MAGIC-TDAS 06 - 10

060722 / WWittek

22 July 2006

Calculation of Confidence Intervals

Wolfgang Wittek (MPI Munich)

(Presentations made at MPI)

Calculation of Confidence Intervals

- Baye's theorem
- The likelihood principle (LP)
- The likelihood ratio test

- Different kinds of intervals (ordering algorithms)

- Bayesian confidence intervals
- Frequentist (classical) confidence intervals
- Likelihood-ratio (*LR*) intervals

- Treatment of nuisance parameters
- Comparison of different approaches

Baye's theorem

$$P(A \mid B) \cdot P(B) \;=\; P(A \cap B) \;=\; P(B \mid A) \cdot P(A)$$

$\uparrow$ conditional probability for A, given B

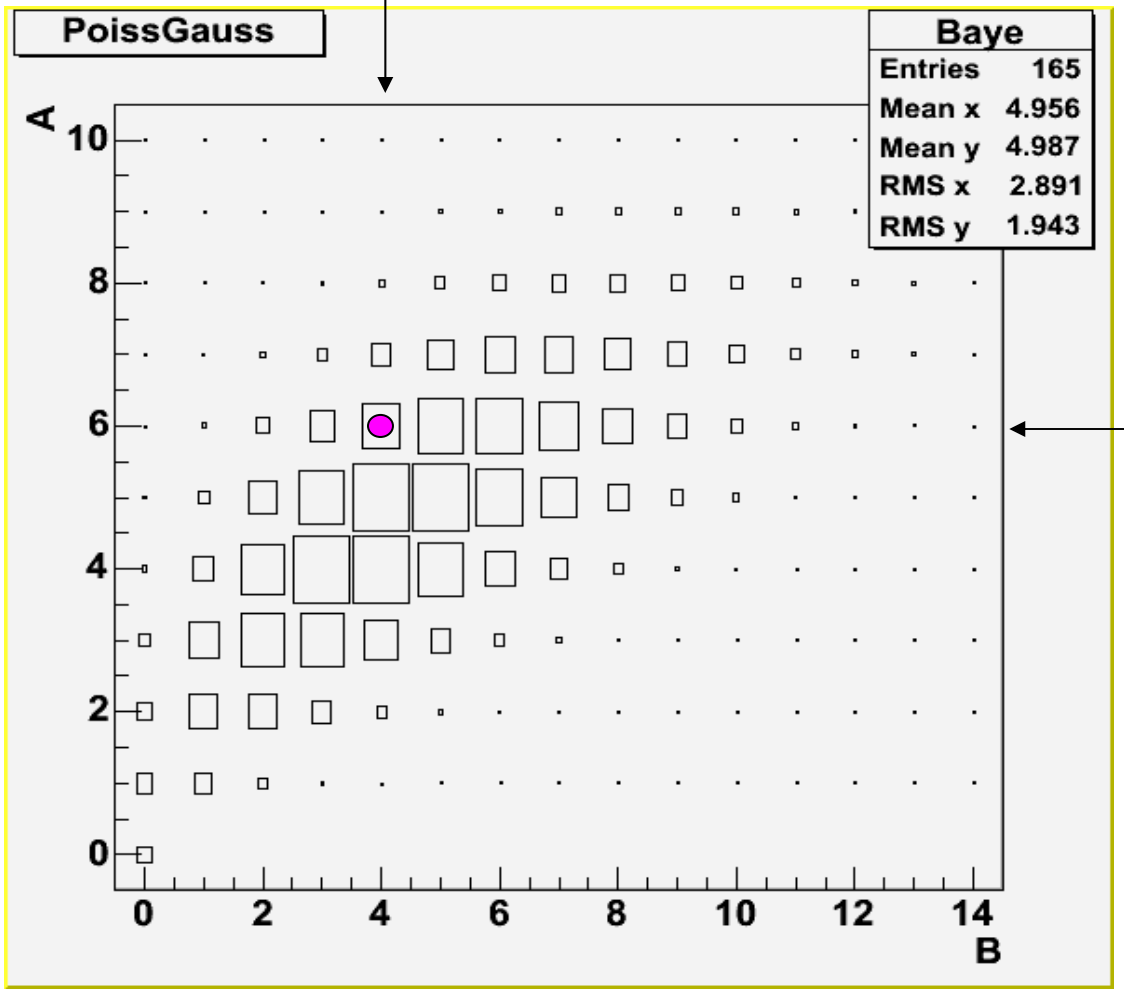
Example : $P(B / A) \;=\; Poiss(B; A) \;=\; \dfrac{A^B}{B!} \cdot \exp(-A)$

$$P(A / B) = \frac{P(B / A) \cdot P(A)}{P(B)} \qquad P(B) = \sum_{\tilde{A}} P(B / \tilde{A}) \cdot P(\tilde{A})$$

$$\text{if } P(A) \;=\; const \quad \Rightarrow \quad P(A / B) \;=\; \frac{Poiss(B; A)}{\sum_{\tilde{A}} Poiss(B; \tilde{A})}$$

Baye's theorem :

$$P(A \mid B) \cdot P(B) \;=\; P(A \cap B) \;=\; P(B \mid A) \cdot P(A)$$



A

2-dim. normalized probability distribution :

Gauss (A; m=5, σ=2) ·

Poisson (B; A)

The likelihood principle (LP)

One assumes that an observation x follows a given probability density $f(x; A)$ with unknown parameter A.

The LP states : *The information contained in an observation* x_0 *with respect to the parameter* A *is summarized by the* ***likelihood function*** $L(A; x_0) = f(x_0; A)$. *All what matters for the parameter inference is* $L(A; x_0)$. *Measurements different from* x_0 *are irrelevant.*

The LP relies on the strict validity of the probability density. Therefore, methods based on the likelihood function often are very sensitive to unknown biases, backgrounds and losses.

The LP is not fully accepted by all statisticians.

Neyman-Pearson lemma

(see K.S. Cranmer, physics/0310108)

Given a certain measurement x, the likelihood ratio

$$LR(x) = L(x; H_0) / L(x; H_1)$$

is the **most powerful** variable or test statistic for **discriminating** between

- a simple null hypothesis (H_0, background only)
- and another hypothesis (H_1, signal plus background)

For a proof see J.Stuart, A.Ord and S.Arnold, "Kendall's Advanced Theory of Statistics", Vol 2A (6th Ed.) (Oxford University Press, New York, 1994)

If W is the acceptance region for H_0 for a confidence level α, the probability for a Type I error is equal to $1-\alpha = 1-\int_W L(x;H_0)\cdot dx$

the probability for a Type II error is $\beta = \int_W L(x;H_1)\cdot dx = \int_W \frac{L(x;H_0)}{LR(x)}\cdot dx$

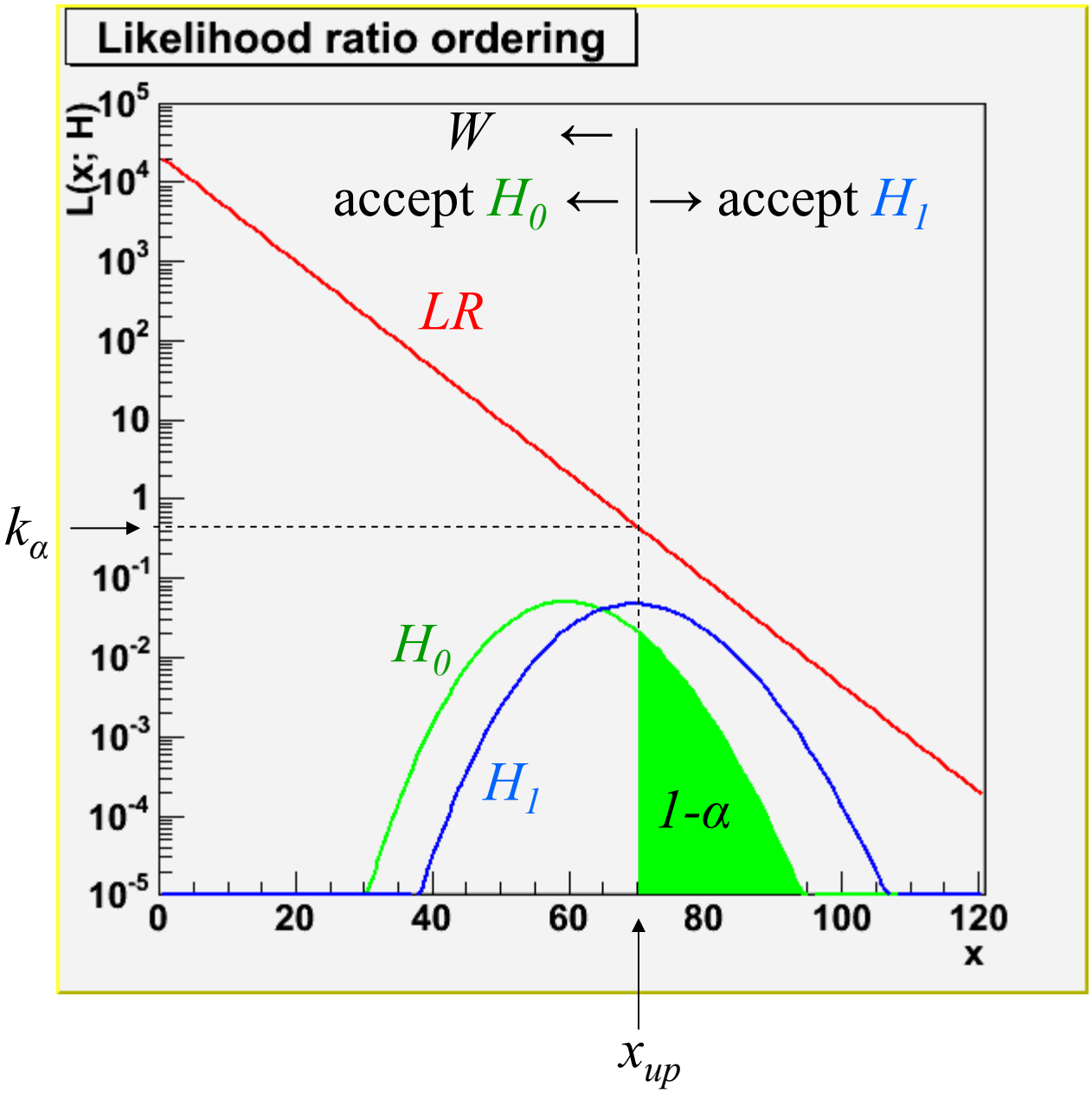
The acceptance region W which minimizes the rate of Type II errors (maximizes the power), mistaking the signal for a background fluctuation, rejecting H_1 although H_1 is true,

is given by : $$W = \left(x \mid LR(x) = \frac{L(x;H_0)}{L(x;H_1)} \geq k_\alpha \right)$$

The constant k_α is determined from $L(x; H_0)$ and the confidence level α :

$$\alpha = \int_W L(x;H_0)\cdot dx = \int_{LR(x)\geq k_\alpha} L(x;H_0)\cdot dx$$

- Note : The LR is used as ordering quantity only. The probability distribution of the LR is not used.



α = confidence level

$LR(x) = L(x;H_0) / L(x;H_1)$

acceptance region for H_0 is defined by

$LR > k_\alpha = LR(x_{up})$ or

$$\alpha = \int_{-\infty}^{x_{up}} L(x;H_0)\cdot dx$$

$$= \int_{LR(x)\geq k_\alpha} L(x;H_0)\cdot dx$$

The likelihood ratio test

Given a certain measurement, calculate the likelihood to obtain this measurement assuming a certain hypothesis.

L_1 maximum likelihood value for the hypothesis H_1

L_0 maximum likelihood value for the **same hypothesis** but with additional constraints (H_0, with k parameters less than H_1)

The likelihood ratio $LR = L_0/L_1$ is then always between 0 and 1.

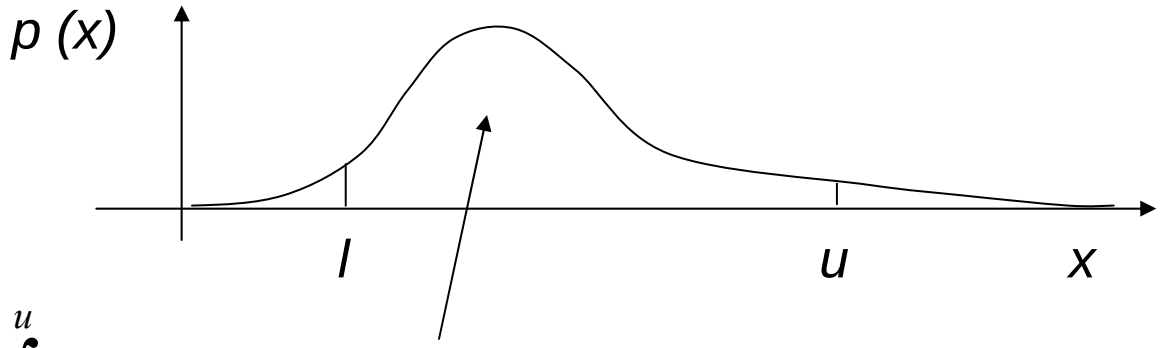
If H_0 is true, the log-likelihood ratio $\Lambda = -2\ ln(LR)$ follows approximately a χ^2 distribution with k degrees of freedom.

The hypothesis H_0 is rejected at the α-confidence level if Λ is larger than the χ^2 value corresponding to the α-quantile of the χ^2 distribution with k degrees of freedom.

Note : the probability distribution of the LR is used to calculate the acceptance region for H_0.

Different kinds of intervals

l = lower edge
u = upper edge of interval



Require $\int_l^u p(x)\,dx = \alpha$ (confidence level)

- central interval : $\int_{-\infty}^{l} p(x)\,dx = \int_u^{\infty} p(x)\,dx = (1-\alpha)/2$
- symmetric interval : $\bar{x} - l = u - \bar{x}$
- highest-probability interval : $p(l) = p(u)$
- selective interval : for example $LR(l) = LR(u)$, LR = likelihood ratio
- minimum-size interval : $|u - l| =$ minimum
- intervals for lower or upper limits : $u = \infty$ or $l = -\infty$

The different intervals are based on different ordering principles,

which define the order in which a region in x is added to the interval (l,u), until $\int_l^u p(x)\,dx = \alpha$.

In the different approaches different probabilities $p(x)$ are used :

- **Bayesian** approach : $p(A) = g(A;a)$ $\int_A g(A;a)\cdot dA = 1$

  to calculate the confidence region of A, given a

- **Frequentist** approach : $p(a) = f(a;A)$ $\int_a f(a;A)\cdot da = 1$

  to calculate the acceptance region of a, given A

where A is the parameter and a the measurement.

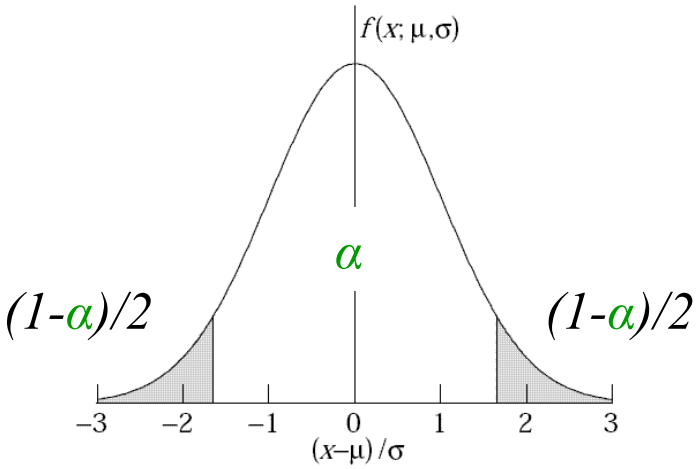
Properties of different intervals

(see G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

- central intervals : - invariant against variable and parameter transformations
  - restricted to the case with 1 variable and 1 parameter
- highest-probability intervals : - not invariant under variable transformations
  - less "biased" than central intervals
- minimum-size intervals : - not invariant under parameter transformations
- symmetric intervals : - not invariant under parameter transformations
- selective intervals : - invariant under transformations of variables and parameters, independent of their dimensions

Understanding of confidence interval (l,u)

If α is the confidence level



(68.27 %, 90 %, 95 %, 95.45 %,
99 %, 99.73 %, 99.9 %, 99.99 %,
99.9937 %, 99.999943 %)

corresponding to

(1 σ, 1.64 σ, 1.96 σ, 2 σ,
2.58 σ, 3 σ, 3.29 σ, 3.89 σ,
4 σ, 5 σ) central intervals

- the true value is contained in the interval (l,u) with a probability α
- if a large number of experiments is performed under identical conditions the true value will be within (l,u) in a fraction α of the experiments

Bayesian approach

The method contains a probability function $\pi(A)$ (prior distribution), reflecting the experimenter's subjective degree of belief about A before the measurement is carried out.

Example : $\pi(A) = const$ for $A_1 \leq A \leq A_2$; $\pi(A) = 0$ otherwise

Problem : $\pi(A)$ depends on the metric of A

should dN/dA be $const$ or dN/dy, where $y = h(A)$

if $y = lnA \quad \rightarrow \quad dN/dA = dN/dy \cdot dy/dA = 1/A$

$\pi(A)$ is appropriate to specify a prior knowledge about A,

it is less suitable to specify ignorance about A

Frequentist (classical) approach

A prior distribution $\pi(A)$ does not appear in this method.

All calculations are based on the probability distribution $f(a; A)$ for a , given the true value A.

Bayesian confidence intervals

$f(a; A)$ probability to measure a, if the unknown parameter has the value A

Example : $f(a;A) = Poisson(a;A) = \frac{A^a}{a!} \cdot \exp(-A)$

a = measured number of signal events

A = true average value of number of signal events

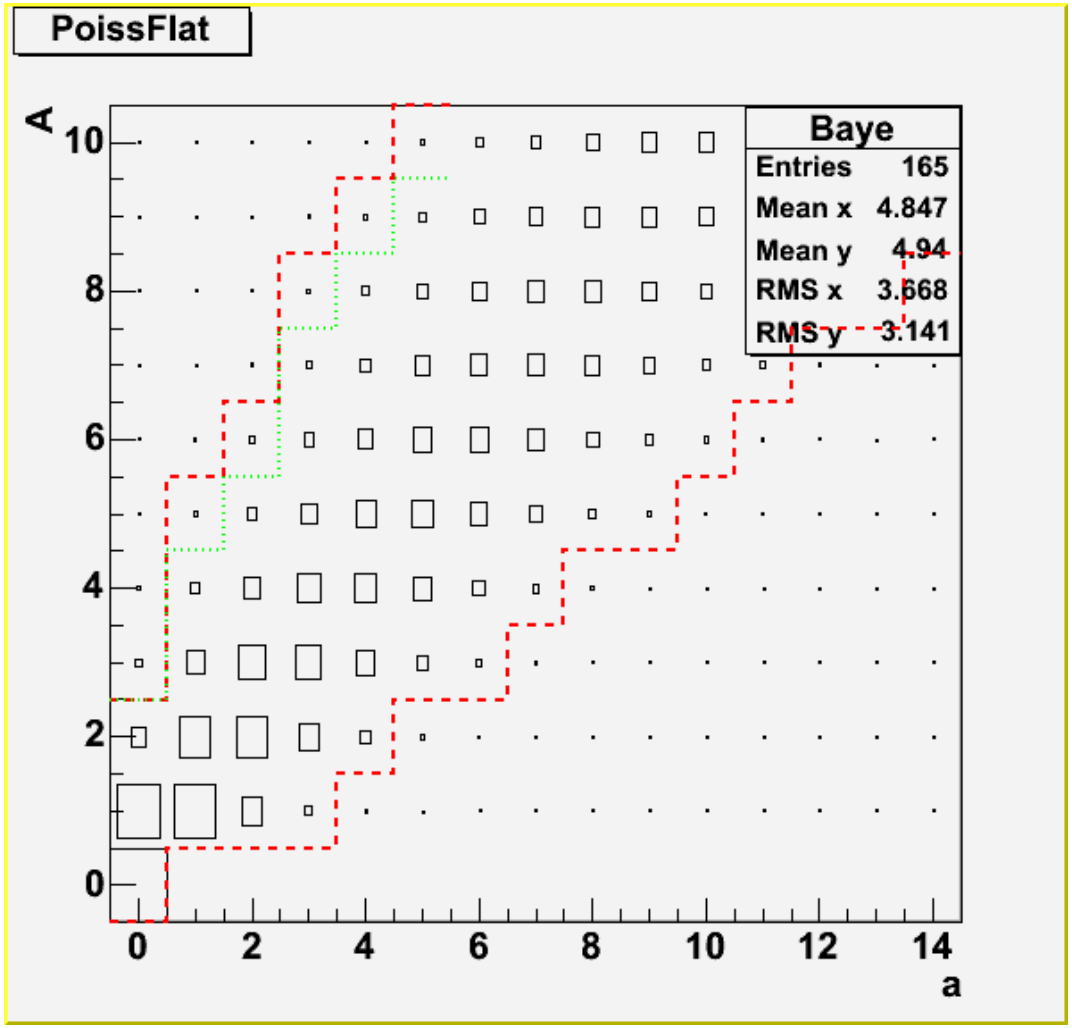
Baye's theorem : $g(A;a) = \frac{f(a;A) \cdot \pi(A)}{\sum_{\tilde{A}} f(a;\tilde{A}) \cdot \pi(\tilde{A})}$ (posterior prob. distr.)

assume $\pi(A) = const \rightarrow g(A;a) = \frac{Poisson(a;A)}{\sum_{\tilde{A}} Poisson(a;\tilde{A})}$

distribution of a, given A :

$f(a; A) = Poisson(a; A)$

$\pi(A) = const$



distribution of A, given a :

$$g(A;a) = \frac{Poisson(a;A)}{\sum_{\tilde{A}} Poisson(a;\tilde{A})}$$

given the measurement a, determine the Bayesian confidence interval $(l(a),u(a))$ for A by the condition

$$\int_{l(a)}^{u(a)} g(A;a)\, dA = \alpha$$

assuming a certain ordering principle (central interval, upper limit, …)

The Bayesian limits ($l(a),u(a)$) for A depend on

- the measurement a
- the ordering principle

  (order in which points are added to the confidence interval)
- the confidence level α
- the prior distribution $\pi(A)$

The limits ($l(a),u(a)$) define the "confidence belt" in the A-a plane.

Measurements different from a don't appear in the calculation of ($l(a),u(a)$).

Bayesian coverage :

One performs a large number of experiments, with a distribution of A values according to $\pi(A)$. The experiments will have A values within ($l(a),u(a)$) in a fraction α of all cases.

Bayesian intervals have Bayesian coverage by construction.

Frequentist confidence intervals

Neyman construction of confidence intervals (see J. Neyman, Phil. Trans. Royal Soc. London, Series A, 236 (1937) 333)

$f(a; A)$ probability for measuring a if the unknown parameter has the value A

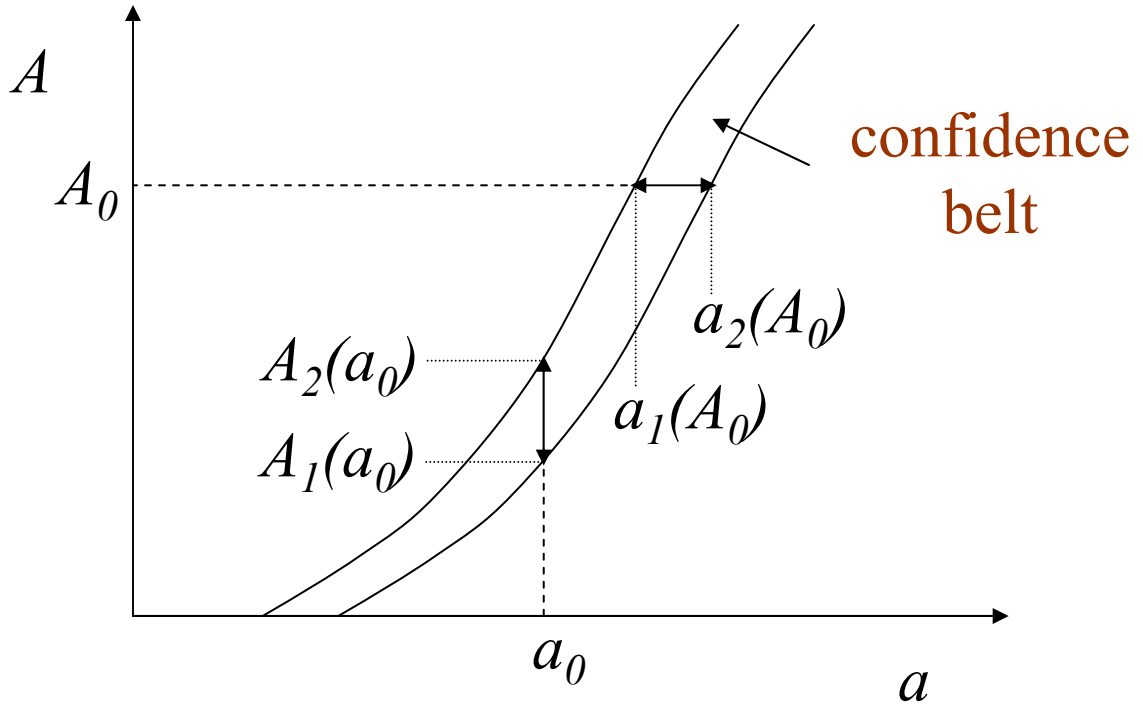
For each value A define an acceptance interval $R_a(A) = (a_1(A), a_2(A))$

of a by the condition $\int_{a_1(A)}^{a_2(A)} f(a; A) \cdot da = \alpha$ (α = confidence level)

The region between the lines $a_1(A)$ and $a_2(A)$, which may also be denoted by $A_1(a)$ and $A_2(a)$, is called "confidence belt".

For a given measurement a define the confidence interval $R_A(a)$ of A as follows: include in $R_A(a)$ all those A whose acceptance interval $R_a(A)$ contains a $\Longrightarrow$ $R_A(a) = (A_1(a), A_2(a))$.

Confidence belt



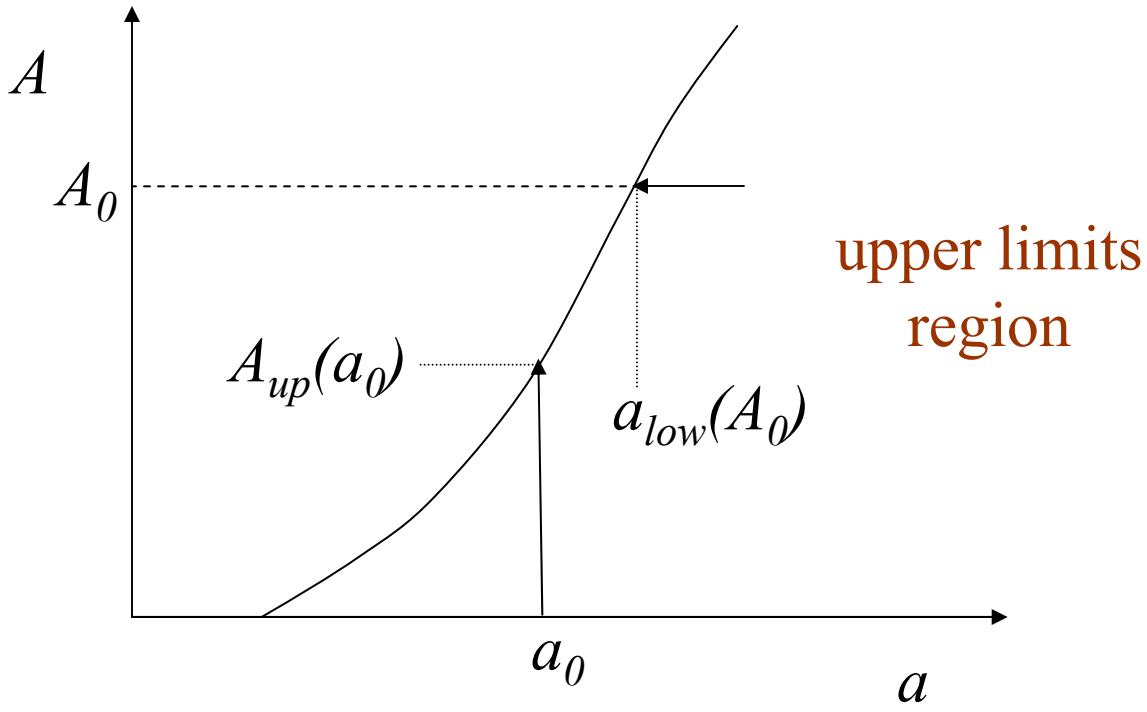
($a_1(A_0)$, $a_2(A_0)$) acceptance interval of a, for $A = A_0$

($A_1(a_0)$, $A_2(a_0)$) confidence interval of A, if measurement is a_0

For a given measurement a_0, A_0 lies in confidence interval $(A_1(a_0), A_2(a_0))$ of A if and only if a_0 lies in acceptance region $(a_1(A_0), a_2(A_0))$ of a .

→ $P(A_1(a_0) < A < A_2(a_0)) = P(a_1(A_0) < a < a_2(A_0)) = \alpha$

Upper limits region



($a > a_{low}(A_0)$) acceptance interval of a, for $A = A_0$

($A < A_{up}(a_0)$) confidence interval of A, if measurement is a_0

For a given measurement a_0, A_0 lies in confidence interval of A

($A < A_{up}(a_0)$) if and only if a_0 lies in acceptance region of a ($a > a_{low}(A_0)$)

→ $P(A_0 < A_{up}(a_0)) = P(a_0 > a_{low}(A_0)) = \alpha$

Different ordering principles in different Frequentist approaches :

- classical : central, symmetric or highest probability intervals
- unified classical : selective intervals, like highest *LR*

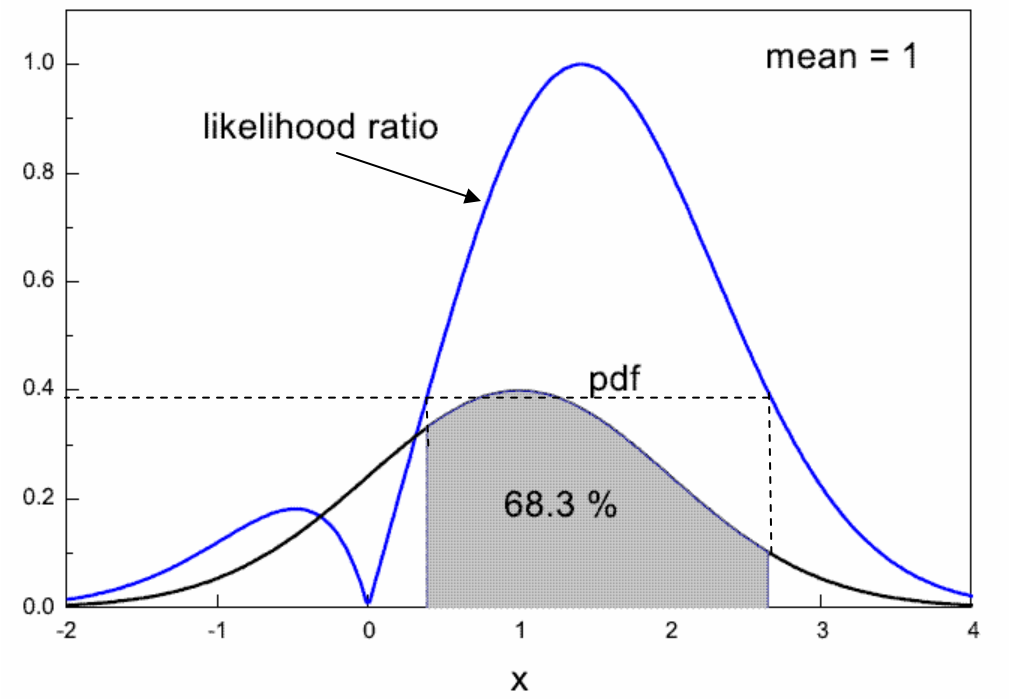


Fig. 5. Likelihood ratio ordering. The likelihood ratios are equal at the limits of the shaded region.

(G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

Advantages of a highest likelihood ratio (*LR*) interval as compared to a central interval :

- it provides a smooth transition from an α-confidence interval to an α-confidence upper limit (??????)
- it often avoids unphysical or empty intervals
- it is also invariant against transformations of variables and parameters (independent of the dimensions)

Flip-flopping "problem"

"If the result x is less than 3, I will state an upper limit. Otherwise I will state a central confidence interval. This policy leads to undercoverage. The experimenter should decide before looking at the data."

F&C, PR D57 (1998) 3873

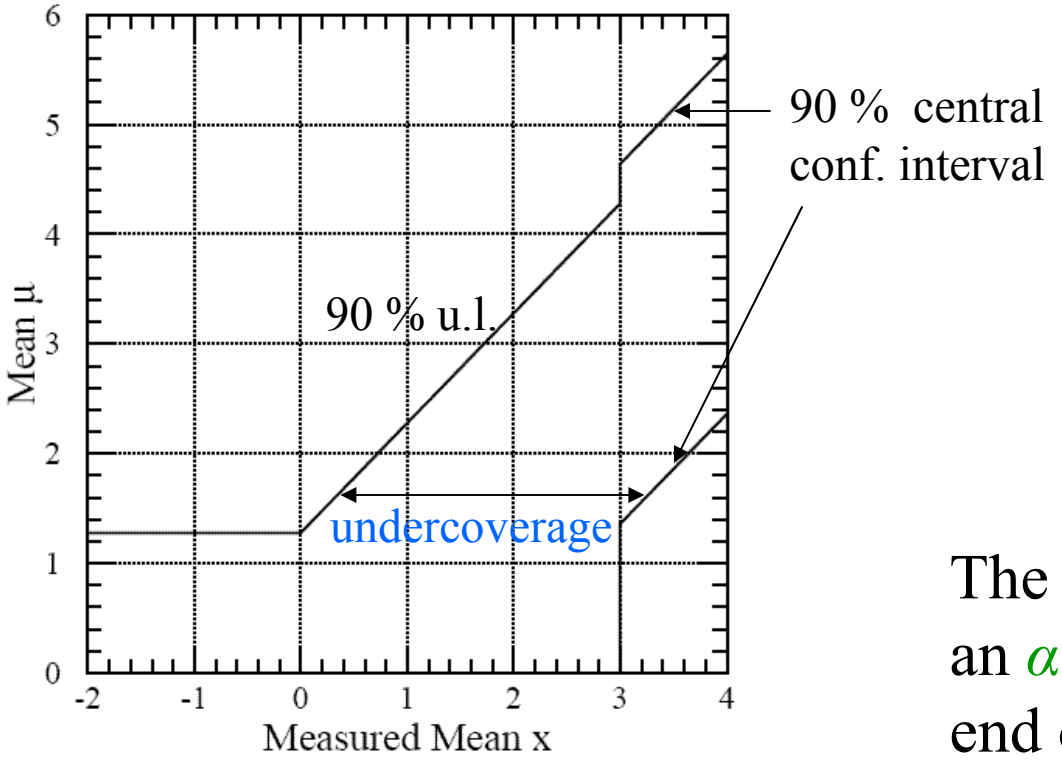
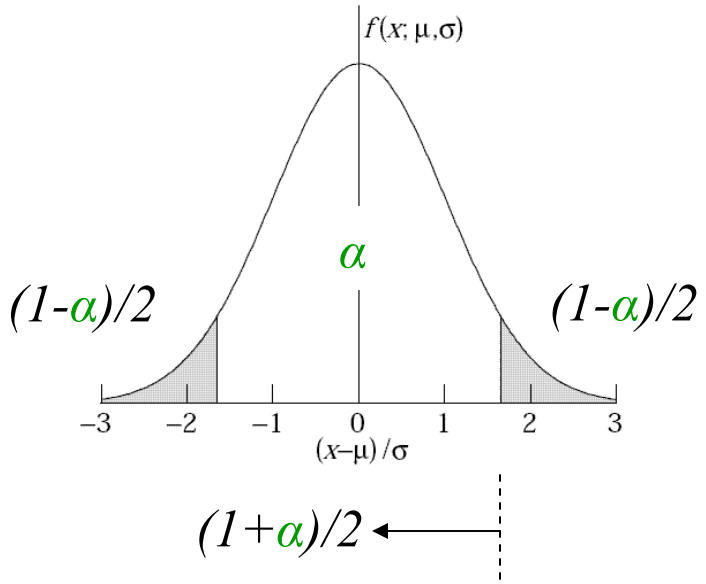


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence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



The "problem" is due to identifying an α-c.l. upper limit with the upper end of an α-c.l. central conf. interval. This is wrong. There is no problem.

More general definition of acceptance interval

- define the acceptance interval $R_a(A)$ of a by a test (χ^2 test, likelihood test): the hypothesis A is accepted if a is within $R_a(A)$
- then proceed as before: given a measurement a, accept those A in the confidence interval $R_A(a)$ of A, for which the measurememt a is contained in the acceptance interval $R_a(A)$

Newman construction in more than one dimension

Assume the measurements (x,y) to be distributed according to $f(x,y;A,B)$, with parameters A and B, whose values are unknown

- For each pair (A,B) determine the acceptance region $R_{xy}(A,B)$ of (x,y), using some ordering algorithm
- Given a measurement (x,y), the confidence region $R_{AB}(x,y)$ for (A,B) is found as follows : include in $R_{AB}(x,y)$ all those pairs (A,B) for which (x,y) is contained in $R_{xy}(A,B)$

W.A. Rolke and A.M. Lopez, NIM A458 (2001) 745

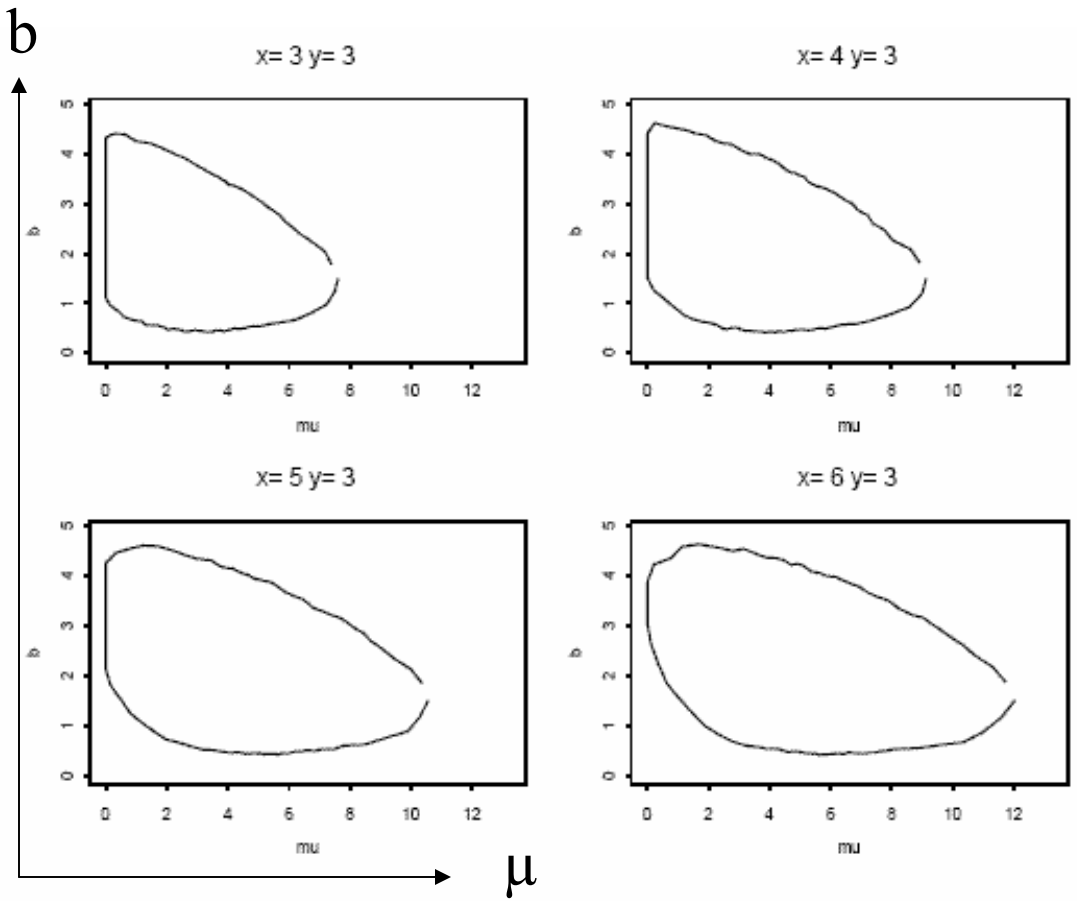


Fig. 1. Two dimensional confidence regions for $x = 3, 4, 5$ and 6 signal events, and $y = 3$ background events. The background region is twice the size of the signal region, and a 90% confidence level was used.

Confidence regions for (μ,b)

for different measurements (x,y),

with $x \approx \text{Poisson}(x; \mu+b)$

$y \approx \text{Poisson}(y; \tau b)$,

$\alpha = 90\ \%, \quad \tau = 2$

Frequentist coverage

If a large number of experiments is performed under identical conditions ($A=A_0$, $B=B_0$)

a confidence region $R_{AB}(x,y)$ of (A,B) is said to have exact coverage if the pair of values (A_0,B_0) is contained in $R_{AB}(x,y)$ in a fraction α of all experiments.

Frequentist confidence regions have **Frequentist coverage** by definition.

Likelihood ratio intervals

$f(a; A)$ probability to measure a, if the unknown parameter has the value A

Define the likelihood function $L(A; a)$ as $L(A; a) = f(a; A)$

For a given measurement a, find $A_{\max}$ which maximizes $L(A; a)$

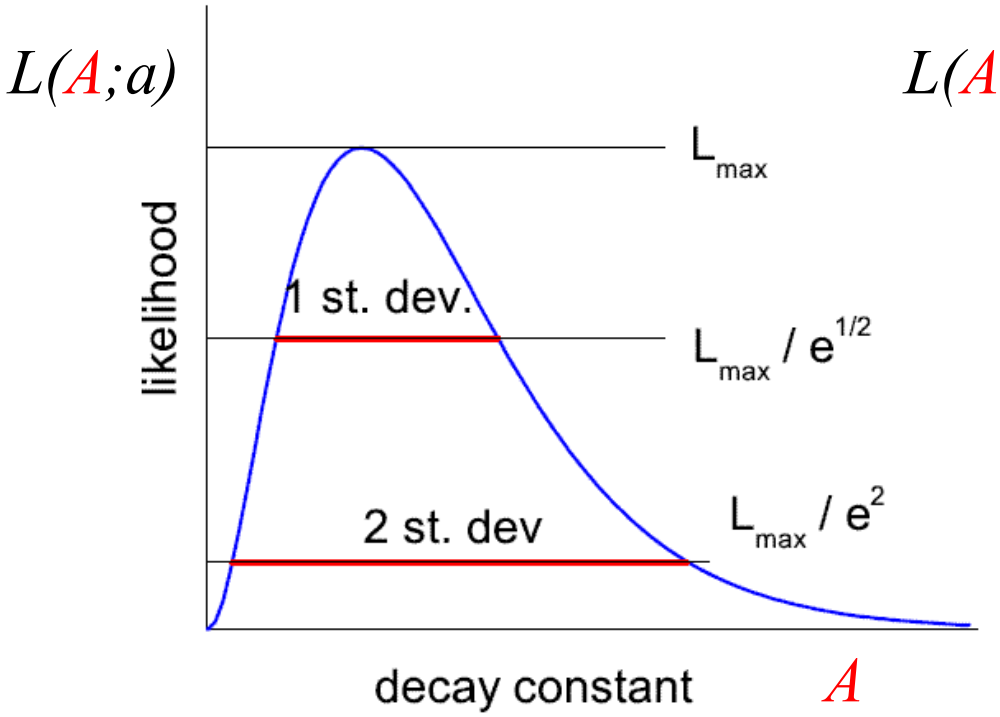
Define the *LR* interval (A_{low}, A_{up}) by the requirement

$$\frac{L(A_{low})}{L(A_{\max})} = \exp(-\Delta) = \frac{L(A_{up})}{L(A_{\max})} \qquad \text{or}$$

$$\ln L(A_{\max}) - \ln L(A_{low}) = \Delta = \ln L(A_{\max}) - \ln L(A_{up})$$

where $\Delta = n^2/2$ for a confidence level of $n \cdot \sigma$

Likelihood ratio intervals



Bayesian interval

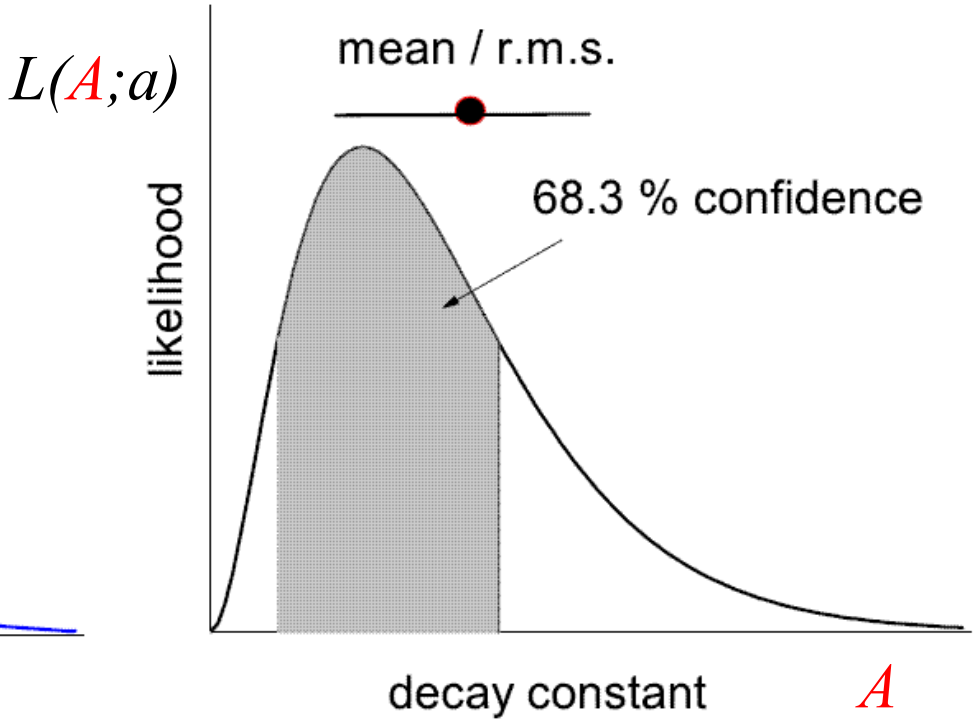


Fig. 18. Likelihood ratio limits (left) and Bayesian limits (right)
(G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

If $L(A; a) = Gauss(a; A,\sigma)$, the *LR* intervals are identical to the Bayesian confidence intervals, obtained with a uniform prior

Example for the construction of a Likelihood ratio interval :

a number of signal events when true average is A (unknown)

b number of background events when true average is B (known)

One measures $x = a+b$. Assume a and b to be distributed like

Poisson $(a; A)$ and *Poisson* $(b; B)$ respectively

The likelihood function is

$$L(A;\ x)\ =\ Poisson(x;\ A+B)\ =\ \frac{(A+B)^x}{x!}\cdot\exp(-A-B)$$

get *LR interval* $(A_{low},\ A_{up})$ for A by determining the A at which

$L(A; x) = maximum\ (= L_{max})$ and $L(A_{low}; x) / L_{max} = exp(-1/2)$

$L(A_{up}; x) / L_{max} = exp(-1/2)$

For $x = 0$ $\rightarrow\ L(A;\ x=0)\ \approx\ \exp(-A)$, independent of B

Comparison of Bayesian, Frequentist and LR intervals

R.D. Cousins (PHYSTAT05)

Table 1. 68% C.L. intervals for the mean μ of a Poisson distribution, based on the single observation $n_0 = 3$, calculated by various methods. Only the frequentist intervals avoid under-coverage for all values of μ. The boldface numbers highlight the fact that the frequentist central interval shares the right endpoint with the Bayesian interval with uniform prior, and the left endpoint with the Bayesian interval with $1/\mu$ prior, explaining why neither set of Bayesian intervals covers for all values of μ.

Method	Prior	Interval	Length
rms deviation	–	(1.27, 4.73)	3.46
Bayesian central	1	(2.09, **5.92**)	3.83
Bayesian shortest	1	(1.55, 5.15)	3.60
Bayesian central	$1/\mu$	(**1.37**, 4.64)	3.27
Bayesian shortest	$1/\mu$	(0.86, 3.85)	2.99
Likelihood ratio	–	(1.58, 5.08)	3.50
Frequentist central	–	(**1.37**, **5.92**)	4.55
Frequentist shortest	–	(1.29, 5.25)	3.96
Frequentist LR ordering	–	(1.10, 5.30)	4.20

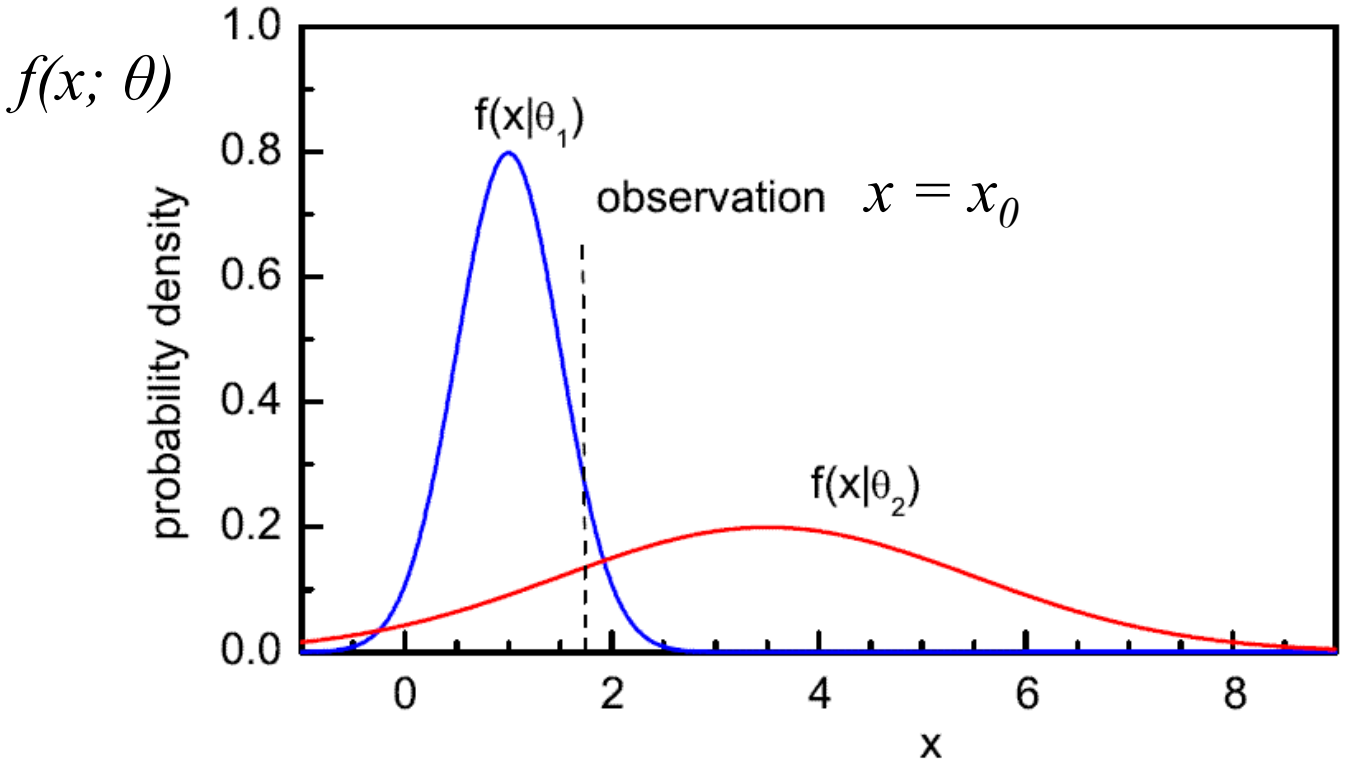
68 % confidence intervals

for the mean *A*

f(x; A) = Poisson(x; A)

with the measurement $x = 3$

Comparison of Frequentist and *LR* approach



Frequentist approach :

considers $f(x; \theta_1)$ and $f(x; \theta_2)$ over a wide range of x ; x_0 is within acceptance region of θ_2

LR approach :

considers $f(x; \theta_1)/f(x; \theta_2)$ at $x = x_0$ only ;

θ_1 is accepted

Fig. 1. The likelihood is larger for parameter θ_1, but the observation is less then 1 st. dev. off θ_2. Classical approaches include θ_2 and exclude θ_1 within a 68.3% confidence interval

(G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

An interesting special example

- Fewer events (n) in signal region than expected from background

  Helene : for $n = 0$, upper limit of $A = 2.3$ (independent of B)

  F&C : for $n = 0$, "upper limit" of A decreases as B increases

  for any n, "upper limit" tends to 1 as B goes to ∞

Suggestion by F&C :

if "upper limit" is less than "sensitivity" give both.

sensitivity(B) = average "upper limit" in an ensemble of experiments with $A = 0$;

This is an average over all possible measurements with $A=0$ and B.

O.Helene,
NIM 212 (1983) 319

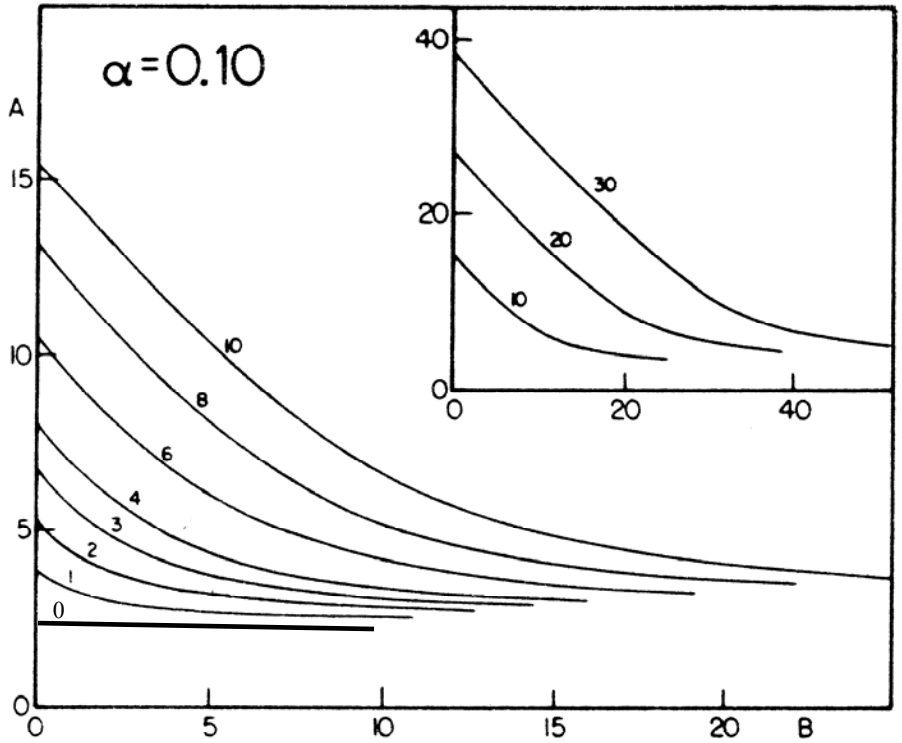


Fig. 2. The same as figure 1, here to 90% ($\alpha = 0.10$) confidence level.

90 % c.l. upper limit

G.J.Feldman and R.D.Cousins,
PR D57 (1998) 3873

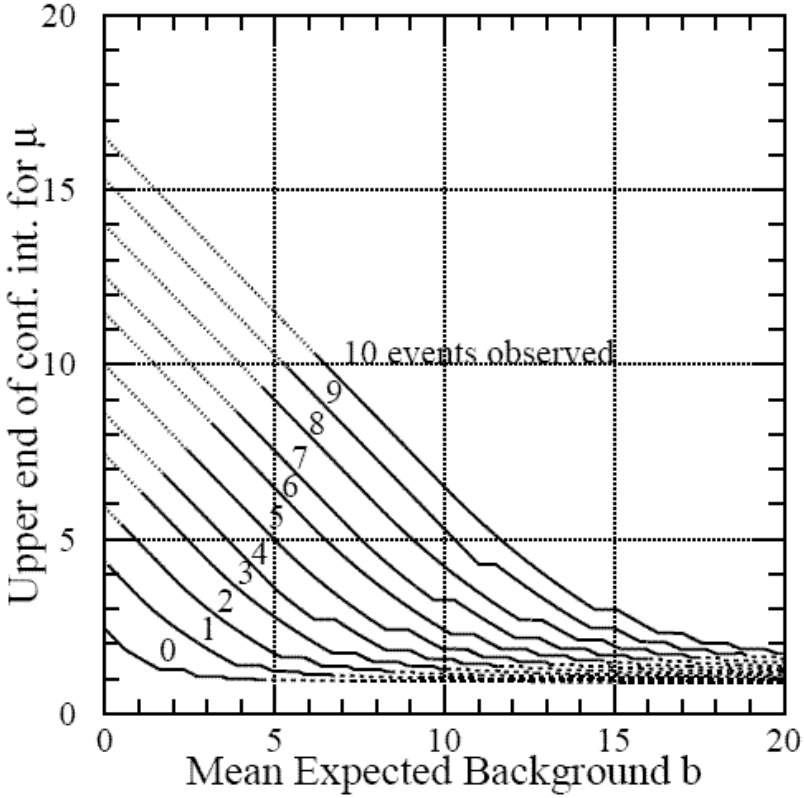
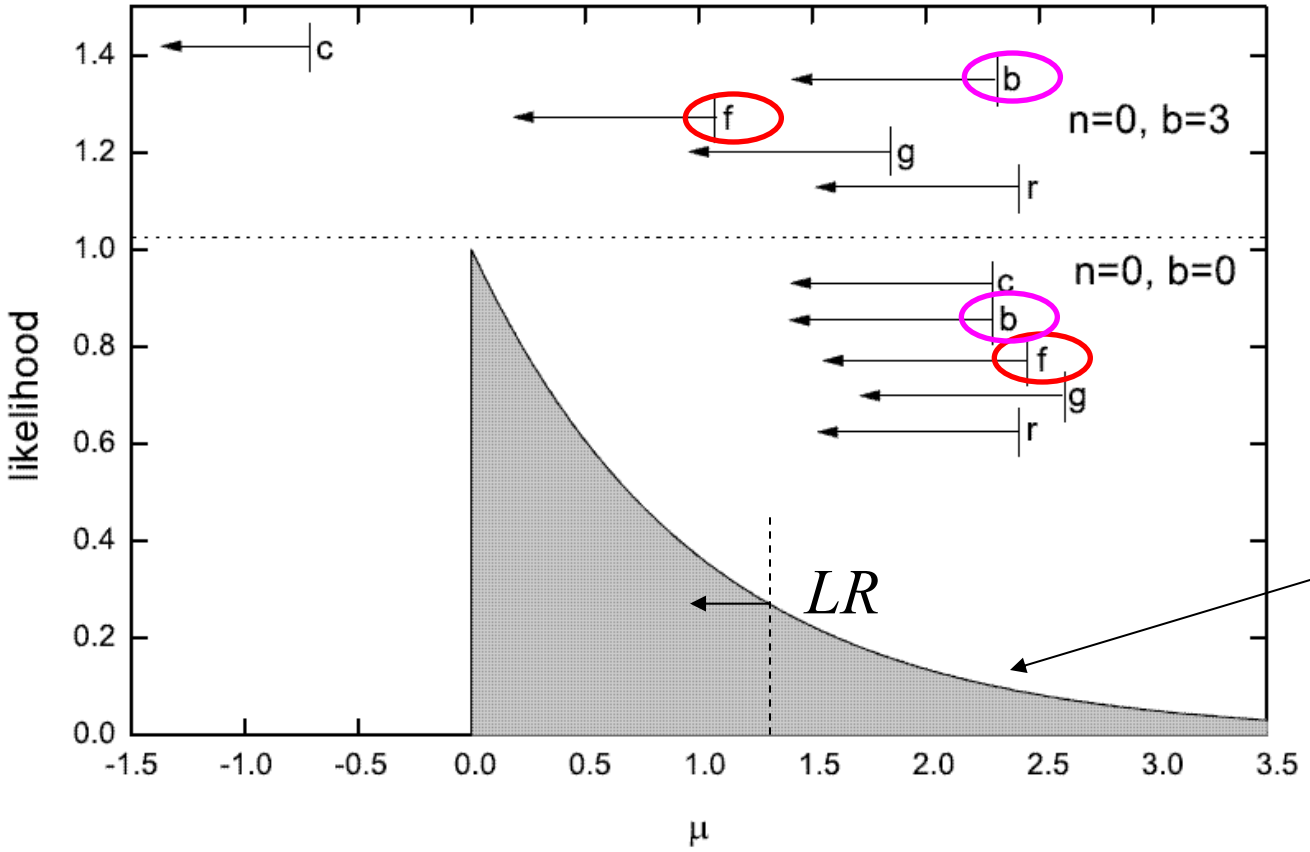


FIG. 8. Upper end μ_2 of our 90% C.L. confidence intervals $[\mu_1, \mu_2]$, for unknown Poisson signal mean μ in the presence of expected Poisson background with known mean b. The curves for the cases n_0 from 0 through 10 are plotted. Dotted portions on the upper left indicate regions where μ_1 is non-zero (and shown in the following figure). Dashed portions in the lower right indicate regions where the probability of obtaining the number of events observed or fewer is less than 1%, even if $\mu = 0$.

upper end of 90 % conf. interval

$n \approx Poisson(n; \mu+b)$ measure $n=0$



90 % upper limits

c: classical

f : unified classical

b : Bayesian

$L(\mu;n=0) \approx exp(-\mu)$

Fig. 16. Likelihood function for zero observed events and 90% confidence upper limits with and without background expectation. The labels refer to [9] (f), Bayesian (b), [41] (g) and [42] (r)

(G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

End of part 1

Nuisance parameters

- parameters whose values have to be known for calculating a result but which carry no information about the result and
- whose uncertainties affect the uncertainty of the final result

Examples of **nuisance parameters** **:**

- background under the signal
- efficiency for measuring the signal (acceptance, A_{eff})

Given the measurements x, y and z of

- the no. of events in the signal region (x)
- the no. of events in the background region (y)
- the no. of signal events surviving the analysis cuts (MC) (z)

What are the limits (l,u) for the true average number A of signal events ?

Treatment of nuisance parameters in the Bayesian approach

1) ***B*** is exactly known (see O.Helene, NIM 212 (1983) 319) :

a number of signal events when true average is A (**unknown**)

b number of background events when true average is B (**known**)

One measures $x = a+b$. Assume x to be distributed like

$f(x; A,B) = Poisson\ (x; A+B)$, where A is unknown and B *is* known

$$g_0(A;\ x, B) \quad = \quad \frac{f(x; A, B)\cdot \pi(A)}{\int_{\tilde{A}} f(x; \tilde{A}, B)\cdot \pi(\tilde{A})\cdot d\tilde{A}}$$

$$= \quad \frac{Poisson\ (x;\ A+B)}{\int_{\tilde{A}} Poisson\ (x;\ \tilde{A}+B)\cdot d\tilde{A}} \qquad (\text{for } \pi\,(A) = const)$$

Given the measurement x and the known true number of background events B, determine the Bayesian confidence interval $(l(x,B),\ u(x,B))$

for A by the condition $\int_{l(x,B)}^{u(x,B)} g_0(A; x, B)\, dA = \alpha$

assuming a certain ordering principle (central interval, upper limit, …)

As long as the nuisance parameter is exactly known (error = 0)

no special procedure is needed for its treatment.

2) B is not exactly known (see O.Helene, NIM 212 (1983) 319) :

a number of signal events when the true average is A (**unknown**)

b number of background events in the signal region when the true average is B (**unknown**)

y number of background events in the background region when the true average is $\tau \cdot B$; τ = (size of bg region) / (size of signal region)

One measures $x = a+b$ and y. Assume x and y to be distributed

like $f(x, y;\ A, B) = Poisson(x;\ A+B) \cdot Poisson(y;\ \tau \cdot B)$

where both A and B are unknown

$$g(A;\ x, y) = \frac{\int_B f(x, y; A, B) \cdot \pi(A) \cdot \pi(B) \cdot dB}{\iint_{\tilde{A}\ B} f(x, y; \tilde{A}, B) \cdot \pi(\tilde{A}) \cdot \pi(B) \cdot dB \cdot d\tilde{A}}$$

with the prior distributions $\pi(A)$ and $\pi(B)$ (Baye's theorem)

Given the measurements x and y, determine the Bayesian confidence interval $(l(x,y),\ u(x,y))$ for A by the condition

$$\int_{l(x,y)}^{u(x,y)} g\ (A; x, y)\ dA \quad = \quad \alpha$$

assuming a certain ordering principle (central interval, upper limit, …)

Treatment of nuisance parameters in the Frequentist approach

- The average value B of the background is **exactly known**
- The average value B of the background is **not exactly known**

The average value B of the background is exactly known

(see G.J. Feldman and R.D. Cousins, PR D57 (1998) 3873)

a number of signal events when the true average is A (**unknown**)

b number of background events when the true average is B (**known**)

One measures $x = a+b$. Assume x to be distributed like

$f(x; A,B) = Poisson\ (x; A+B)$, where A is unknown and B is known

For each value A define an acceptance interval $R_x(A) = (x_1(A), x_2(A))$ of x by the condition $\int_{x_1(A)}^{x_2(A)} f(x; A, B) \cdot dx = \alpha$ (α = confidence level)

using as ordering quantity the likelihood ratio $LR = f(x; A,B) / f(x; A_{best},B)$,

where A_{best} is that (physically allowed) value of A which maximizes $f(x;A,B)$

Determine the confidence interval $R_A(x)$ of A in the usual way:

$R_A(x)$ contains all those A whose acceptance region $R_x(A)$ contains the measurement x

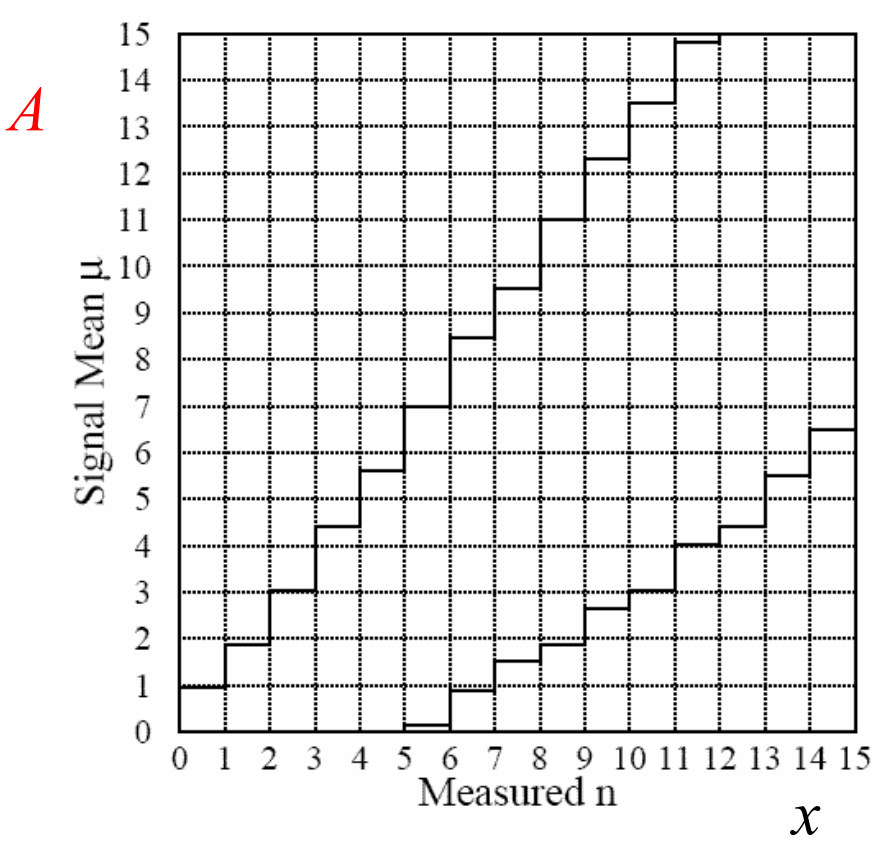


FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$.

$B = 3$, $\alpha = 90\ \%$

the ordering quantity is the likelihood ratio

$$LR = f(x; A,B) / f(x; A_{best},B),$$

As long as the nuisance parameter is exactly known (error = 0)
no special procedure is needed for its treatment.

G.J.Feldman and R.D. Cousins,
PR D57 (1998) 3873

Critical remarks to the paper of F & C

- The flip-flopping problem : "In the classical approach you may decide to quote an upper limit or a confidence interval, after having had a look at the data. This policy leads to undercoverage, in general."

  This statement is wrong : coverage has nothing to do with the above decision. The experimenter can quote an upper limit or a confidence interval, or both, without violating coverage.

  In F&C's paper the apparent flip-flopping problem arises from the fact that they consider as altenatives the upper end of a 90% c.l. central interval with a 90% c.l. upper limit. The problem would not arise if they considered as alternatives the upper end of a **90% c.l. central interval** with a **95% c.l. upper limit**.

- Unified approach : “Using the LR as ordering quantity one obtains intervals which automatically change from two-sided intervals to upper limits. This eliminates undercoverage caused by basing this choice on the data (flip-flopping).”

  It is true that the upper end of any confidence interval can be understood as an upper limit, however, at **different confidence levels**, in general. Even in the unified approach the acceptance intervals are two-sided, implying that the upper end of the confidence interval is an upper limit at a different confidence level than the confidence interval. Even if the confidence interval is one-sided (because its lower end coincides with the lowest physically allowed value, for example) the upper end doesn’t become an upper limit at the same c.l. as the confidence interval.

The average value *B* of the background is

not exactly known

I. Neyman construction in more dimensions + projection

II. *Profile LR* used as ordering quantity (Cranmer, Punzi);
Neyman construction in 2 dimensions (physics p. and nuisance p.)

III. *Profile LR* used to construct confidence interval (Rolke & Lopez);
Neyman construction in 1 dimension (physics parameter only)

IV. Mixed Frequentist-Bayesian approach (Cousins & Highland)

I. Neyman construction in more dimensions + projection

(see G. Punzi, "Including systematic uncertainties in Confidence Limits", CDF report (2003))

x number of signal events when the true average is A (**unknown**)
y number of background events in the signal region
b number of background events when the true average is B (**unknown**)

one measures $a = x+y$ and b ;
assume (a,b) to be distributed like $f(a,b; A,B)$

For each point (A,B) define an acceptance region $R_{ab}(A,B)$ of (a,b) by

the condition $\int_{R_{ab}} f(a,b;A,B)\cdot da\ db = \alpha$ (α = confidence level)

using some ordering algorithm.

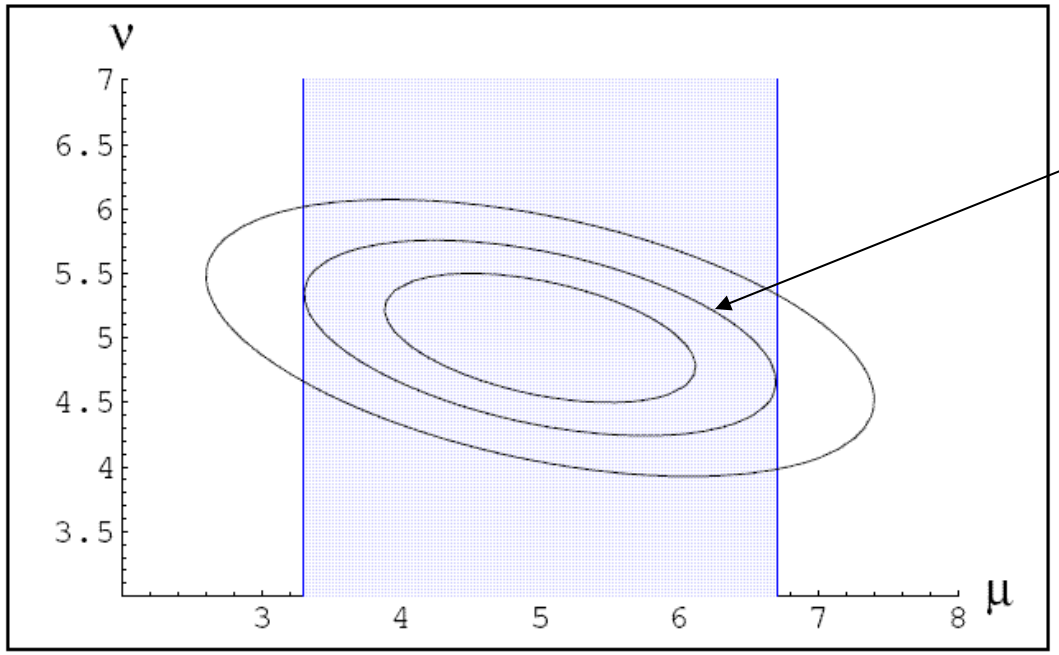
Then proceed as usual to find, for a given measurement (a,b), the confidence region $R_{AB}(a,b)$ of (A,B) : include in $R_{AB}(a,b)$ all those (A,B) for which (a,b) is within $R_{ab}(A,B)$.

In order to get the limits for the physical parameter A, one needs to "project" the confidence region in the (A,B) plane onto the A axis.

"Projection" of the confidence region in the (μ,ν) plane onto the μ axis :

the confidence region in (μ,ν) has exact coverage, by construction;

by the extension to the shaded area (blue) the coverage is increased



68 % confidence region for a given measurement

μ is physical parameter

ν is nuisance parameter

Figure 3: Likelihood ratio contours, and CR on μ obtained from either P or LR-ordering in the (μ, ν) space (F–C)

(G. Punzi, "Including systematic uncertainties in Confidence Limits", CDF report (2003))

Problems with this procedure :

- the projection often leads to large overcoverage
- the procedure is sensitive to the choice of the ordering algorithm
- often the limits obtained with small systematics are quite different from the ones obtained in the absence of that systematics
- the calculations can be complex and CPU-time consuming

II. Profile Likelihood ratio used as ordering quantity

(M. Kendall and A. Stuart, "The Advanced Theory of Statistics" (1961),
K.S. Cranmer, PHYSTAT2003, G. Punzi, PHYSTAT05)

The *profile LR* is defined as

$$l(a,b,A) = \frac{f(a,b;\ A,\hat{\hat{B}})}{f(a,b;\ \hat{A},\hat{B})}$$

where $\hat{\hat{B}}$ maximizes $f(a,b;\ A,B)$ with fixed a,b,A
and $\hat{A},\hat{B}$ maximize $f(a,b;\ A,B)$ with fixed a,b

$l(a,b;\ A)$ is a good test statistic for the hypothesis A.

Do the Neyman construction in 2 dimensions (A,B). Use *profile LR* (which is independent of B) as ordering quantity, for each value of B.
To be used together with the pdf $f(a,b;\ A,B)$ (which depends on B)
for constructing the acceptance regions $R_{ab}(A,B)$:

$$R_{ab}(A,B) = \{\, a,b \mid l(a,b;A) > l_\alpha(A,B) \,\} \qquad \int_{R_{ab}} f(a,b;A,B)\cdot da\, db = \alpha$$

Note : acceptance regions for different B, for the same A, will only differ in $l_\alpha(A,B)$, and will therefore be included in one another.

Aim : by using an ordering quantity which is independent of B construct acceptance regions $R_{ab}(A,B)$ which depend only little on B.

$$R_{ab}(A,B) = \{ a,b \mid l(a,b;A) > l_\alpha(A,B) \}$$

is independent of B if $l_\alpha(A,B)$ is independent of B

Modify the above definition of $R_{ab}(A,B)$ by excluding those b which are extremely unlikely: $f(a,b;A,B) < \varepsilon$

Special ordering function

G.Punzi, physics/0511202, PHYSTAT05

$$\int_{f(x,e;\mu,\epsilon)>c(\mu,\epsilon)} p(x,e|\mu,\epsilon)\,dx\,de \geq CL$$

$$f(x,e;\mu) = \int_{f_0(x')<f_0(x)} p(x'|e;\mu,\hat{\epsilon}(e))\,dx'$$

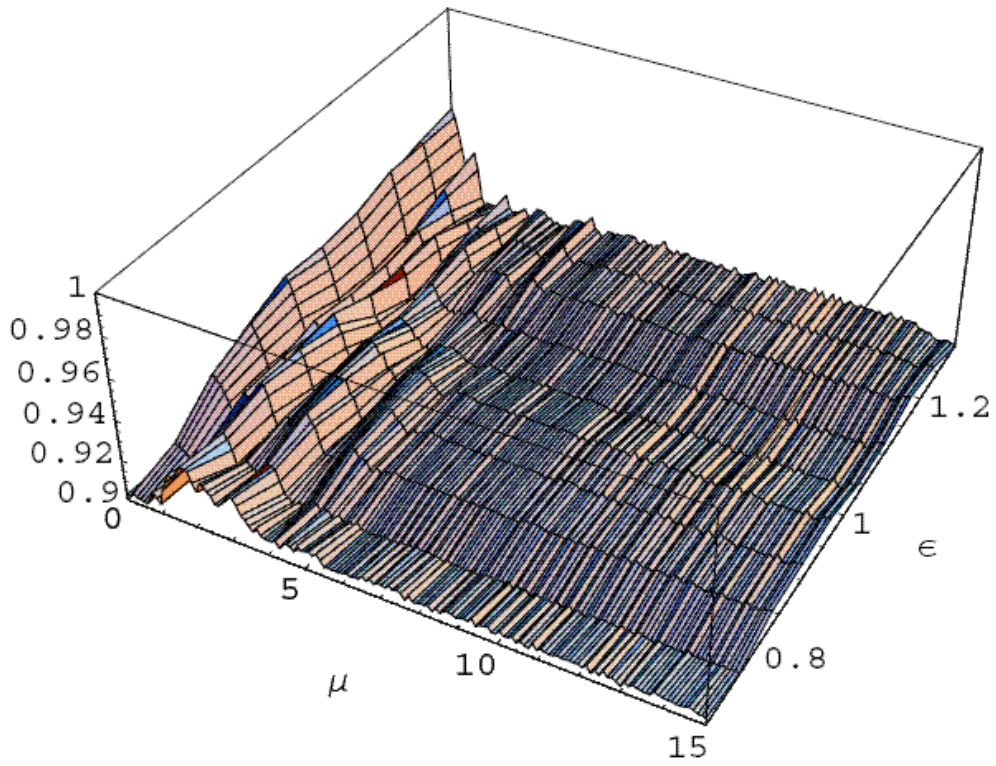


Fig. 2. Coverage plot for Unified limits, Gaussian uncertainty, $b = 3, \sigma = 0.1$.

With $f_0(x) = f(x;\mu) / f(x;\mu_{best})$, this ordering principle is approximately equivalent to an ordering based on the *profile LR*

Example :

$f(x,e;\mu,\varepsilon,b) = \text{Poisson}(x;\ \varepsilon\mu+b) \cdot \text{Gauss}(e;\ \varepsilon,\sigma)$

coverage is close to $\alpha = 90\ \%$

III. Profile Likelihood ratio used to construct confidence intervals

(G. Feldman, Fermilab workshop 2000, W.A.Rolke et al., NIM A458 (2001) 745)

Eliminate nuisance parameter by constructing the acceptance region of A for the optimum value of the nuisance parameter.

a number of signal events when the true average is A (**unknown**)

b number of background events in the signal region when the true average is B (**unknown**)

y number of background events in the background region when the true average is $\tau \cdot B$; τ = (size of bg region) / (size of signal region)

One measures $x = a + b$ and y. Assume x and y to be distributed like $f(x, y; A, B) = Poisson(x; A + B) \cdot Poisson(y; \tau \cdot B)$ where both A and B are unknown.

Case $x >> y/\tau$

(no. of observed events in signal region >> average expected background)

- Define the likelihood ratio $\Lambda(A; x, y) = \dfrac{f(x, y; A, \hat{B})}{\max\limits_{A,B}(f(x, y; A, B))}$

  $\hat{B}(A; x, y)$ is that B which maximizes $f(x, y; A, B)$, at fixed A, x, y

  (profile likelihood function)

$\Lambda(A; x,y)$ is a good test statistic for the hypothesis A. At fixed A, i.e. if A is the true value of the parameter, $L(A; x,y) = -2\ ln\ \Lambda(A; x,y)$ is approximately χ^2-distributed with

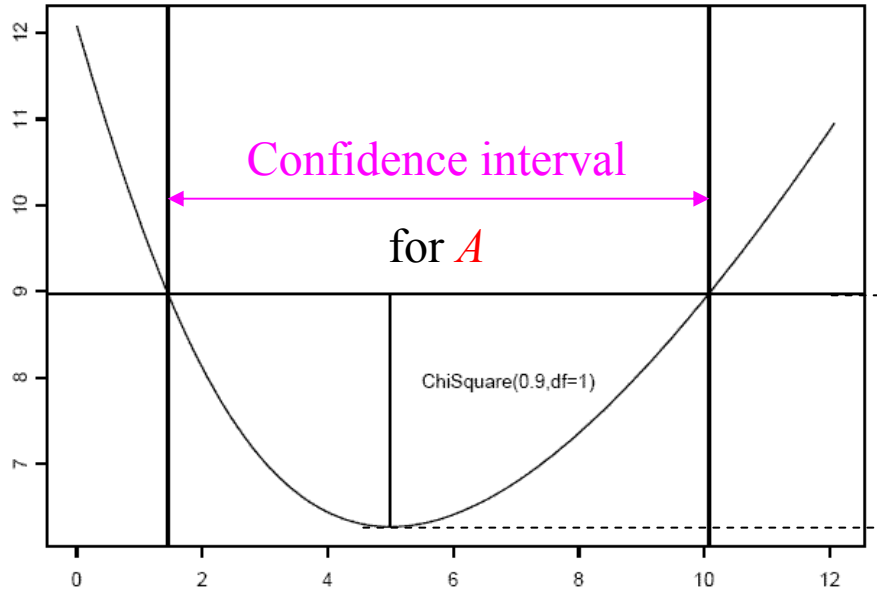
no.of degrees of freedom = npar_{tot}-$\text{npar}_{\text{fixed}}$ $(x >> y/\tau)$

- Define acceptance regions $R_{xy}(A)$ using the distribution of

$\Lambda(A; x,y)$: $R_{xy}(A) = \{ x,y; \mid L(A; x,y) < \chi^2_{cut}(\alpha) \}$

- Given a measurement (x,y), accept all those A in the confidence interval $R_A(x,y)$ of A for which (x,y) is acceptable, i.e for which $L(A; x,y) < \chi^2_{cut}(\alpha)$.

-2 ln Λ(A; x,y) vs. A



$(x,y) = (6,2),\ \tau = 2,\ \alpha = 90\ \%$

$\rightarrow \quad 6 = x > y/\tau = 1$

$\chi^2 = 2.706 = \chi^2_{cut}(\alpha)$

$\chi^2 = 0$

W.A. Rolke and A.M. Lopez, NIM A458 (2001) 745

Fig. 2. Profile likelihood function with two-sided limits for the case of $x = 6$ events in the signal region and $y = 2$ events in the background region. The background region is twice the size of the signal region ($\tau = 2$). The nominal coverage probability is 0.9.

- Since $\chi^2_{cut}(\alpha)$ is independent of A only the actual measurement (x,y) enters. The method thus obeys the **likelihood principle**.
- The acceptance intervals are determined using the pdf of $L(A; x,y)$ at fixed A, not the pdf of $L(A; x,y)$ at fixed (x,y). Thus the method is a **Frequentist method**.
- Although the construction of the acceptance intervals is done effectively in 1 dimension (the physics parameter), **coverage** is tested in 2 dimensions (physics + nuisance parameter). The method yields confidence intervals with good coverage, throughout the parameter space, even at its boundaries.

In contrast to Cranmer, Punzi $\Lambda(A; x,y)$ is not used as ordering quantity but rather as a test statistic to define the acceptance regions $R_{xy}(A)$ in (x,y) and the confidence interval $R_A(x,y)$ of A.

The same method is also used in the minimization program package MINUIT (F. James).

Case $x \leq y/\tau$

(no. of observed events in signal region ≤ average expected background)

- Determine for each (A,B) the acceptance region $R_{xy}(A,B)$ of (x,y)
- Given the measurement (x,y), define the confidence region $R_{AB}(x,y)$ of (A,B)
- Determine the limits of A as the intersection points of the profile likelihood curve $\hat{B}(A)$ with the contour of $R_{AB}(x,y)$

W.A. Rolke and A.M. Lopez, NIM A458 (2001) 745

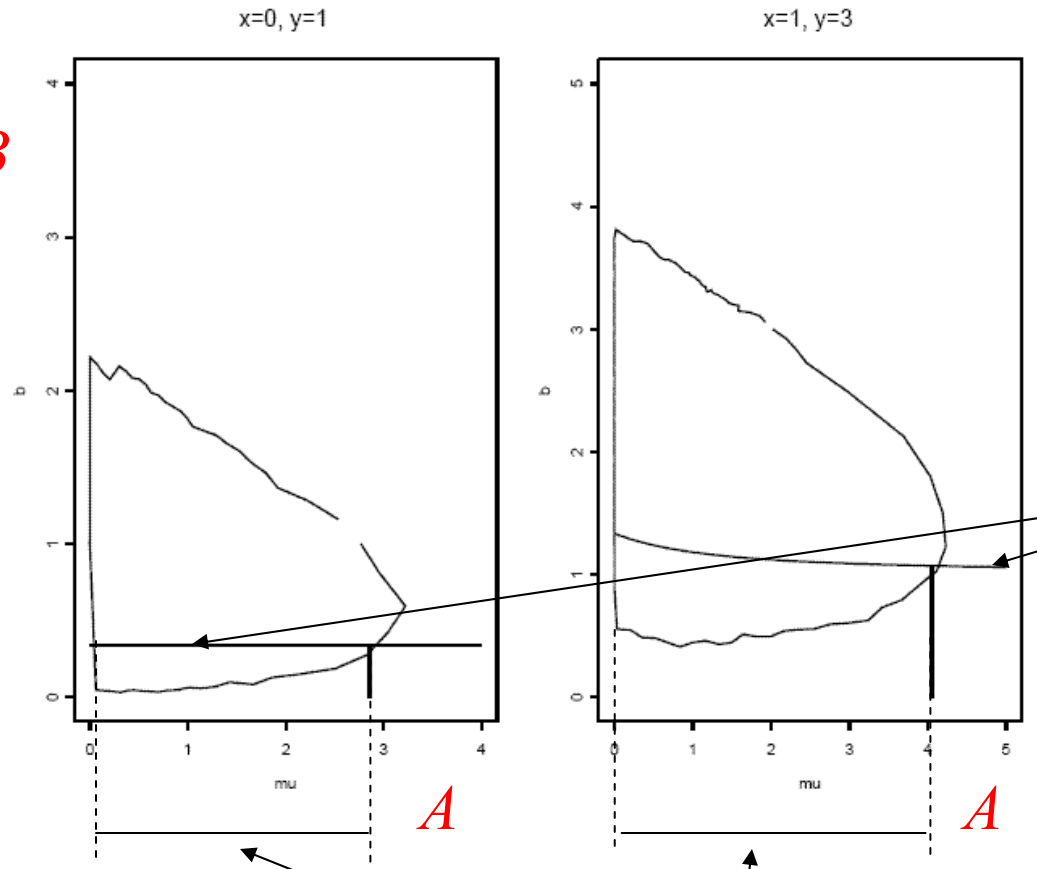


Fig. 3. Confidence region with profile likelihood curve and upper bounds for the cases $x = 0, y = 1$ and $x = 1, y = 3$.

$\tau = 2, \ \alpha = 90\ \%$

$(x,y) = (0,1) \ \rightarrow \ 0 = x < y/\tau = 0.5$

$(x,y) = (1,3) \ \rightarrow \ 1 = x < y/\tau = 1.5$

profile likelihood function

$\hat{B}(A; x, y)$

(it maximizes $f(x,y; A,B)$)

confidence regions of A

IV. Mixed Frequentist-Bayesian approach

(R.D. Cousins and V.L. Highland, NIM A320 (1992) 331,

G.C. Hill, Comments on "Including systematic uncertainties in confidence interval construction for Poisson statistics",

PR D67 (2003) 118101; J. Conrad et al., PR D67 (2003) 012002)

$f(x; A) = Poisson(x; A)$

Assume that there is an additional systematic normalization error in the order σ_s. Take this into account by replacing $f(x; A)$ by

$$f(x; A, \sigma_s) = \int f(x; s \cdot A) \cdot Gauss(s;\ 1, \sigma_s) \cdot ds$$

Use this probability distribution for constructing the acceptance interval $R_a(A)$, assuming a certain ordering quantity O :

F&C : $O_1 = f(x; A, B) / f(x; A_{best}, B)$ where B is known and fixed

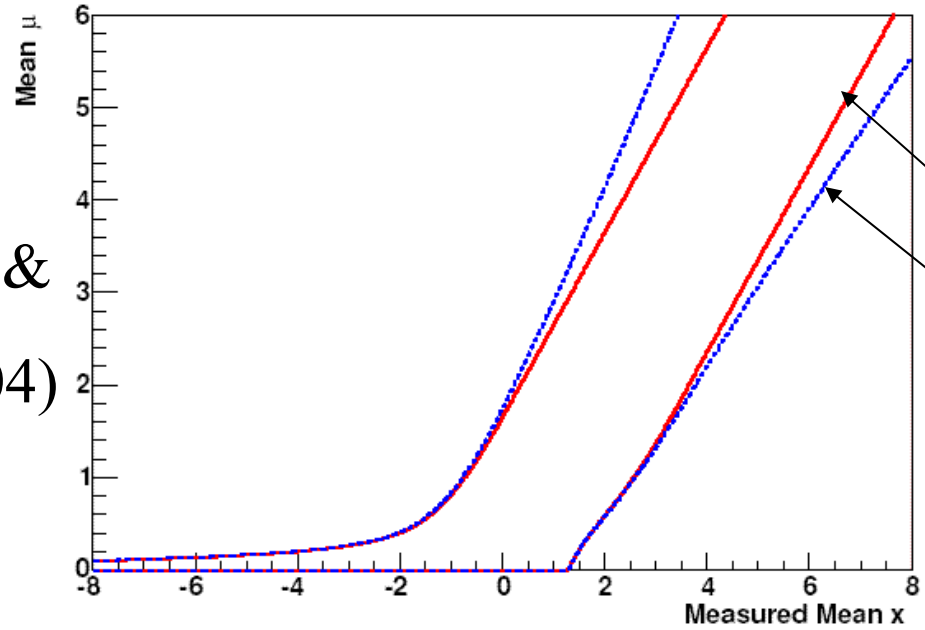
here : $O_2 = f(x; A, \sigma_s) / f(x; A_{best}, \sigma_s)$ yields unsatisfactory results

better : $O_3 = f(x; A, \sigma_s) / f(x; A_{best})$

$$f(x; A, \sigma_s) = \int Poisson(x; s \cdot A) \cdot Gauss(s; 1, \sigma_s) \cdot ds$$

ordering quantity $LR = \dfrac{f(x; A, \sigma_s)}{Poisson(x; A_{best})}$, $\alpha = 90\%$

A



best *LR* confidence intervals for A :

$\sigma_s = 0$

$\sigma_s = 20\ \%$

U.Schwanke &

T.Lohse (2004)

Figure 7: *Confidence intervals calculated in the Feldman Cousins approach for a CL of 90 %. The solid red curves correspond to the case of no systematic error. The blue dotted curves were calculated using the Likelihood Ratio of Eq. (7) with a systematic error of* $\sigma_s = 20\,\%$.

End of part 2

Critisism of Bayesian approach

- prior distribution represents a subjective belief about A; the results is therefore not an objective answer to the problem
- the use of a uniform prior distribution (to express ignorance about A) is problematic (dependence on metric)
- Bayesian confidence intervals have bad Frequentist coverage (over and/or undercoverage)

Main advantages :

- easy treatment of nuisance parameters
- Bayesian approach obeys the likelihood principle (only the actual measurements enter in the calculations)

Critisism of Frequentist (classical) approach

- elimination of nuisance parameters is problematic
- the likelihood principle is violated; this means that some available information is ignored
- results are sometimes counterintuitive :
  - with 0 observed events upper limit decreases with increasing average background level
  - adding additional information causes the limits to widen dramatically (G. Punzi, Durham 2002)

Main advantages :

- confidence intervals have exact Frequentist coverage (except with discrete measurements)
- invariance against variable and parameter transformations, independent of the dimensions (in the unified classical approach)

Critisism of *LR* approach

- elimination of nuisance parameters is problematic
- if maximum of *LR* is outside (or close to the border of) the physical region errors become unrealistic
- simple *LR* intervals are not confidence intervals

Main advantages :

- invariance against variable and parameter transformations
- combination of measurements (without loss of information) is straightforward : add their log-likelihood functions
- the *LR* approach is able to handle a discrete sample space

Criteria used when comparing different approaches

- Consistency : Support of a hypothesis must not be affected by information judged intuitively to be irrelevant (example : Poisson case with expected average background and 0 events observed).
- Precision : The interval should represent a measure for the relative precision of different experimental results; the interval should serve as a check of the compatibility of the measurement with a theoretical prediction.
- Universality : The method should cope with all special cases, such as elimination of nuisance parameters, discrete and continuous measurements and parameters, … .
- Objectivity : The results should not depend on the experimenter's subjective believe.

- Coverage : Good (Frequentist) coverage means that the true value is within the interval with high probability. It does not mean that any parameter within the interval is true with high probability.
- Invariance against transformations of measurements and parameters : The intervals should be the same using the original variables or the transformed ones.
- Nuisance parameters : Elimination of nuisance parameters should be possible.
- Combining data : The combination of confidence intervals should be possible.
- Likelihood principle : The intervals should not violate this principle.
- Bias : the expectation value of the estimator of a parameter should be equal to the parameter.

- Error propagation : requires not only an error interval but also a parameter estimate
- The limits given should effectively convey the information content of the experiment

Comparison of different approaches

Table 4. Comparison of different approaches to define error intervals,

method:	classical	unified classical	likelihood ratio	Bayesian u.p.	Bayesian a.p.
consistency	- -	- -	++	+	+
precision	- -	-	+	+	+
universality	- -	- -	-	+	++
simplicity	-	- -	++	+	+
variable transform.	-	++	++	- -	- -
nuisance parameter	-	-	-	+	+
error propagation	-	-	+	+	+
combining data	-	-	++	+	-
coverage	+	++	- -	- -	- -
objectivity	-	-	++	+	-
discrete hypothesis	-	-	+	+	+

(G. Zech, "Frequentist and Bayesian confidence intervals", hep-ex/0106023)

uniform prior

arbitrary prior

Choice of the "best" approach

B, F, LR

It depends on the experimentor's intention :

- is one interested in the uncertainty of a measurement (B, LR)
- does one want to verify or reject a theory (F)
- does one want to estimate the parameter in addition to determining an error interval (LR)
- does one want to combine measurements (LR)

"You see, a question has arisen, about which we cannot come to an agreement, probably because we have read too many books"

(Bertold Brecht, "Leben des Galilei")

L. Lyons, "Bayes or Frequentism" ? CDF report (2002)

- It is not a question of which of the two approaches (Bayesian or Frequentist) is correct, but rather the consumer should be aware exactly what each method has to offer, and what are its limitations and pitfalls.
- Because there are so many options in calculating ranges and even more so for limits, it is crucial to state clearly what procedure was used.
- It is useful to provide to the reader more information than just the final results on the ranges or limits.

Table 1: Comparison of Bayes and Frequentist philosophies

Method	Bayes	Frequentist
Probability	Degree of belief	Limit of frequency ratio
$P(\mu)$?	Yes	Anathema
Need for prior?	Yes	No
What is used?	$L(\mu\|x)$ Only observed data	$f(x\|\mu)$ Also other possible data
$\mu_l \leq \mu \leq \mu_u$	μ = random variable μ_l, μ_u fixed by this expt	μ = unknown but fixed μ_l , μ_u = random variables
Nuisance parameters	Marginalise (integrate)	Maximise wrt them
Prob of which data?	Only what you observed	Also more extreme
Empty intervals?	Intervals always physical	Can happen
Problem with $\mu \geq 0$?	Limits always physical	Can give empty interval
Concept of coverage	Not seen as relevant Achieves average coverage	Basic importance Covers (or overcovers) for any μ
Need to define ensemble?	No	Yes
Decisions	Requires cost function	No. Needs prior and cost fn.

L. Lyons, CDF report (2002)

Methods not recommended

- multidimensional Neyman construction with simple projection onto the space of physical parameters (strong overcoverage)

Reasonable methods

- Bayesian approach (Helene)
- Frequentist approaches using the profile *LR* as ordering rule (Cranmer, Punzi)
- Approaches which use the pdf of the profile *LR* to calculate confidence intervals (MINUIT, Rolke)
- ……