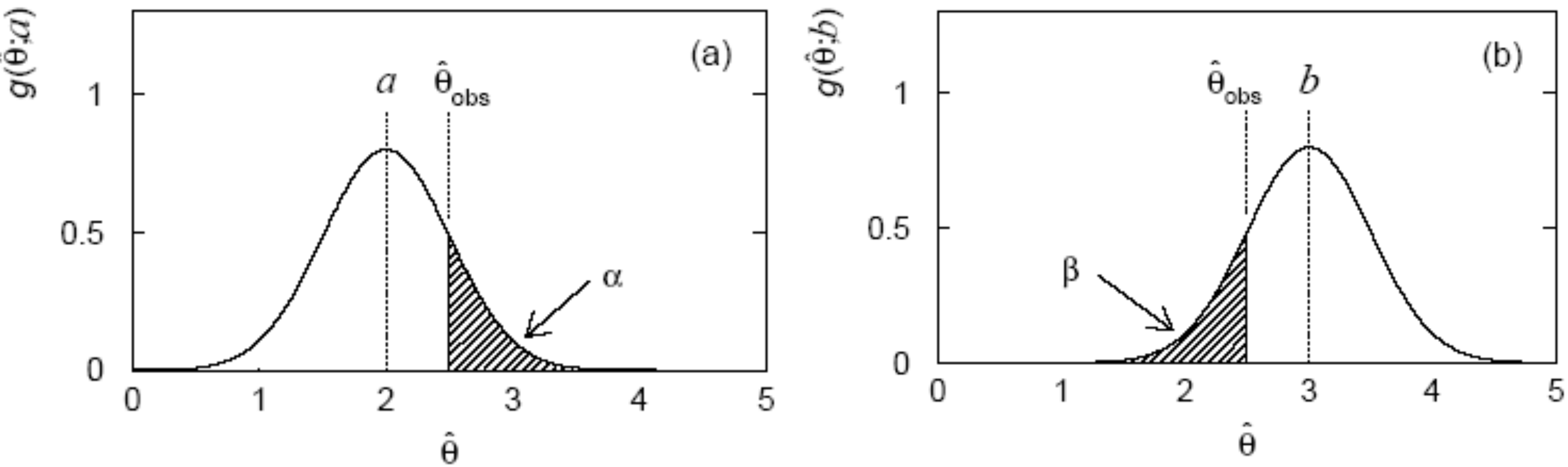
$$\beta = \int_{-\infty}^{\hat{\theta}_{obs}} g(\hat{\theta}; b) d\hat{\theta} = G(\hat{\theta}_{obs}; b).$$

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

→ *b* is 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:

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

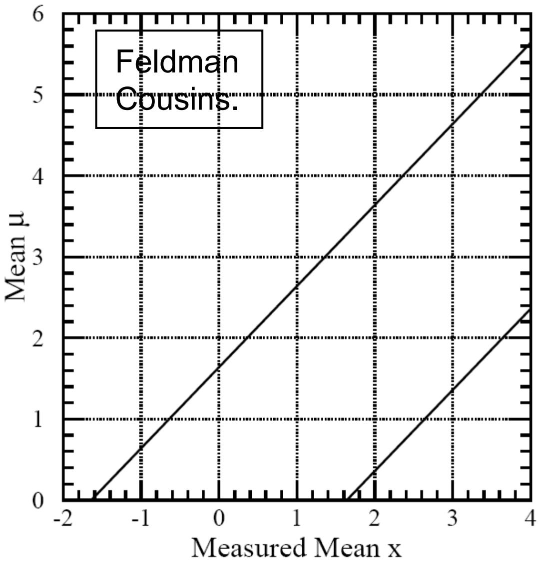
$$b = \hat{\theta}_{obs} + \sigma_{\hat{\theta}}\Phi^{-1}(1-\beta)$$

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 = μ±1,64σ

FIG. 3. Star
Gaussian, in uni

Two sided $\Phi^{-1}(1-\gamma/2)$	$1-\gamma$	One sided $\Phi^{-1}(1-\alpha)$	$1-\alpha$
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:

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

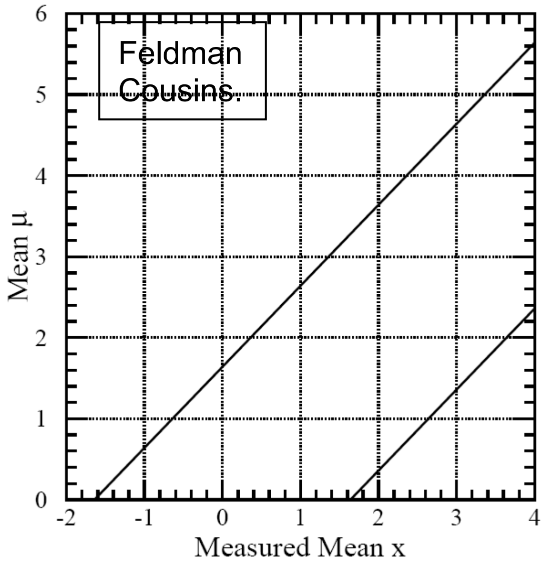
$$b = \hat{\theta}_{obs} + \sigma_{\hat{\theta}}\Phi^{-1}(1-\beta)$$

For $\alpha=\beta= 0.16$

1-σ Intervall

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

Confidence belt for σ=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

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

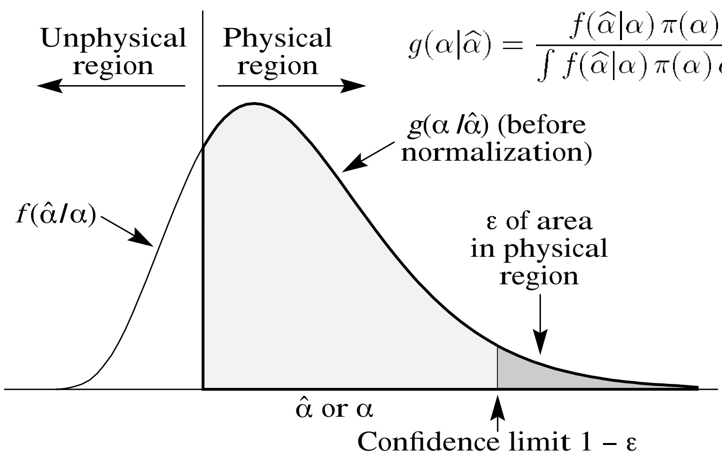
$$L(x;\mu) = \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}}$$

A-priori probability

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

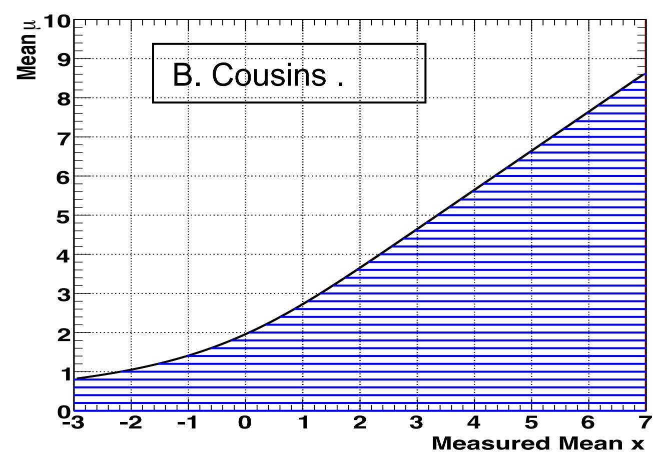
Yields ratio of two integrasl over Gauss PDF starting at physical boundary

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



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

Bayesian upper limit at 95% CL



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

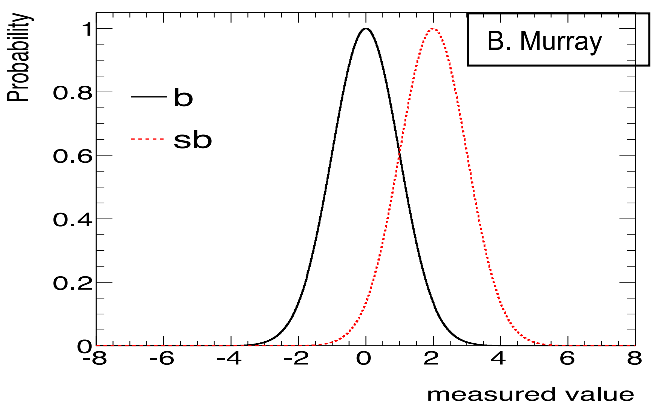
Gaussian example: small (large) value of x inconsistent with μ (μ=0) hypothesis

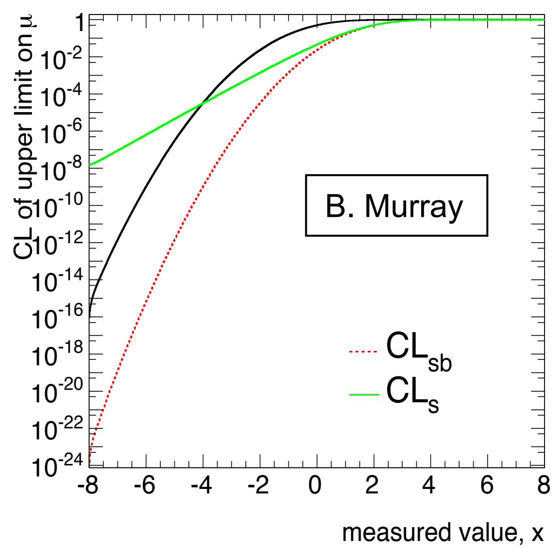
$$CL_S = \frac{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}}} dx}{\int_{-\infty}^{x_{obs}} dx \exp^{-\frac{(x-0)^2}{\sqrt{2\pi\sigma^2}}} dx}$$

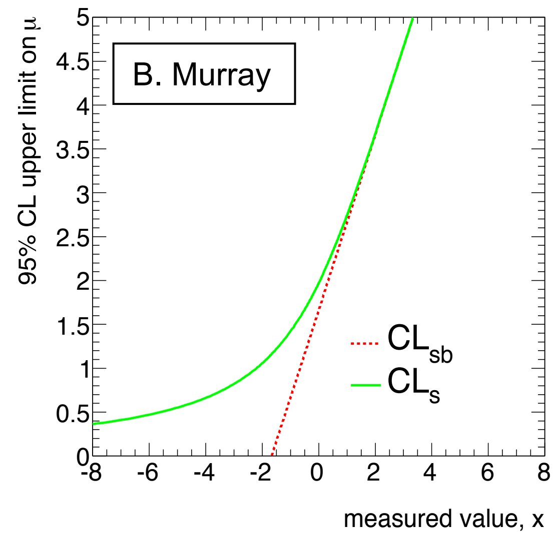
Numerically identical to Bayesian limit

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

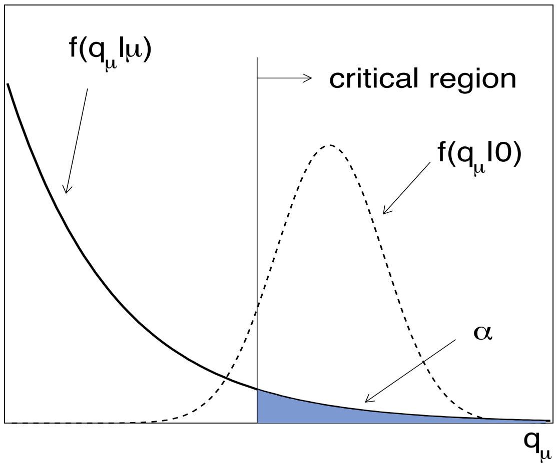
CL_{SB} = classical limit
CL_S = CL_S limit







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

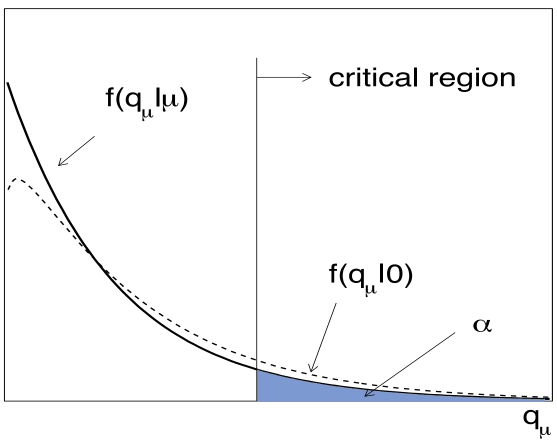


Upper limit from inversion of hypothesis test

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

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

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

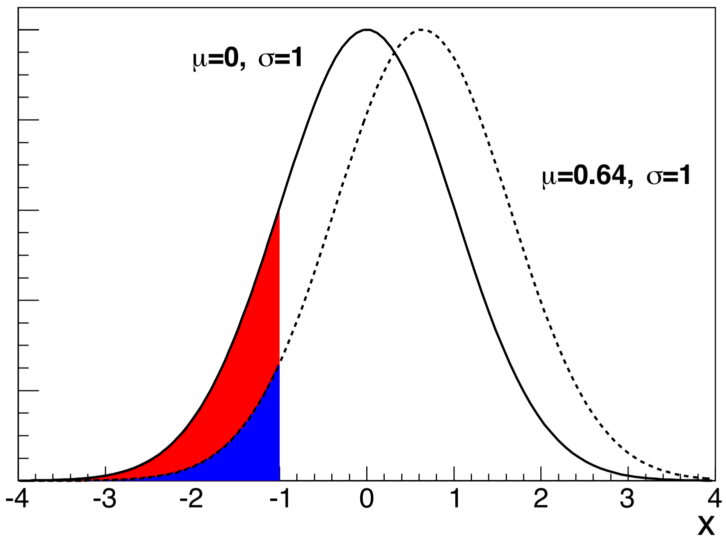
Procedure: determine "usual" upper limit μ_{up}

Find minimal μ value which has minimal power M_{min} μ_{min}

The PCL μ^*_{up} is then given by larger of the two: $\mu^*_{\rm up} = \max(\mu_{\rm up}, \mu_{\rm min})$

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

PCL for Gauss-PDF with μ‘ = 0



Critical region in a test of μ with size α

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

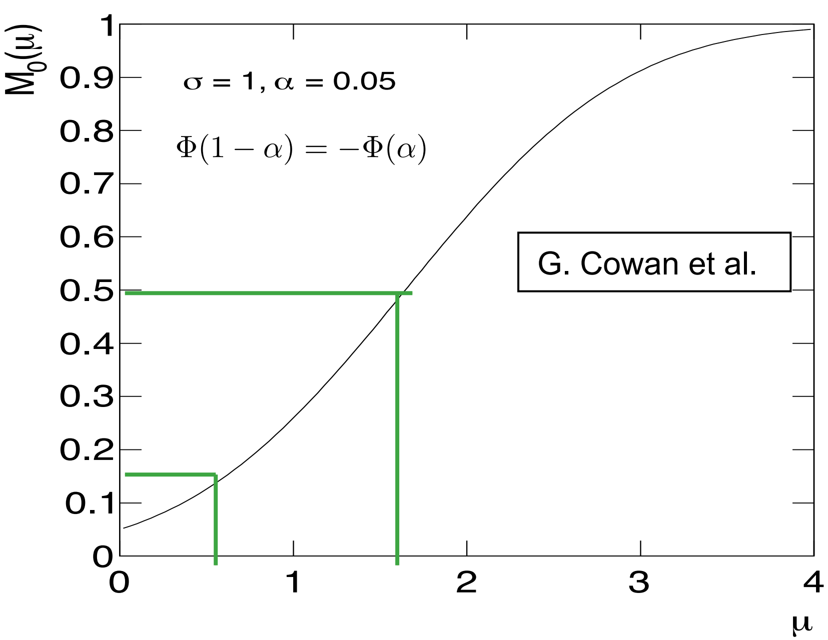
The “usual” limit is then given by:

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

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

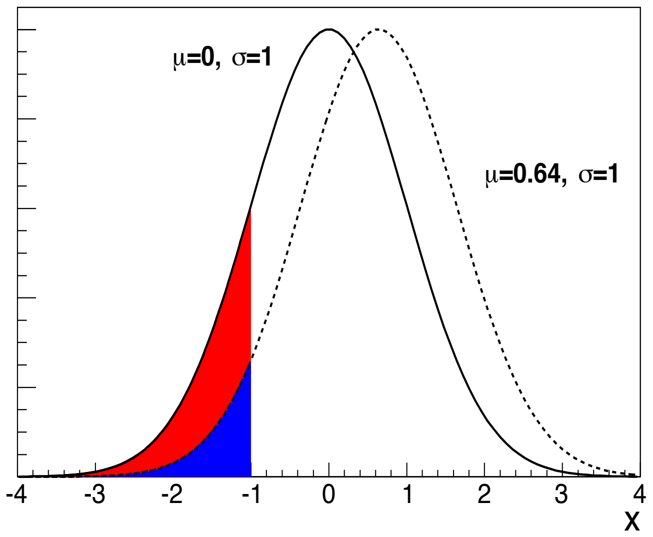
$$M_0(0) = \alpha$$

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

M_{min}= 16% → μ_{min} = 0.64

M_{min} = 50% → μ_{min} = 1.64

PCL for Gauss-PDF with μ' = 0

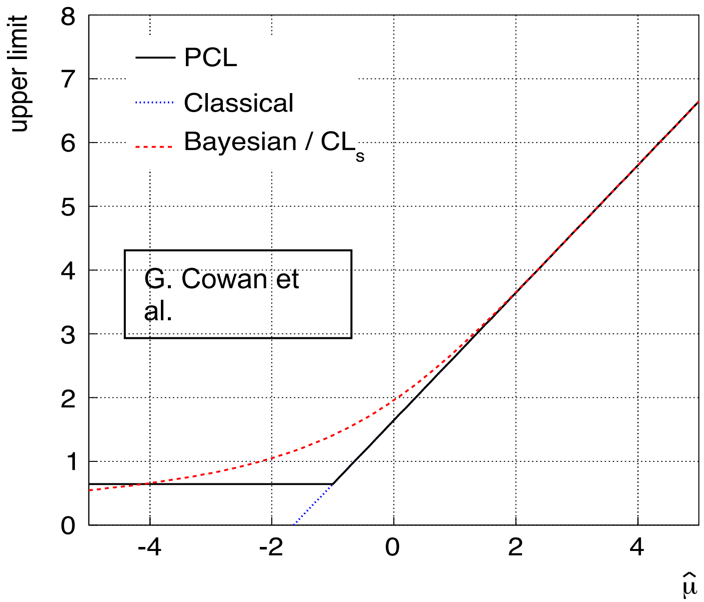


Requiring a minimal power $\Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right) \geq M_{\min}$

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

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

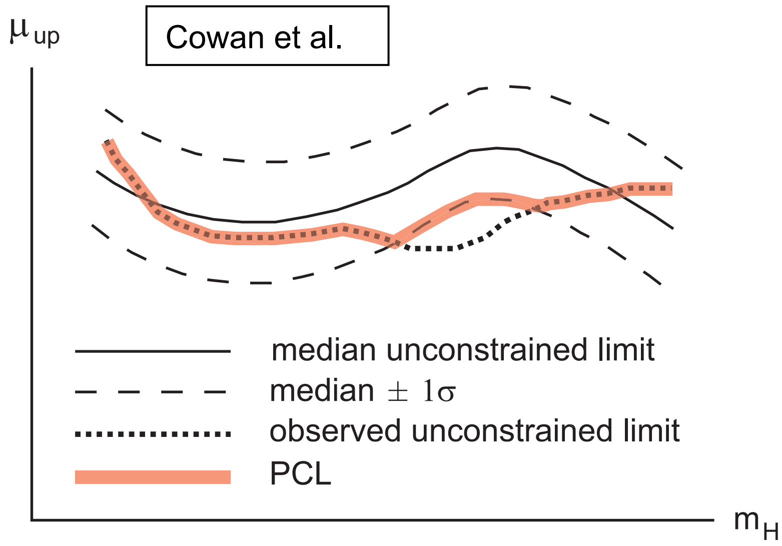
For $\alpha = 0.05$ $M_{min} = 16\%$ $\sigma = 1$

μ_{up} = μ_{meas} + 1,64 μ_{min} = -1+1.64 = 0.64

μ^*_{up} = max (-1, μ_{meas}) + 1.64

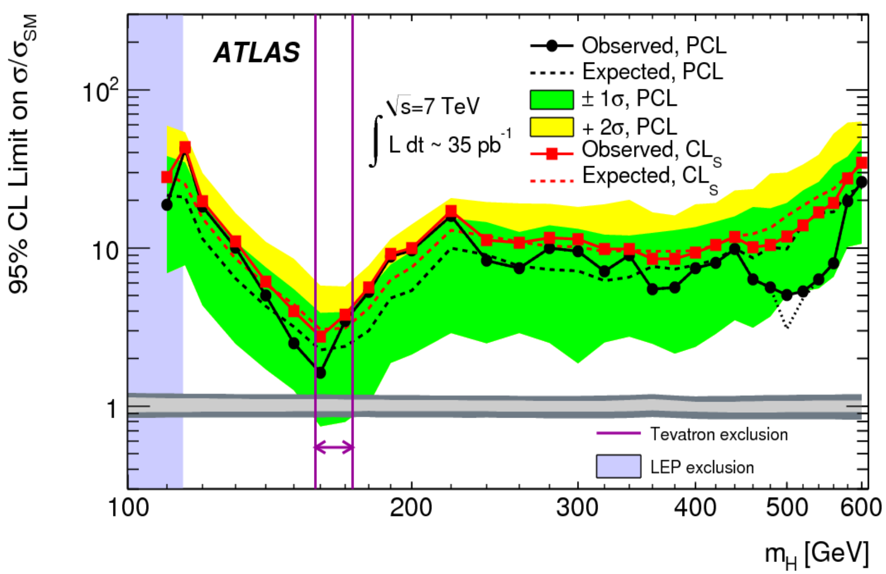
$$\alpha = 0.05 \text{ gives } \Phi^{-1}(1-\alpha) = 1.64$$

Power Constraint Limits at Work



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

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



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

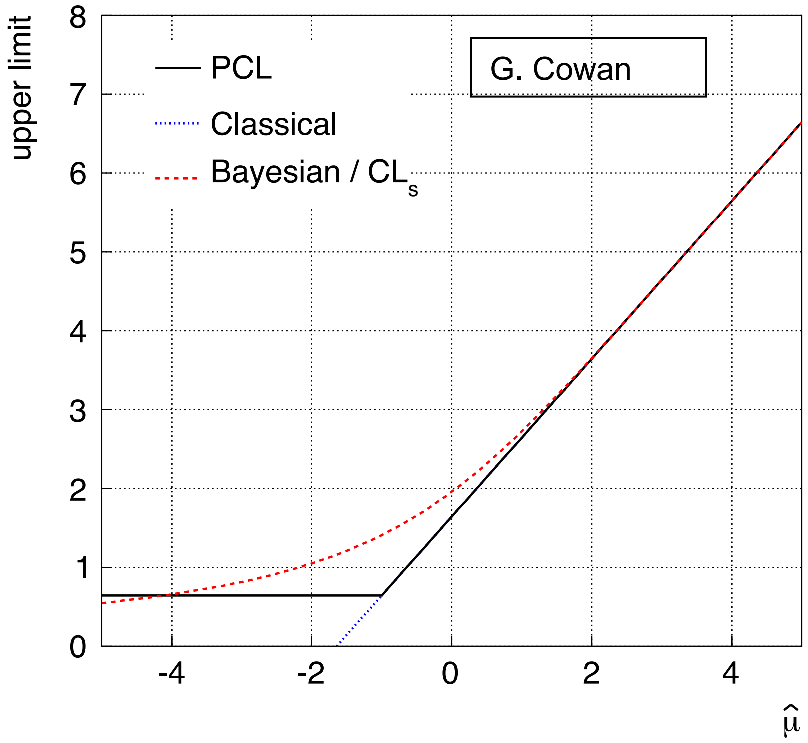
expected limit: median value of μ which will be excluded under BG-only green and yellow bands are 68% (95%) confidence intervals around this

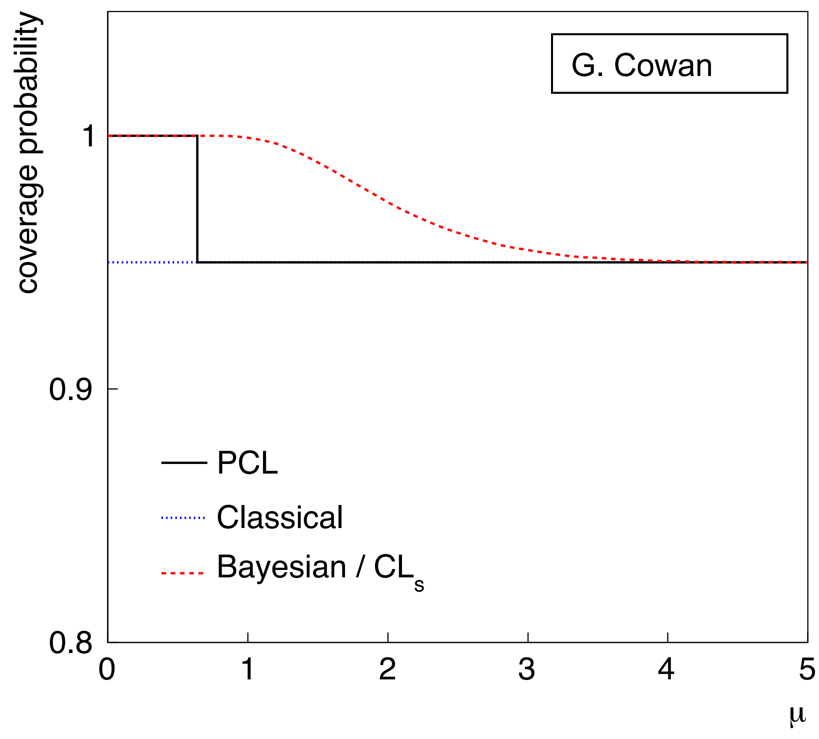
expected CL_S limit worse due to division by 1-p-value(b-only) = 0.5 on average

Different 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

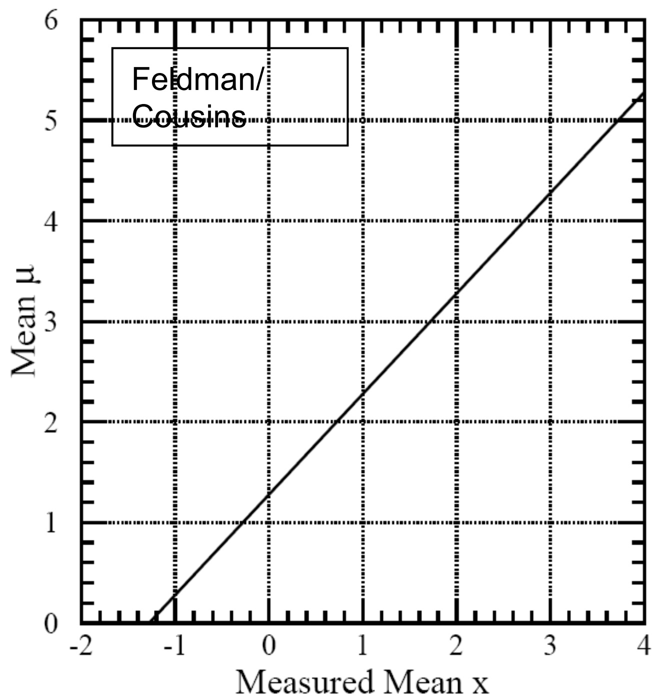
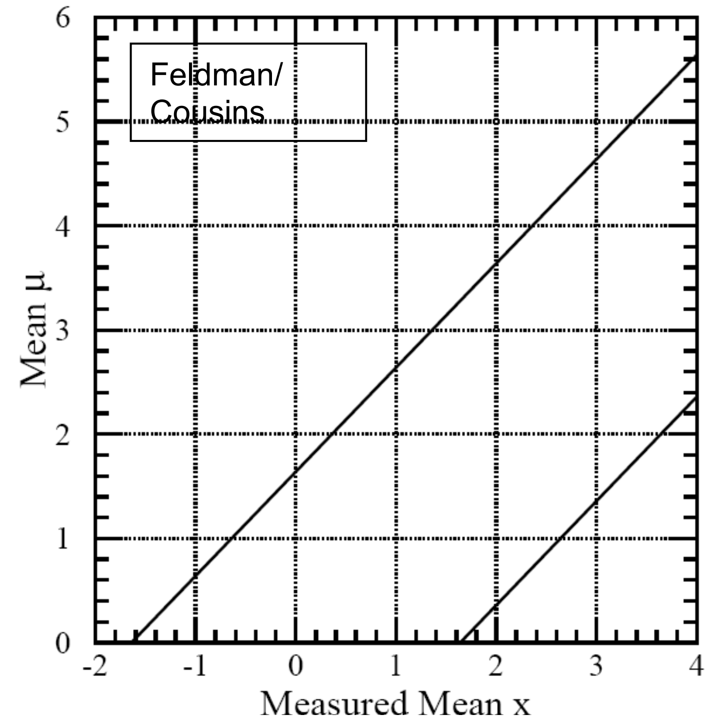


FIG
units o

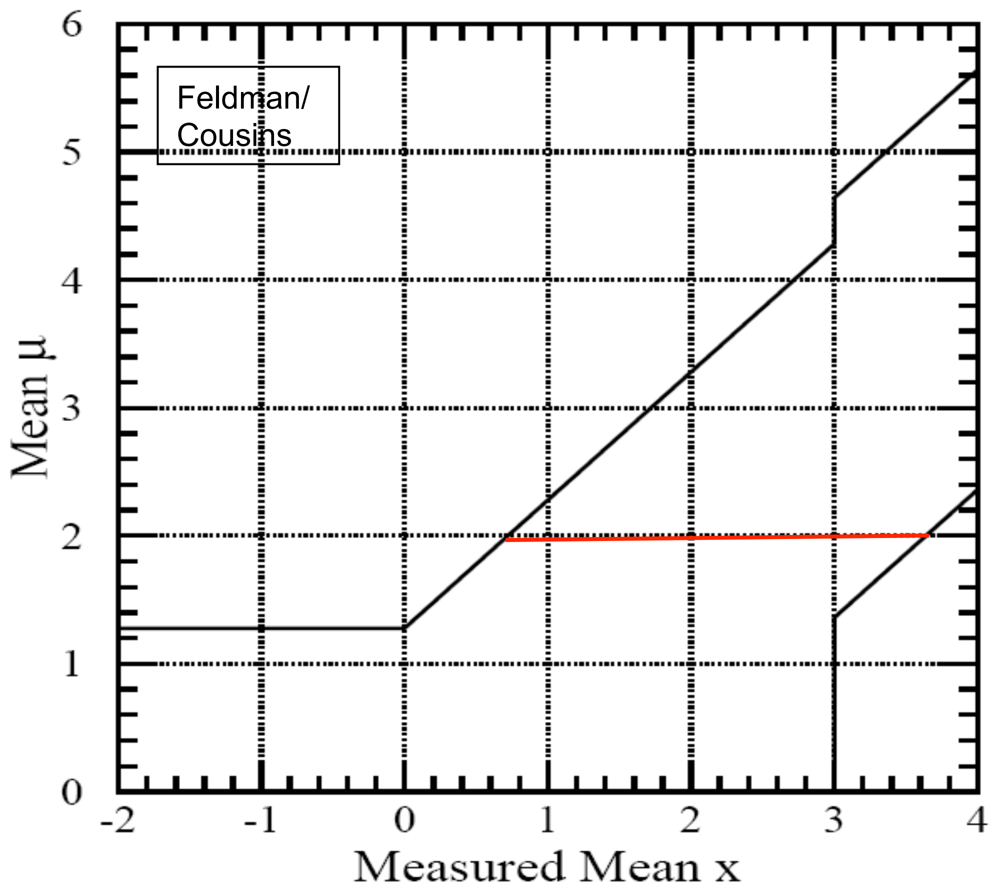
one-sided: measured x + 1.28



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:

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

Likelihood for x assuming μ_{best}

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

Likelihood ratio R(x) defined according to :

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

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 μ≥0
Confidence belt at 90% CL

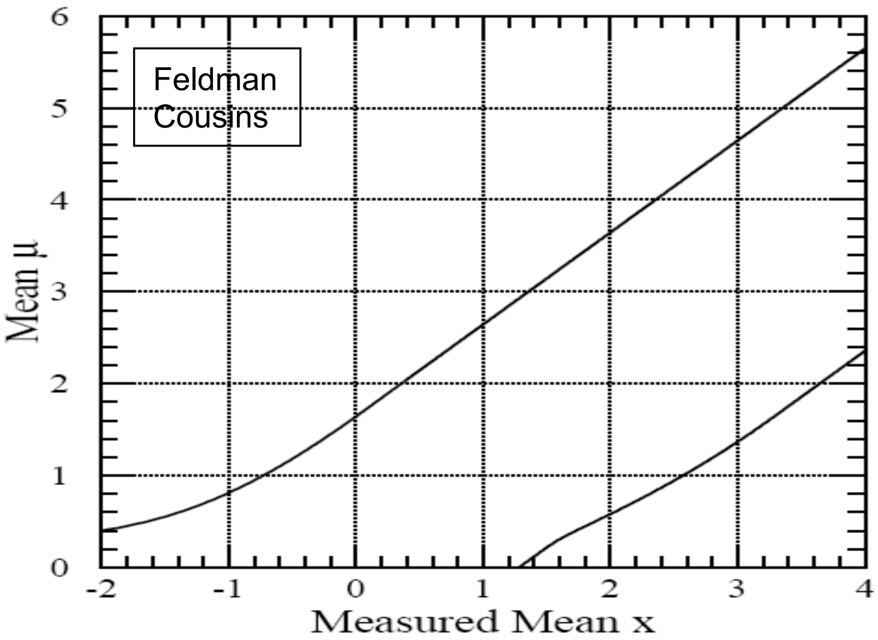


FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

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

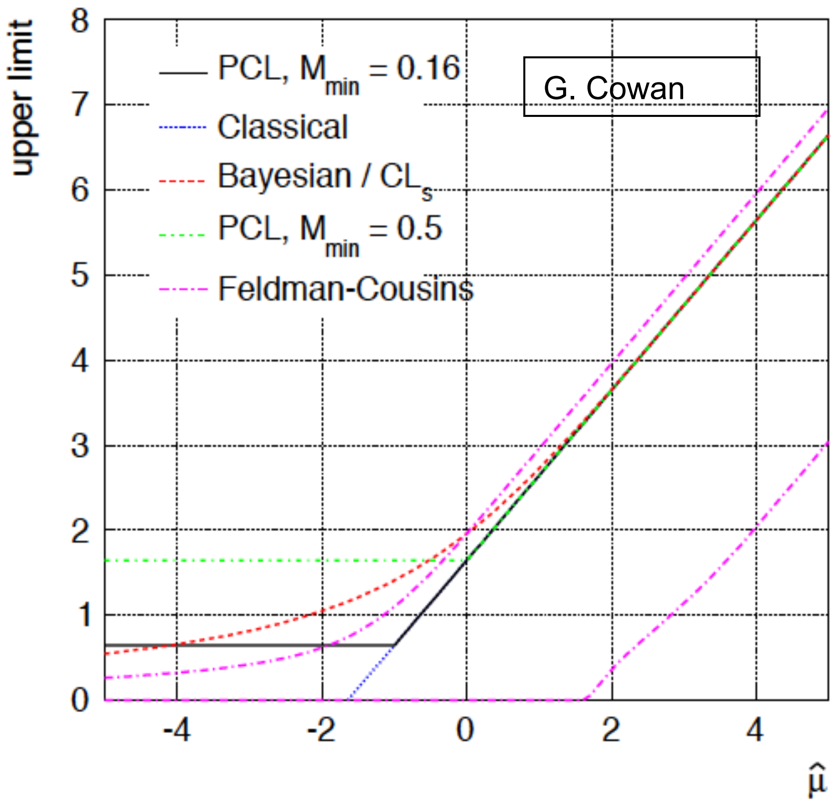
$$\int_{x_1}^{x_2} P(x|\mu)dx = \alpha.$$

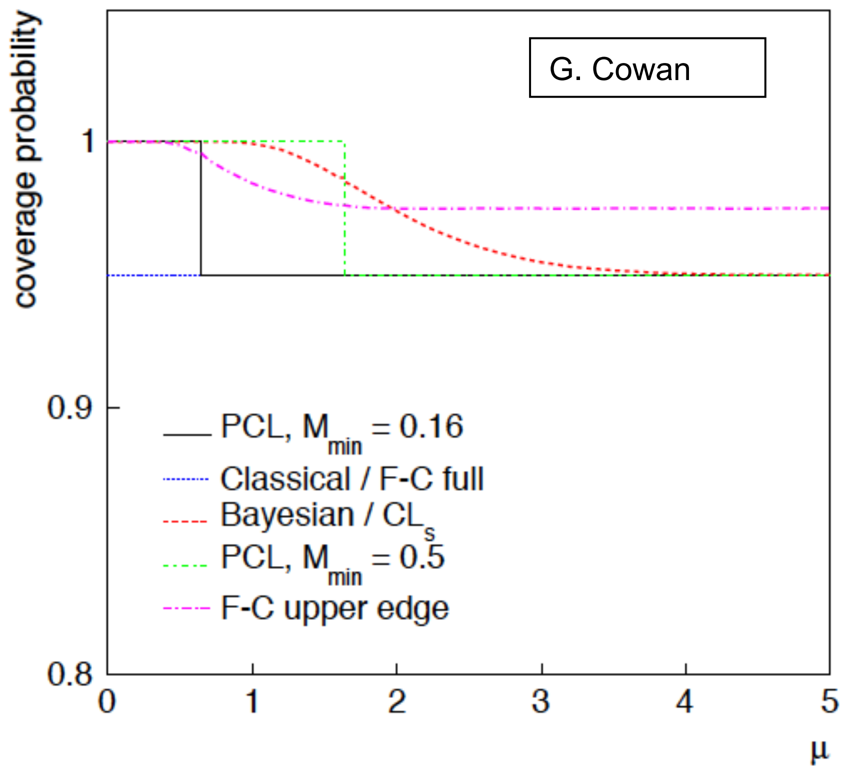
$$R(x_1) = R(x_2)$$

- 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)

Different Upper Limits and Their Coverage

Gauss PDF with variance =1 ,   physical region $\mu \geq 0$
Upper limit at 95% CL and coverage
(PCL with M_{min}=16% (50%), equivalent to replacing observation by -1 if < -1 (0 if < 0))

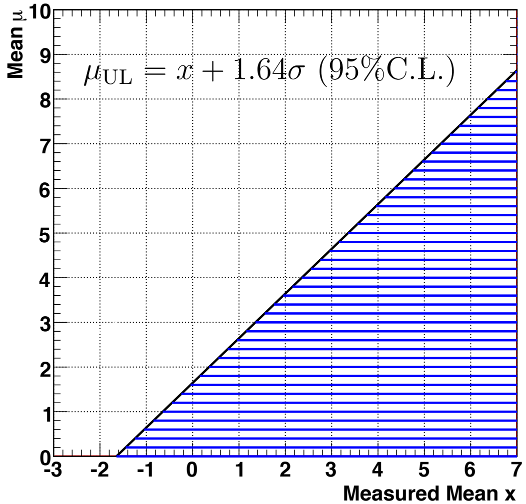




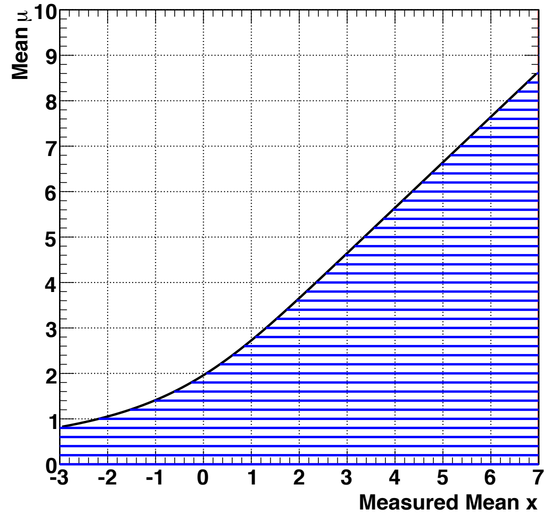
FC gives smallest upper limits for large negative values
FC/unified approach can be supplemented by power constraint

Upper Limits for Gauss-PDF at 95% CL

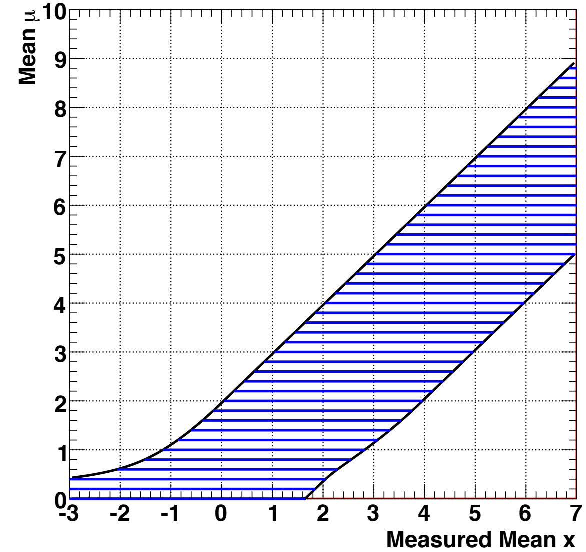
Classical



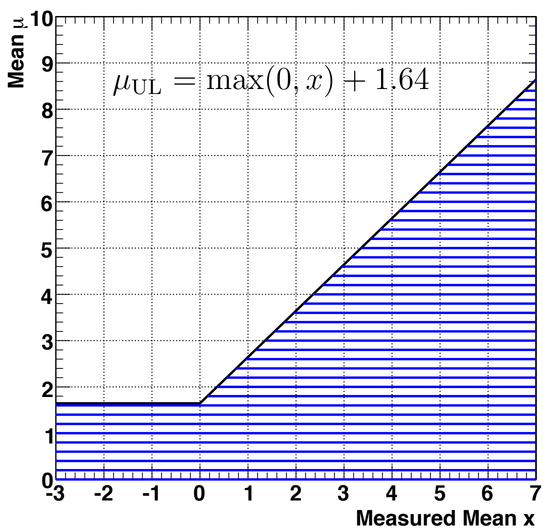
Bayesian / CL_S



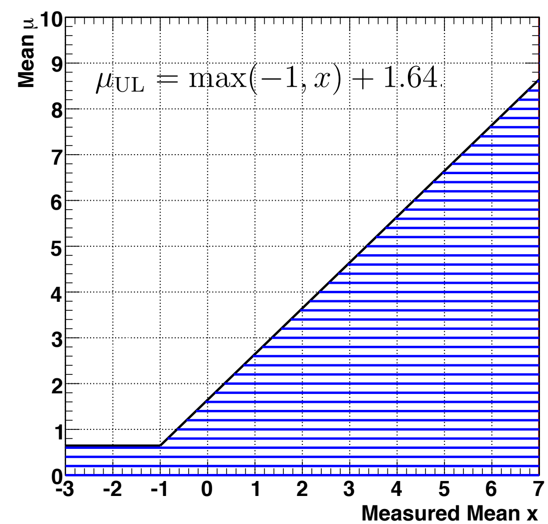
FC unified



PCL 50%



PCL 16 %



Cousins

Confidence Intervals for Poisson-PDF

$$f(n;\lambda) = \frac{\lambda^n}{n!}\exp(-\lambda)$$

n = observed events = ML estimate $\hat{\lambda}$ for λ
Target: confidence interval for λ

Due to the discreteness of n the „confidence belt“ equations can not be fulfilled exactly e.g.

$$\alpha = P(\hat{\lambda} \geq u_\alpha(\lambda))$$

„Conservative“ modification of equations e.g:

$$\alpha \geq P(\hat{\lambda} \geq u_\alpha(\lambda))$$

Hence over-coverage per construction

$$P(l_\beta(\lambda) < \hat{\lambda} < u_\alpha(\lambda) > 1 - \alpha - \beta$$

$$P(a \leq \lambda \leq b) \geq 1 - \alpha - \beta$$

Inversion of test

$$\alpha = \sum_{n=n_{obs}}^{\infty} f(n;a) = 1 - \sum_{n=0}^{n_{obs}-1} f(n;a) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{a^n}{n!}e^{-a},$$

Solve numerically the equtions →

$$\beta = \sum_{n=0}^{n_{obs}} f(n;b) = \sum_{n=0}^{n_{obs}} \frac{b^n}{n!}e^{-b}.$$

Determination of CI for Poisson-Parameter

Simple case: no observed event $\beta = e^{-b} \Longrightarrow b = -\log\beta$

hence at CL = 95% $b = -\log(0.05) = 2.996 \approx 3.$

For general case use relation btw. Poisson-PDF and Chi[2]-PDF

$$\sum_{n=0}^{n_{obs}} \frac{\lambda^n}{n!} e^{-\lambda} = \int_{2\lambda}^{\infty} f_{\chi^2}(z; n_{dof} = 2(n_{obs}+1))dz$$

$$= 1 - F_{\chi^2}(2\lambda; n_{dof} = 2(n_{obs}+1)),$$

The borders of the CI are obtained via the cumulative of the Chi[2]-PDF

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

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

	untere Schranke a		obere Schranke b	
n_{obs}	$\alpha = 0.1$ $CL = 90\%$	$\alpha = 0.05$ $CL = 95\%$	$\beta = 0.1$ $CL = 90\%$	$\beta = 0.05$ $CL = 95\%$
0	-	-	2.30	3.00
1	0.105	0.051	3.89	4.74
2	0.532	0.355	5.32	6.30
3	1.10	0.818	6.68	7.75
4	1.74	1.37	7.99	9.15
5	2.43	1.97	9.27	10.51
6	3.15	2.61	10.53	11.84
7	3.89	3.29	11.77	13.15
8	4.66	3.98	12.99	14.43
9	5.43	4.70	14.21	15.71
10	6.22	5.43	15.41	16.96

Upper limit for Poisson-PDF with Background

Upper limit s at CL=1-γ given by solving the equation from test inversion

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

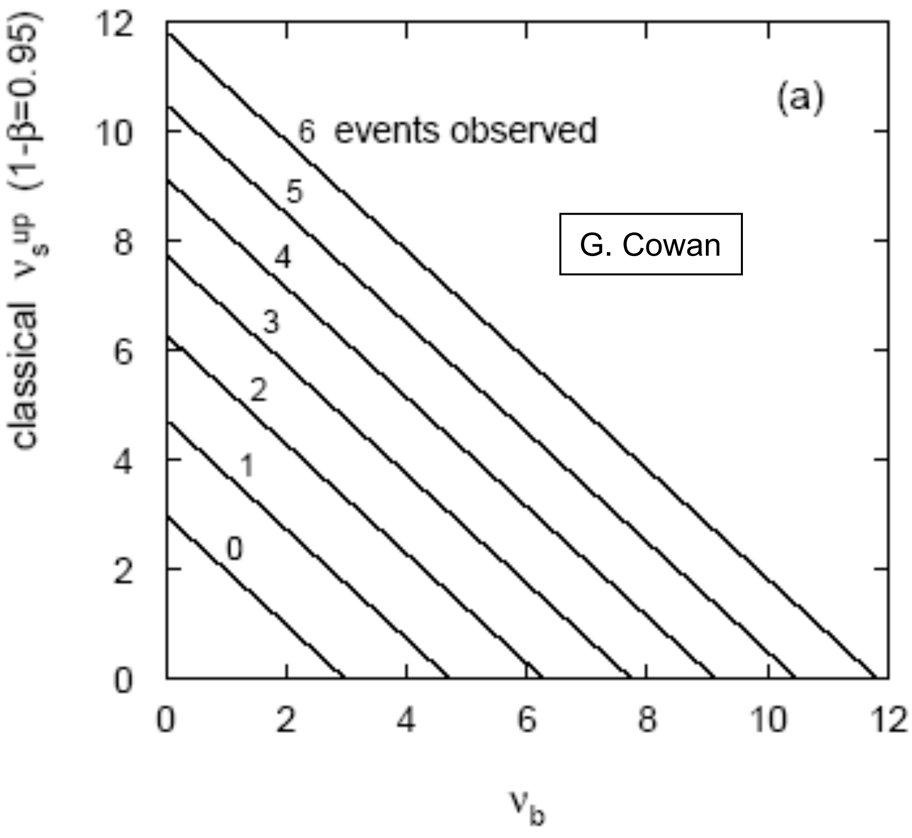
Boundaries of CI s_{lo}, s_{up} determined using Chi²-PDF:

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

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

same as for „b=0" – b
→ called „background subtraction"

$n \leq b$ can yield $s_{up} < 0$



For uncertainty in background use profile likelihood ratio test statistic and Wilk's and Wald's approximation if b not too small.

Expected Limit at Physical Boundary

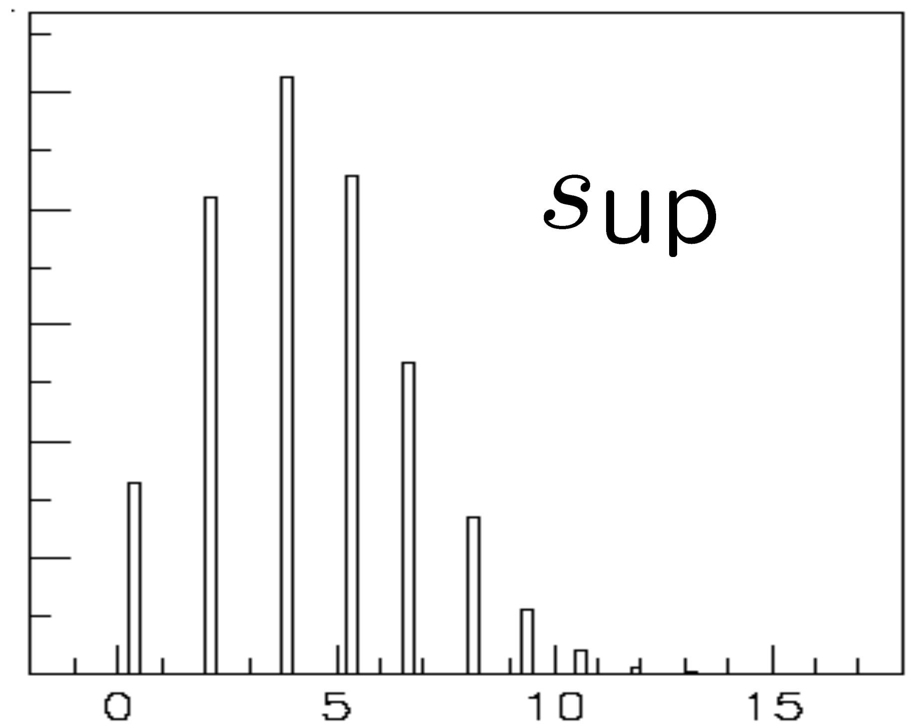
e.g. for b = 2.5 and n = 0 we find upper limit of $s_{up} = -0.197 \quad (CL = 0.90)$

increase CL to 0.95 yields s_{up} = 0.496

„cheating" with CL = 0.917923 yields $s_{up} = 10^{-4}$!

naive argument: for b = 2.5 → variance is $\sqrt{2.5}$ = 1.6.   how can limit be so small?

MC simulation:
determine median limit under
„b-only" hypothesis (s = 0)
→ expected limit



distribution of 95% CL upper limits
for b = 2.5, s = 0.

→ Median s_{up} = 4.44

Bayesian Upper Limit for Poisson-PDF

Bayesian upper limit to CL = 1-α to be derived from

$$1-\alpha = \int_{-\infty}^{s_{up}} p(s|n)ds = \frac{\int_{-\infty}^{s_{up}} L(n|s)\,\pi(s)\,ds}{\int_{-\infty}^{\infty} L(n|s)\,\pi(s)\,ds}$$

.

with likelihood function

$$L(n|s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

and uniform prior in physical region

$$\pi(s) = \begin{cases} 1 & s \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Posterior probability:

$$p(s|n) = \frac{(s+b)^n e^{-(s+b)}}{\Gamma(b,n+1)}$$

$$\Gamma(b,n+1) = \int_b^\infty x^n e^{-x}\,dx$$

Need so solve:

$$1-\alpha = \int_0^{s_{up}} p(s|n)\,ds$$

$$\int_0^a x^n e^{-x} dx = \Gamma(n+1) F_{\chi^2}(2a, 2(n+1))$$

Upper limit given by

$$s_{up} = \frac{1}{2} F_{\chi^2}^{-1}\left[p, 2(n+1)\right] - b$$

Frequentist formula modified by replacing (1-α) by p

$$p = 1-\alpha\left(1 - F_{\chi^2}\left[2b, 2(n+1)\right]\right)$$

Pseudo-Frequentist or Zech‘s Interpretation

Bayesian limit with uniform prior first proposed by O. Helene (1983)
Condition can be rewritten as

$$\alpha = e^{-s_{\mathrm{up}}} \frac{\sum_{m=0}^{n} (s_{\mathrm{up}} + b)^m / m!}{\sum_{m=0}^{n} b^m / m!}$$

Numerical identical result derived by G. Zech (1988) in different context

$$P(n; s+b) = \frac{e^{-(s+b)}(s+b)^n}{n!}$$

stems from

$$P(n; s+b) = \sum_{n_b=0}^{n} \sum_{n_s=0}^{n-n_b} P(n_b; b) P(n_s; s)$$

If N< b we know background in data < b
→ renormalilze background PDF
and replace it in compound PDF

$$P'(n_b; b) = P(n_b; b) / \sum_{n_b=0}^{N} P(n_b; b)$$

Find upper limit s by solving (with $\varepsilon=\alpha$)

$$\epsilon = \sum_{n=0}^{N} P(n; s+b) \Bigg/ \sum_{n_b=0}^{N} P(n_b; b)$$

Zech‘s interpetation →
(not accepted by many Frequentist
as one conditions on data, but known
as the PDG formula for many years)

different. The limit in the “frequency interpretation” can be stated as follows: for an infinitely large number of experiments, looking for a signal with expectation s and Poisson distributed background with mean b, where the background is restricted to values of less than or equal to N, the frequency of observing N or less events is ϵ.

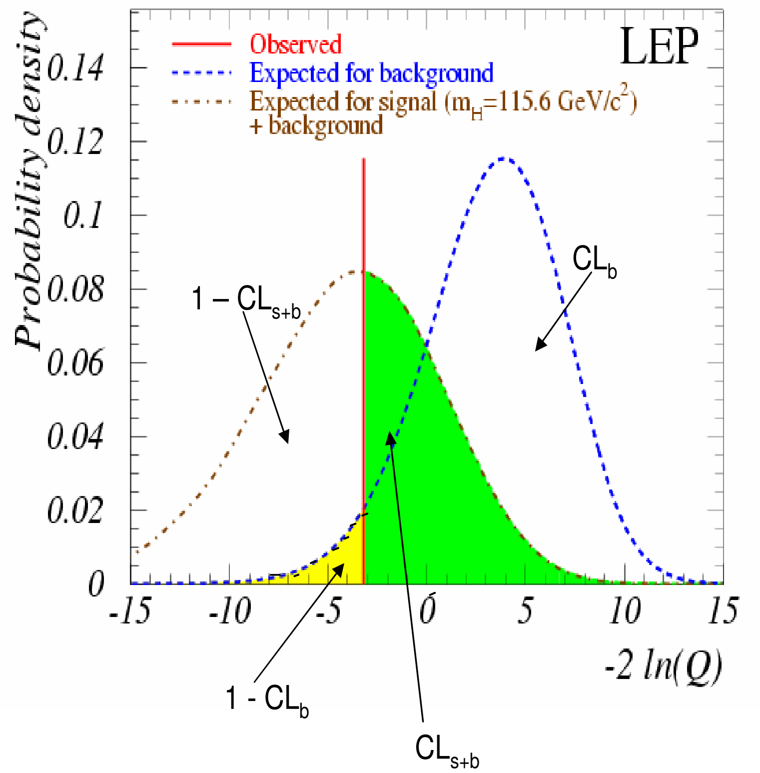
CL_S Limit for Poisson PDF

A. Read (1997): applied Zech's "background conditioning" to the LEP test statistic Q
CL_S ≈ "confidence in the signal-only hypothesis"

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs})$$

$$CL_b = P_b(Q \leq Q_{obs})$$

$$CL_s \equiv CL_{s+b}/CL_b.$$



A hypothesis is exlcuded at confidence level CL if

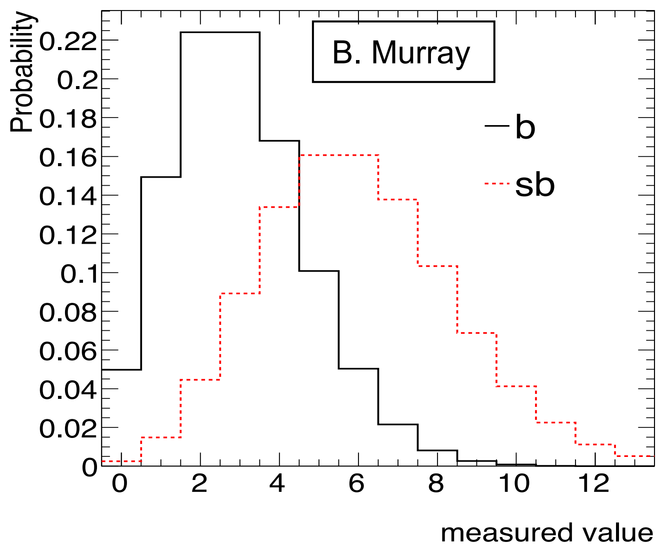
$$1 - CL_s \leq CL$$

Applied to Poisson case yields Zech's formula:

$$CL_s = \frac{P(X \leq X_{obs})}{P(X_b \leq X_{obs})} = \frac{P(n \leq n_{obs})}{P(n_b \leq n_{obs})} \qquad CL = 1 - \frac{\sum_{n=0}^{n_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{e^{-b}b^n}{n!}}.$$

Remark: denominator is <u>not</u> 1-p-value for the b-only hyp. The sum would only run from 0 up to n_{obs}-1. Calling it the power is correct (I think)

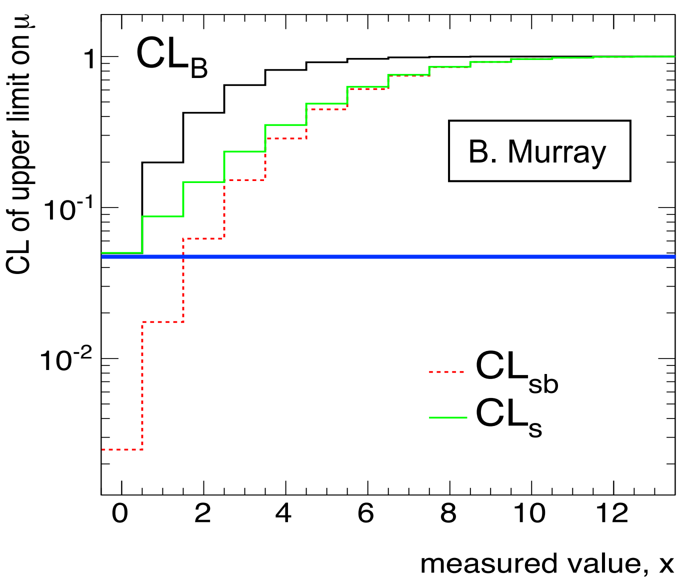
Classical and CL_S Limit compared for Poisson PDF

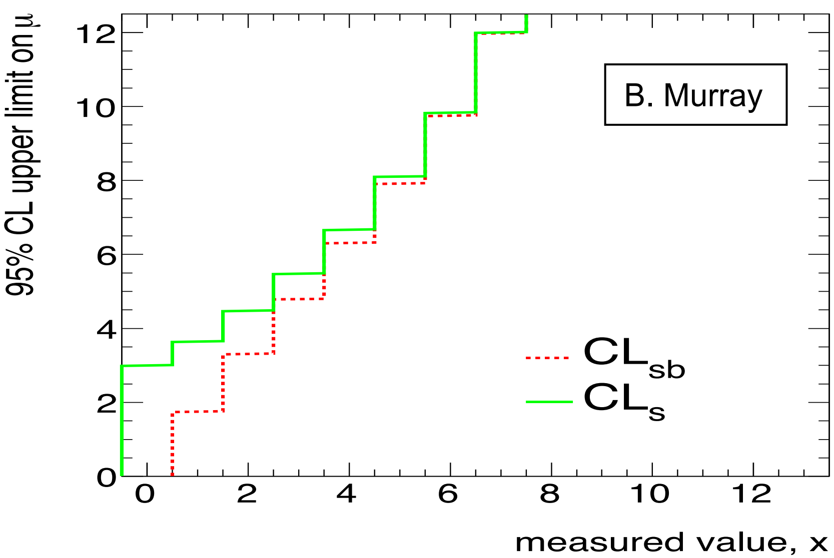


Expected background b = 3
Expected signal yield s = 3

Bottom left: cumulative of Poisson distributions and their ratio CL_S

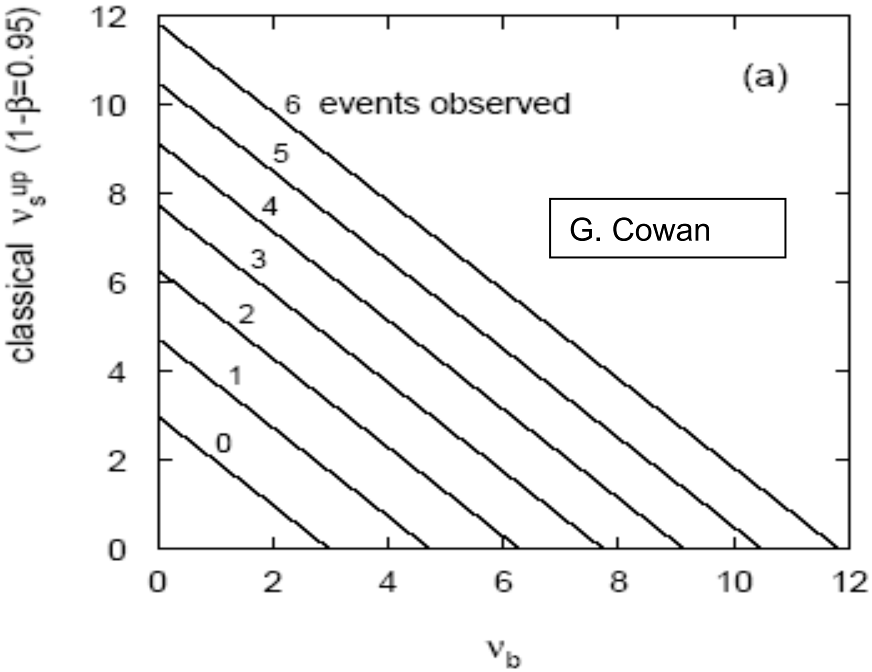
Bottom right: upper limits from classical approach CL_{sb} CL_S technique



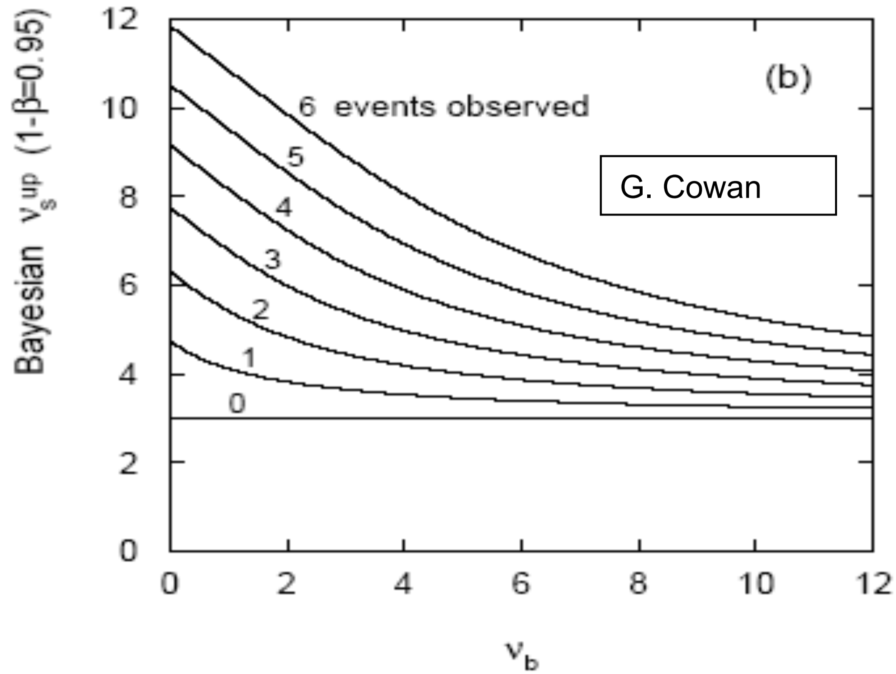


Classical and Bayesian/CL_S Limits at 95% CL

Classical



Bayesian/Zech/CL_S



Upper limit can be „0“

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

Upper limit always ≥ 3 $p = 1 - \alpha\left(1 - F_{\chi^2}[2b, 2(n+1)]\right)$

$$s_{up} = \frac{1}{2} F^{-1}_{\chi^2}[p, 2(n+1)] - b$$

for b= 0 identical

for n>>b also identical

other b values Bayesian> classical limit →“conservative” coverage > CL

independent on b for n= 0 .

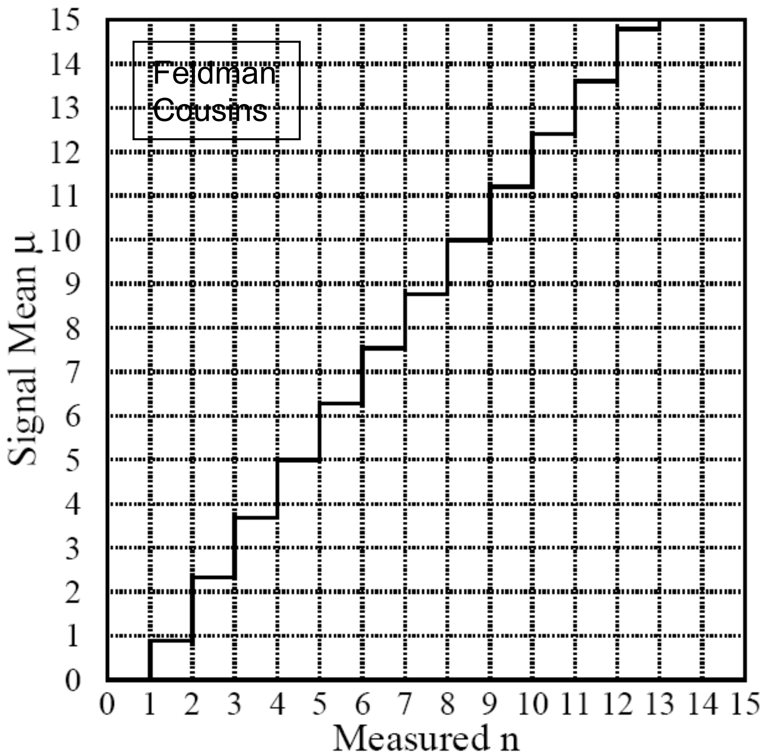
Flip-Flop-Problem for Poisson-Parameter s=μ

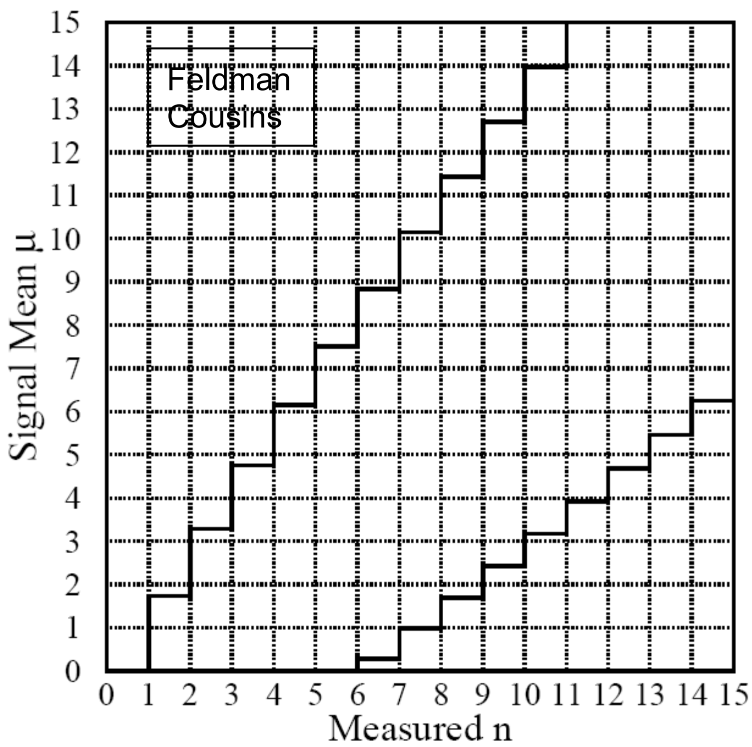
Known background =3

$$P(n|\mu) = (\mu + b)^n \exp(-(\mu + b))/n!$$

One-sided CI at CL=90%

Two-sided CI at CL=90%





For „Flip-Flop“ again to small coverage

Construction of confidence belt via likelihood ratio

$$l(s) = \frac{L(n|s,b)}{L(n|\hat{s},b)} \quad \text{where} \quad \hat{s} = \begin{cases} n - b & n \geq b, \\ 0 & \text{otherwise} \end{cases}$$

„Unified approach“: Poisson-CI at 90% CL

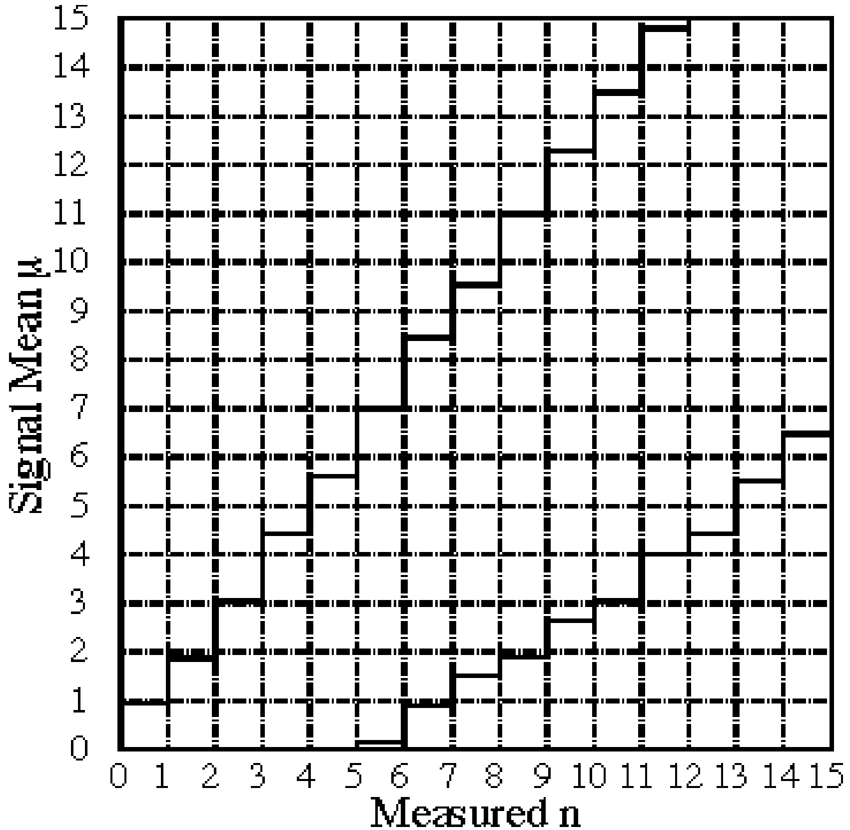
Construction of confidence belt
for μ=0.5, b=3

confidence belt for b=3

$$R = P(n|\mu)/P(n|\mu_{\text{best}})$$

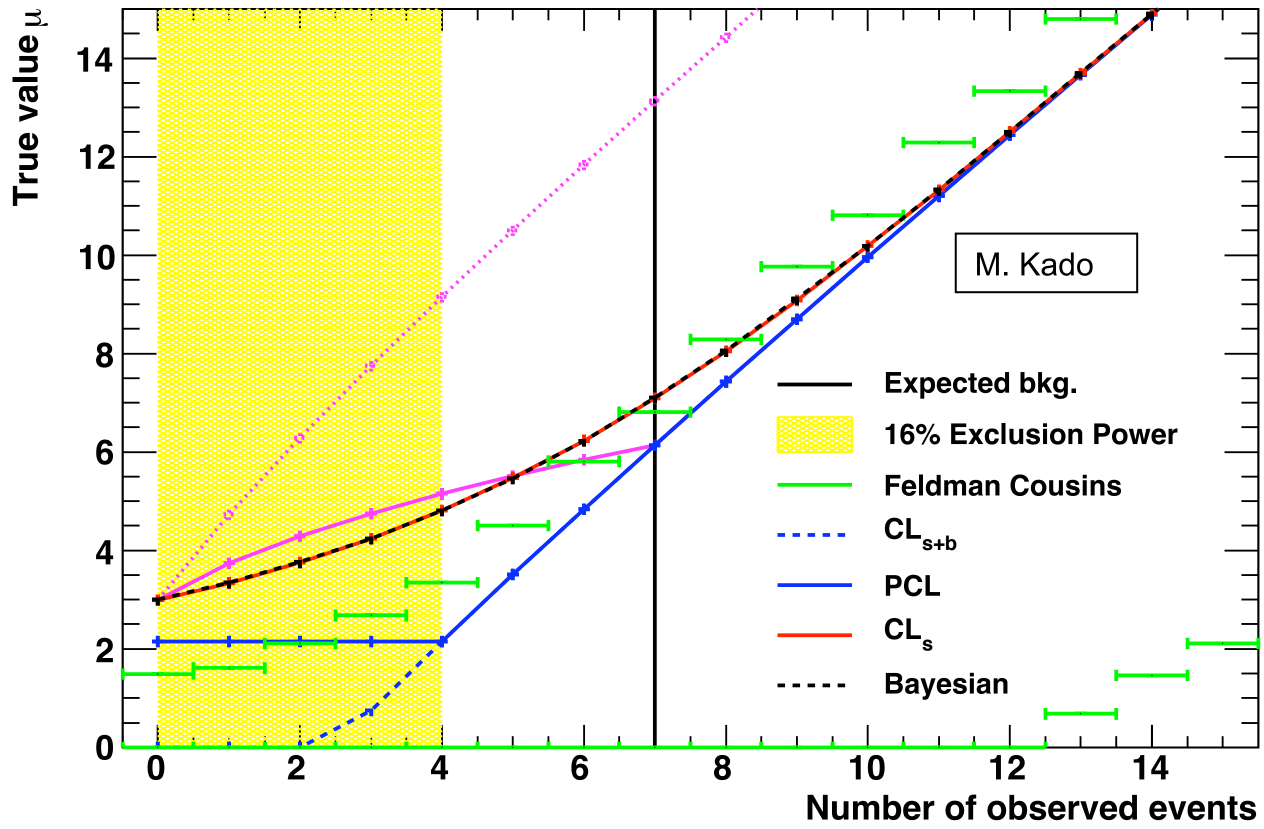
Standard

x	$P(x\|\mu)$	$\hat{\mu}$	$P(x\|\hat{\mu})$	R	rank	U.L.	C.L.
0	0.030	0.0	0.050	0.607	6 •		
1	0.106	0.0	0.149	0.708	5 •	•	•
2	0.185	0.0	0.224	0.826	3 •	•	•
3	0.216	0.0	0.224	0.963	2 •	•	•
4	0.189	1.0	0.195	0.966	1 •	•	•
5	0.132	2.0	0.175	0.753	4 •	•	•
6	0.077	3.0	0.161	0.480	7 •	•	•
7	0.039	4.0	0.149	0.259		•	•
8	0.017	5.0	0.140	0.121		•	



Comparison of Different Limit Derivations at 95% CL

Simple counting experiment with exactly known background expectation of 7 events



- if CL_S<5% we call a μ hypothesis excluded at 95% CL (true coverage larger)
- CL_S and Bayesian limit with flat prior in signal rate mathematically identical in praxis also very similar results for test statistics used at LHC (Tevatron, LEP)
- PCL= power constrained limit: require that power ≥ 16% (cut off at expected -1σ)

Conclusions

probably its already quite late when we arrive here

→ thanks for your attention and please ask questions now or at the bar

Appendix

Literature

Wilks and Wald theorems

S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-2.
A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society, Vol. **54**, No. 3 (Nov., 1943), pp. 426-482.

Profile Likelihood

Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

PCL

Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, "Power-Constrained Limits", `arXiv:1105.3166v1 [physics.data-an]`.

CL_S method

A.L. Read, J. Phys. G **28**, 2693 (2002).
T. Junk, Nucl. Instrum. Methods Phys. Res., Sec. A **434**, 435 (1999).

Zechs Interpretation

G. Zech, Nucl. Instr. and Meth. **A277** (1988) 608

Bayesian Limit for Poisson with flat prior

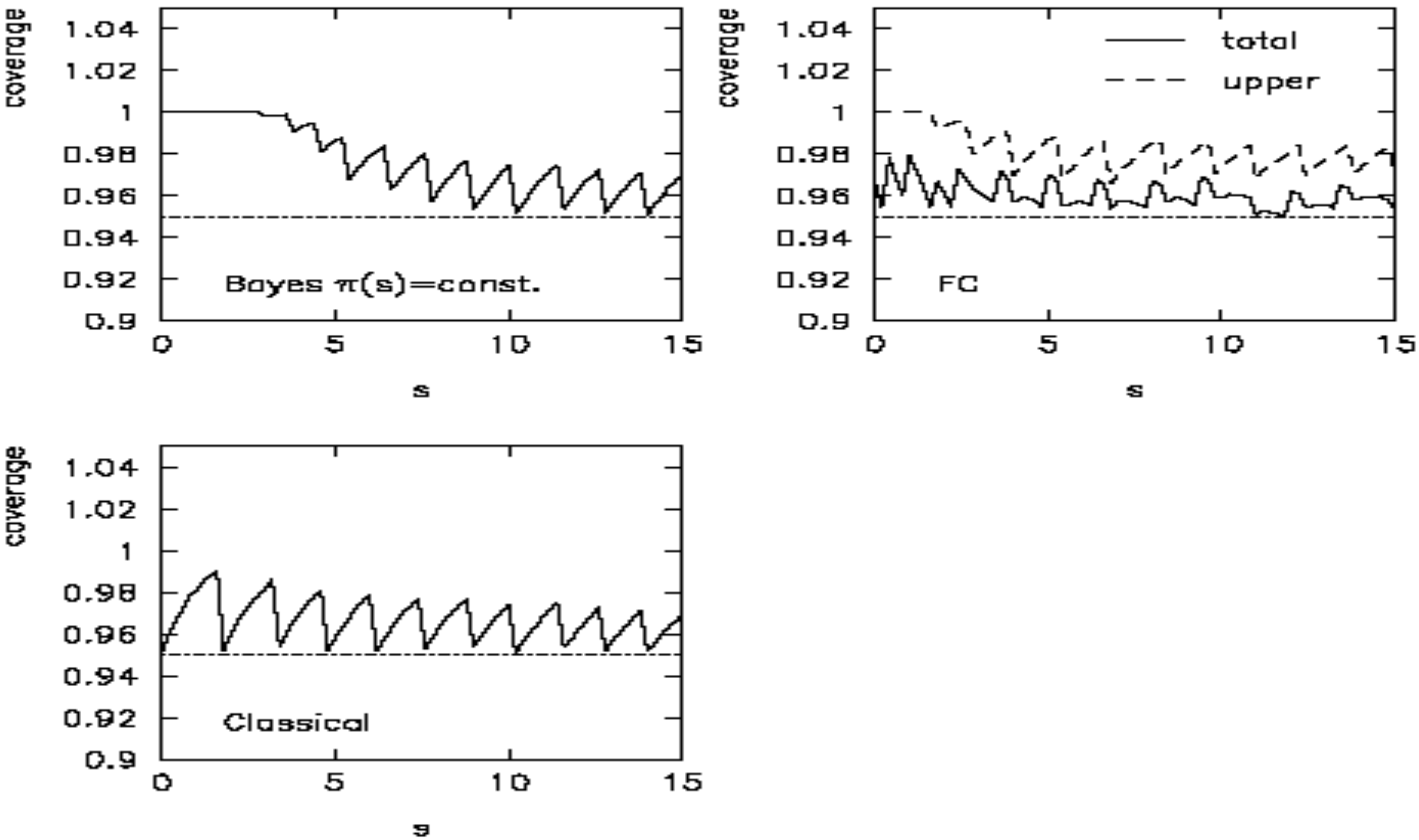
O. Helene, Nucl. Instr. and Meth. 212 (1983) 319

FC/Unified Intervalls

G. J. Feldman and R. D. Cousins, "A Unified Approach to the Classical Statistical Analysis of Small Signals", *Phys. Rev.* **D57** (1998) 3873–3889,

Coverage of CI for Poisson-PDF

Due to discrete nature of Poisson random variable the coverage is per construction larger than quoted CI also for Frequentist methods for most true values



PDF for Exclusion Test Statistic

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)^2\right]$$

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$

$$F(q_\mu|\mu') = \Phi\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)$$

$$p_\mu = 1 - F(q_\mu|\mu) = 1 - \Phi\left(\sqrt{q_\mu}\right)$$