# BAYES and FREQUENTISM: The Return of an Old Controversy

Louis Lyons

Oxford University

CERN, October 2006

Lecture 4

# Topics

- The ideologies, with examples
- Upper limits
- Feldman and Cousins
- Ordering Rule
- Systematics
- Summary

It is possible to spend a lifetime analysing data without realising that there are two very different approaches to statistics:

Bayesianism and Frequentism.

# How can textbooks not even mention

# Bayes/ Frequentism?

For simplest case $(m \pm \sigma) \leftarrow Gaussian$

with no constraint on $m(true)$ then

$$m - k\sigma < m(true) < m + k\sigma$$

at some probability, for both Bayes and Frequentist

(but different interpretations)



See Bob Cousins "Why isn't every physicist a Bayesian?" Amer Jrnl Phys 63(1995)398

## We need to make a statement about
## Parameters, Given Data

The basic difference between the two:

Bayesian : Probability (parameter, given data)
(an anathema to a Frequentist!)

Frequentist : Probability (data, given parameter)
(a likelihood function)

# PROBABILITY

## MATHEMATICAL

Formal

Based on Axioms

## FREQUENTIST

Ratio of frequencies as $n \rightarrow$ infinity

Repeated "identical" trials

Not applicable to single event or physical constant

## BAYESIAN Degree of belief

Can be applied to single event or physical constant

(even though these have unique truth)

Varies from person to person ***

Quantified by "fair bet"

# Bayesian versus Classical

## Bayesian

$P(A \text{ and } B) = P(A;B) \times P(B) = P(B;A) \times P(A)$

e.g. A = event contains t quark

B = event contains W boson

or A = I am in Switzerland

B = I am giving a lecture

Completely uncontroversial, provided….

$$P(A;B) = P(B;A) \times P(A) / P(B)$$

Bayesian

$$P(A;B) = \frac{P(B;A) \text{ x } P(A)}{P(B)}$$

Bayes Theorem

$$\underbrace{P(hyothesis; data)}_{\text{posterior}} \;\alpha\; \underbrace{P(data; hypothesis)}_{\text{likelihood}} \text{ x } \underbrace{P(hypothesis)}_{\text{prior}}$$

Problems: P(hyp..) true or false

"Degree of belief"

Prior What functional form?

Coverage

Goodness of fit

P(hypothesis…..) True or False

"Degree of Belief"

credible interval

Prior: What functional form?

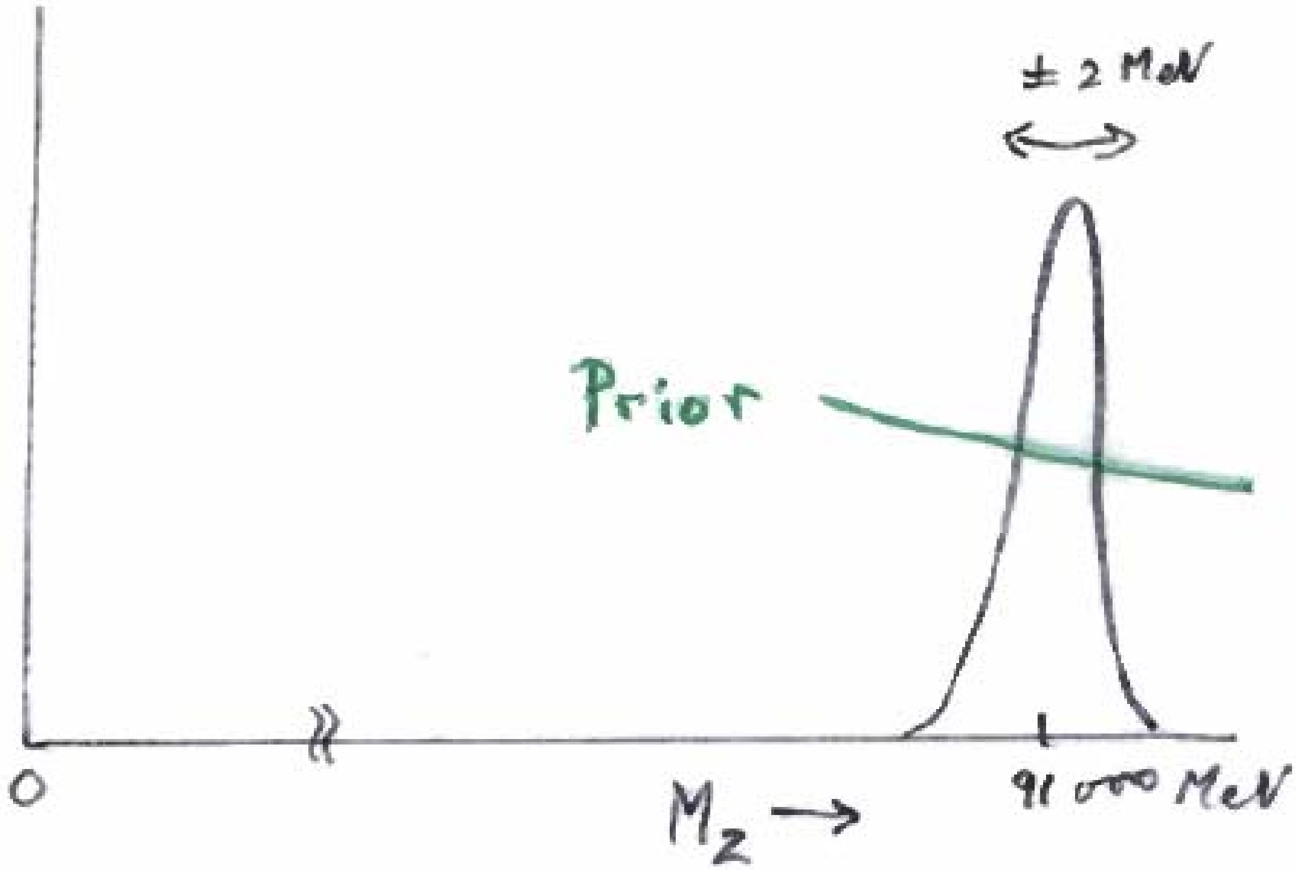
Uninformative prior: flat?

In which variable? e.g. $m, m^2, \ln m, ....?$

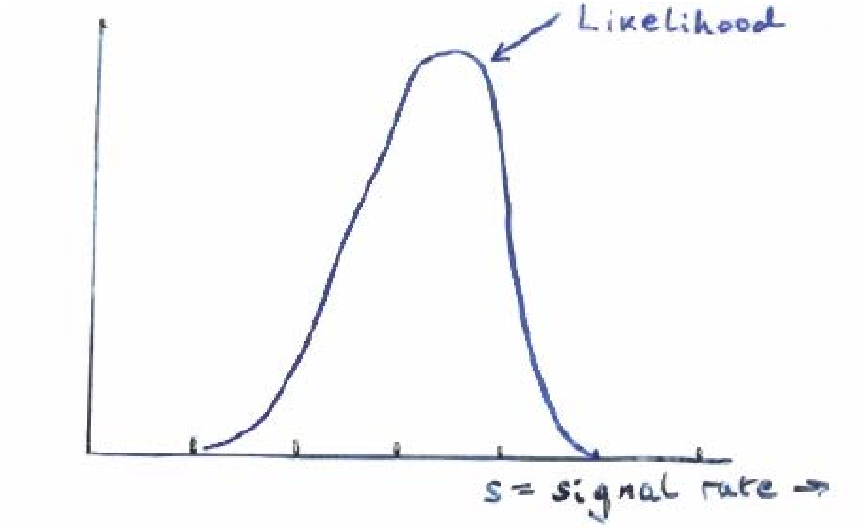
Unimportant if "data overshadows prior"

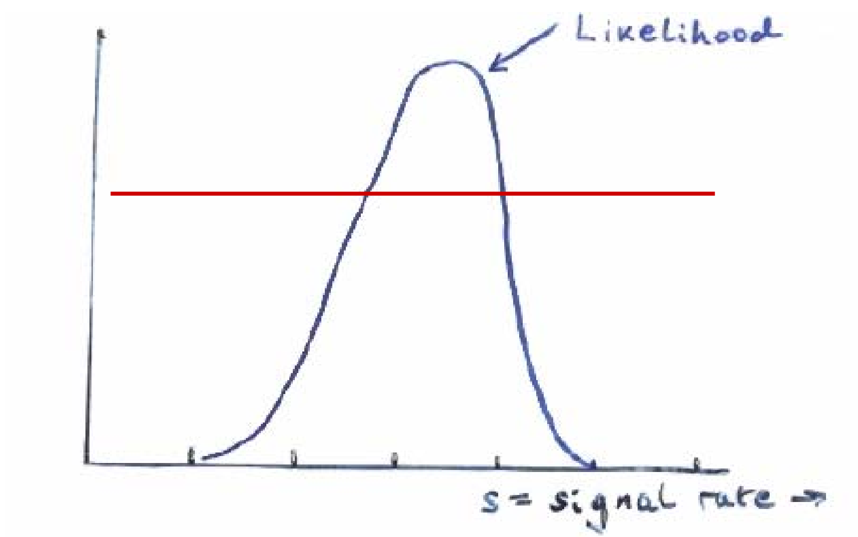
Important for limits

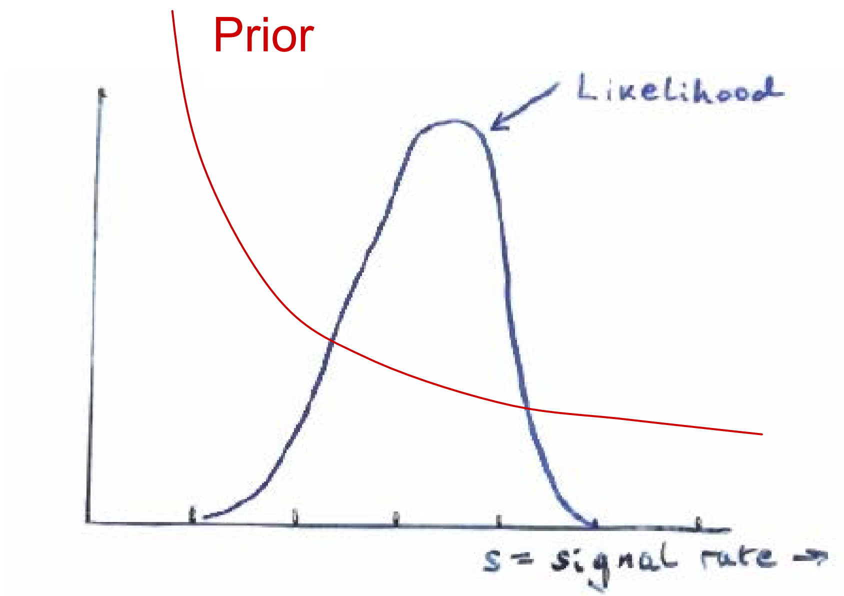
Subjective or Objective prior?




Data overshadows the Prior


Likelihood
s = signal rate →


Likelihood
s = signal rate →


Prior
Likelihood
s = signal rate →

P (Data;Theory) $\neq$ P (Theory;Data)

HIGGS SEARCH at CERN

Is data consistent with Standard Model?

or with Standard Model + Higgs?

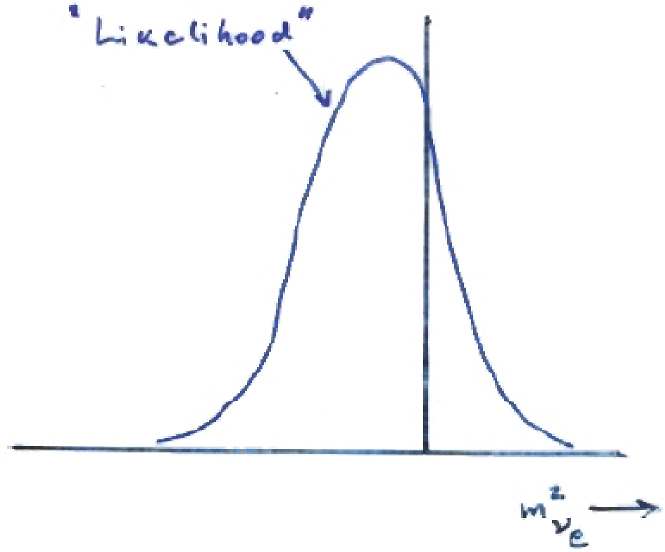
End of Sept 2000 Data not very consistent with S.M.

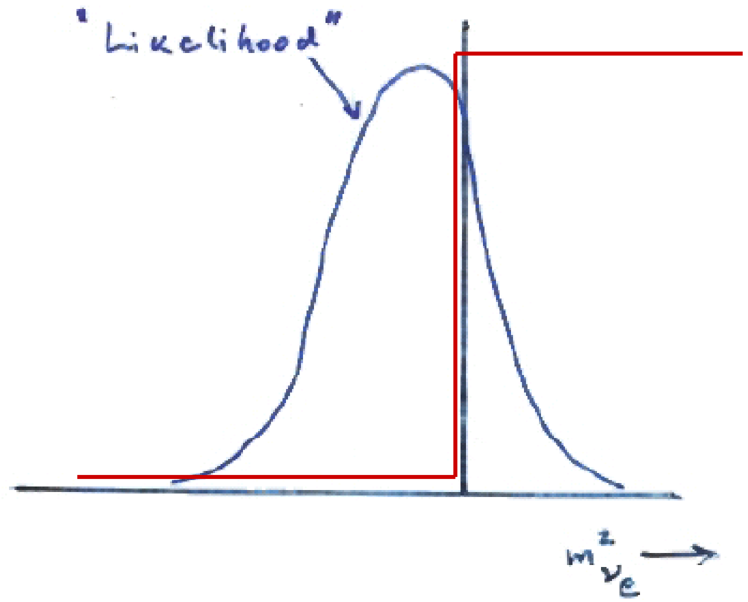
Prob (Data ; S.M.) < 1% valid frequentist statement

Turned by the press into: Prob (S.M. ; Data) < 1%

and therefore Prob (Higgs ; Data) > 99%

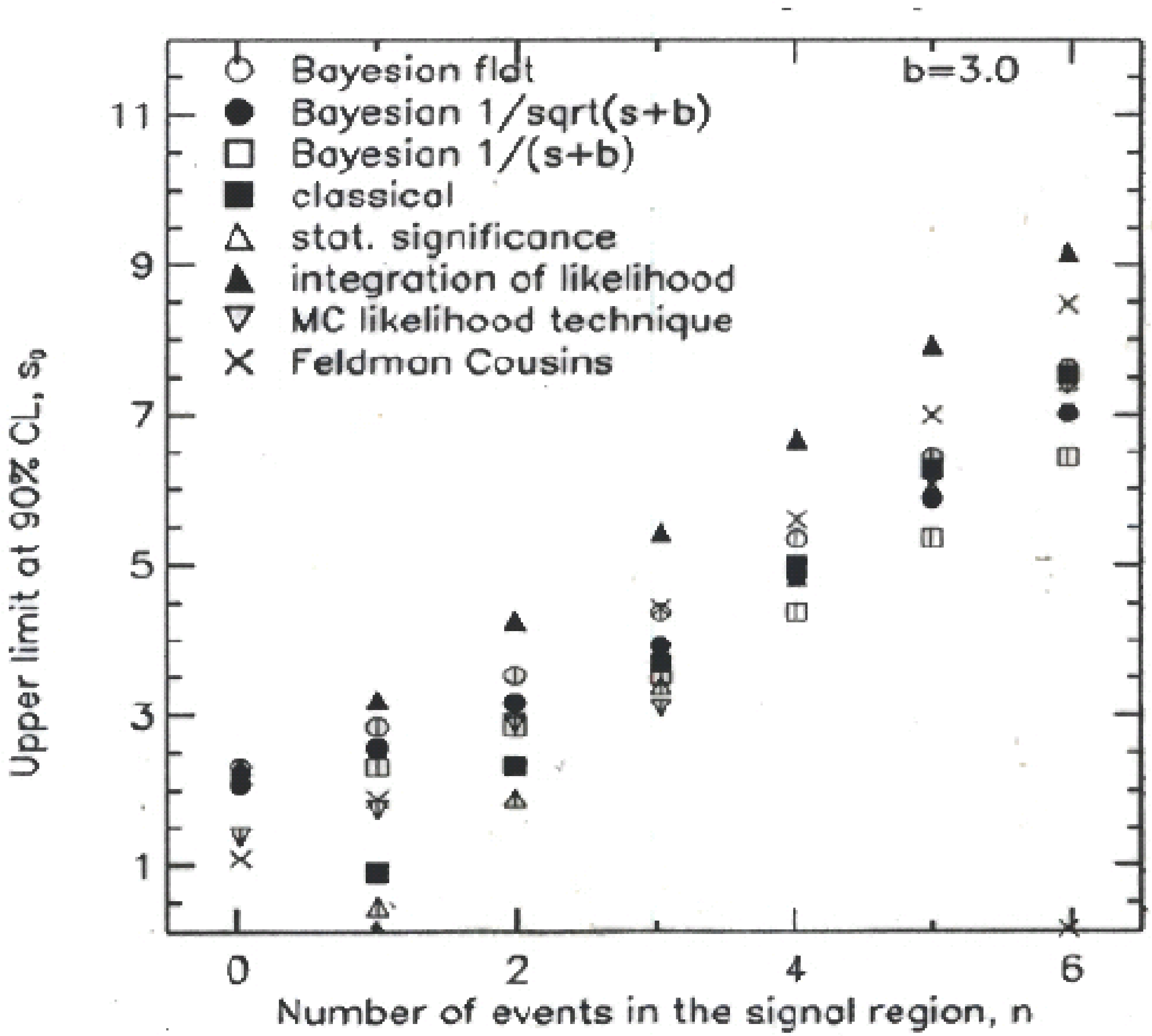
i.e. “It is almost certain that the Higgs has been seen”




Prior = zero in unphysical region

# Ilya Narsky, FNAL CLW 2000



$P(\text{Data;Theory}) \neq P(\text{Theory;Data})$

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

$$P\,(\text{Data;Theory}) \neq P\,(\text{Theory;Data})$$

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

but

P (female ; pregnant) >>>3%

Example 1 : Is coin fair ?

Toss coin: 5 consecutive tails

What is P(unbiased; data) ? i.e. $p = ½$

Depends on Prior(p)

If village priest prior ~ $\delta(1/2)$

If stranger in pub prior ~ 1 for $0<p<1$

(also needs cost function)

## Example 2 : Particle Identification

Try to separate $\pi$ and protons

probability (p tag;real p) = 0.95

probability ( $\pi$ tag; real p) = 0.05

probability (p tag ; real ( $\pi$) = 0.10

probability ( $\pi$ tag ; real $\pi$ ) = 0.90

Particle gives proton tag. What is it?

Depends on prior = fraction of protons

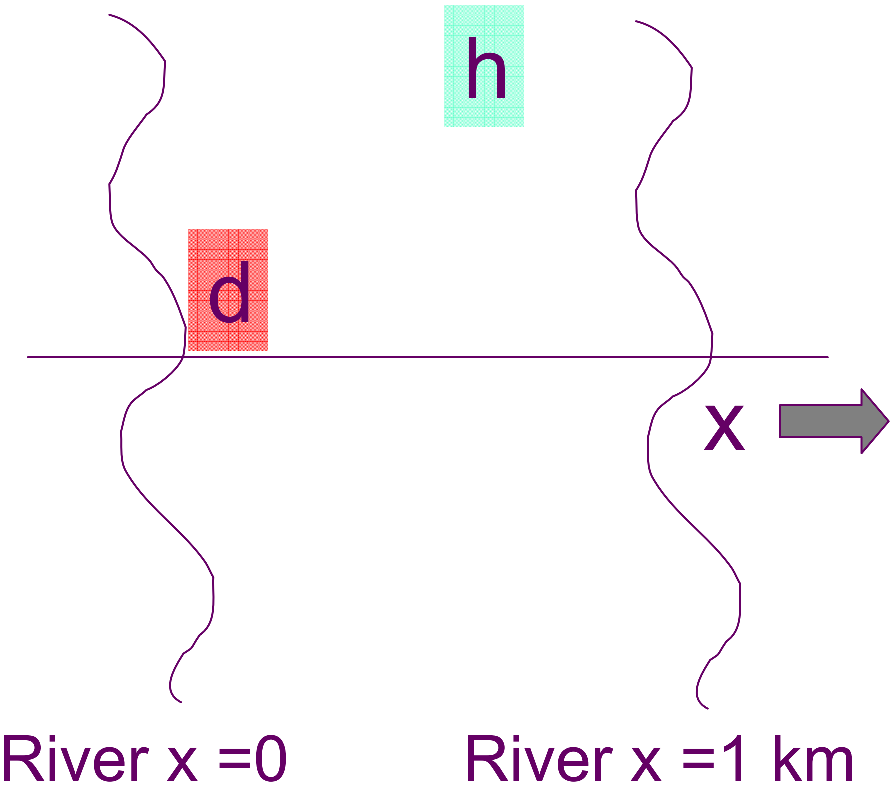
If proton beam, very likely

If general secondary particles, more even

If pure $\pi$ beam, ~ 0

# Hunter and Dog

1) Dog d has 50% probability of being 100 m. of Hunter h
2) Hunter h has 50% probability of being within 100m of Dog d



Given that: a) Dog d has 50% probability of being 100 m. of Hunter

Is it true that b) Hunter h has 50% probability of being within 100m of Dog d ?

Additional information

- Rivers at zero & 1 km. Hunter cannot cross them.

$$0 \leq h \leq 1 \, \text{km}$$

- Dog can swim across river - Statement a) still true

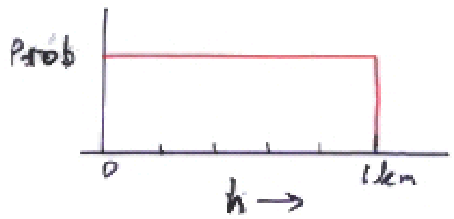
If dog at –101 m, hunter cannot be within 100m of dog

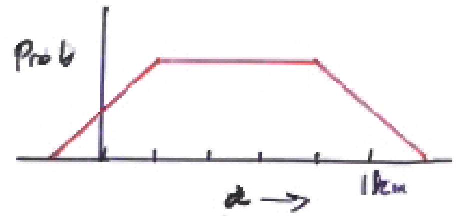
Statement b) untrue

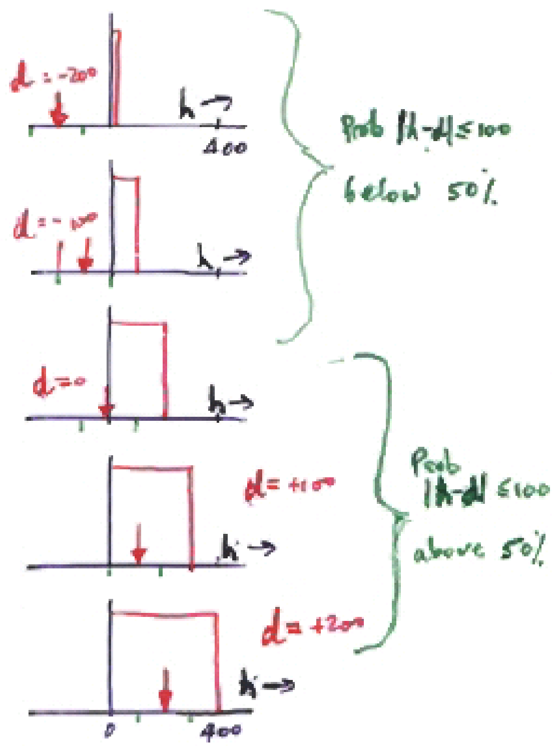
1) More specific on statement ①:

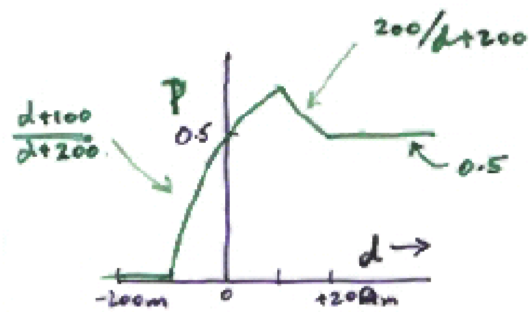
$$\text{Prob}(d-h) = \begin{cases} \text{const} & \text{for } |d-h| < 200\,\text{m} \\ 0 & \text{for } |d-h| > 200\,\text{m} \end{cases} \quad [\text{L'Hood}]$$

2) Hunter $h$ uniform in $0 \rightarrow 1$ km [PRIOR]




$P$ = prob $|h-d| \leq 100$ m

BF7 9''

# Classical Approach

Neyman "confidence interval" avoids pdf for $\mu$

uses only P( x; $\mu$ )

Confidence interval $\mu_1 \rightarrow \mu_2$ :

P( $\mu_1 \rightarrow \mu_2$ contains $\mu$ ) = $\alpha$ True for any $\mu$

Varying intervals from ensemble of experiments

fixed

Gives range of $\mu$ for which observed value $x_0$ was "likely" ($\alpha$ )

Contrast Bayes : Degree of belief = $\alpha$ that $\mu_t$ is in $\mu_1 \rightarrow \mu_2$

CLASSICAL (NEYMAN) CONFIDENCE INTERVALS

Uses only P (data | theory)

FIGURES

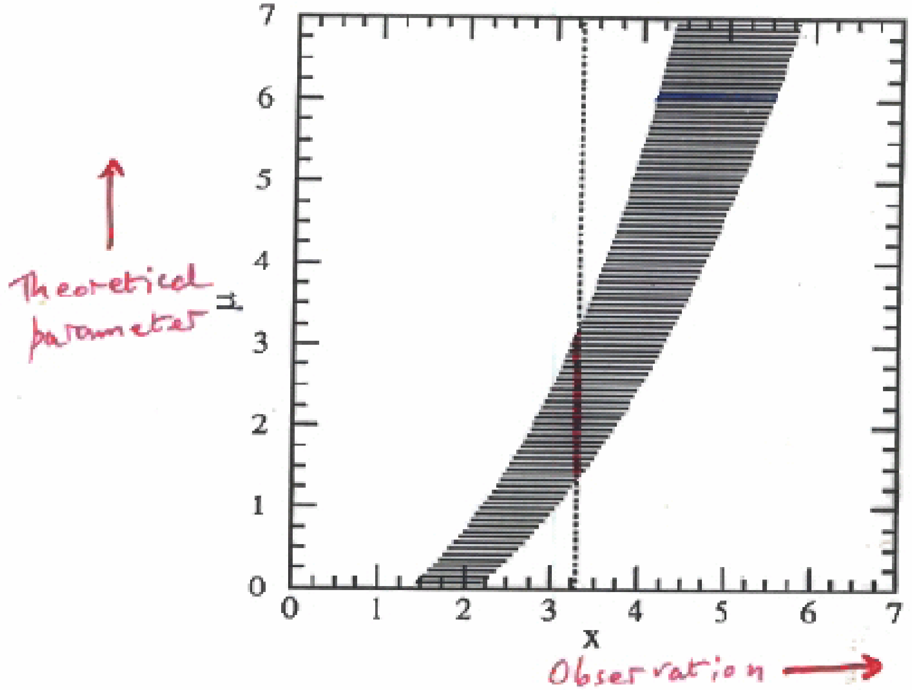



FIG. 1. A generic confidence belt construction and its use. For each value of $\mu$, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] | \mu) = \alpha$. Upon performing an experiment to measure $x$ and obtaining the value $x_0$, one draws the dashed vertical line through $x_0$. The confidence interval $[\mu_1, \mu_2]$ is the union of all values of $\mu$ for which the corresponding acceptance interval is intercepted by the vertical line.

NO PRIOR INVOLVED

# 90% Classical interval for Gaussian

$\sigma = 1$ $\mu \geq 0$ e.g. $m^2(\nu_e)$

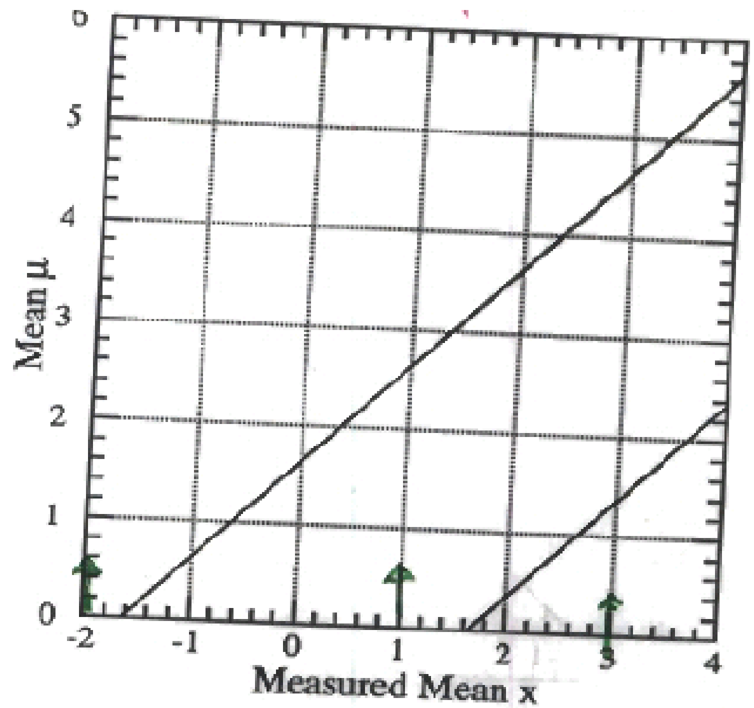



FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$x_{obs} = 3$ Two sided limit

$x_{obs} = 1$ Upper limit

$x_{obs} = -2$ No region for $\mu$

$\mu_l \leq \mu \leq \mu_u$ at 90% confidence

Frequentist

$\mu_l$ and $\mu_u$ known, but random

$\mu$ unknown, but fixed

Probability statement about $\mu_l$ and $\mu_u$

Bayesian

$\mu_l$ and $\mu_u$ known, and fixed

$\mu$ unknown, and random

Probability/credible statement about $\mu$

# Classical Intervals

- Problems

Hard to understand e.g. d'Agostini e-mail
Arbitrary choice of interval
Possibility of empty range
Over-coverage for integer observation
e.g. # of events
Nuisance parameters (systematic errors)

- Advantages

Widely applicable
Well defined coverage

# FELDMAN - COUSINS

WANT TO AVOID EMPTY CLASSICAL INTERVALS ⇒

USE "$\mathcal{L}$ RATIO ORDERING PRINCIPLE"

TO RESOLVE AMBIGUITY ABOUT "WHICH 90% REGION?" ⇒

[NEYMAN-PEARSON SAY $\mathcal{L}$ RATIO IS BEST FOR HYPOTHESIS TESTING]

NO FLIP-FLOP PROBLEM ⇒

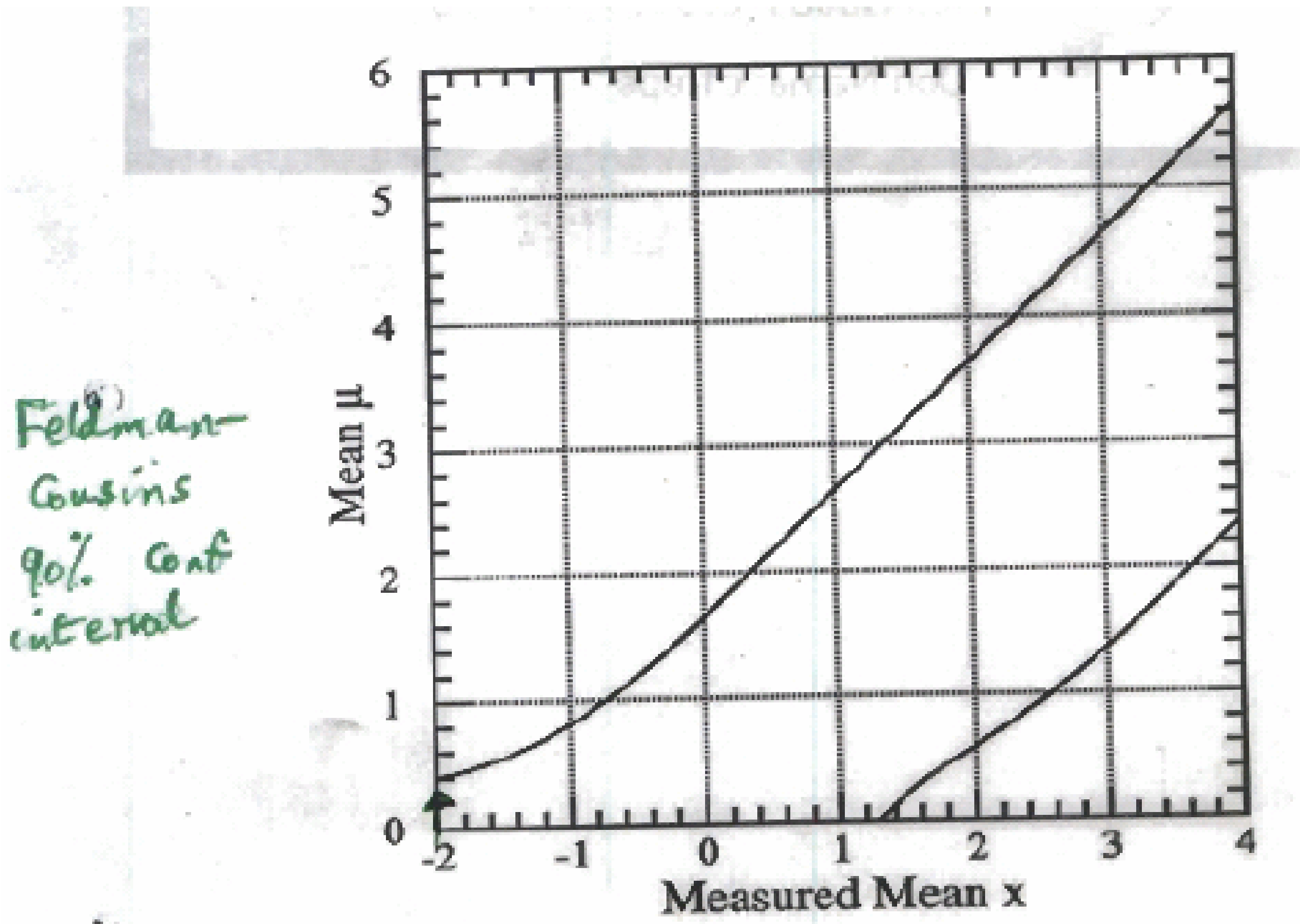
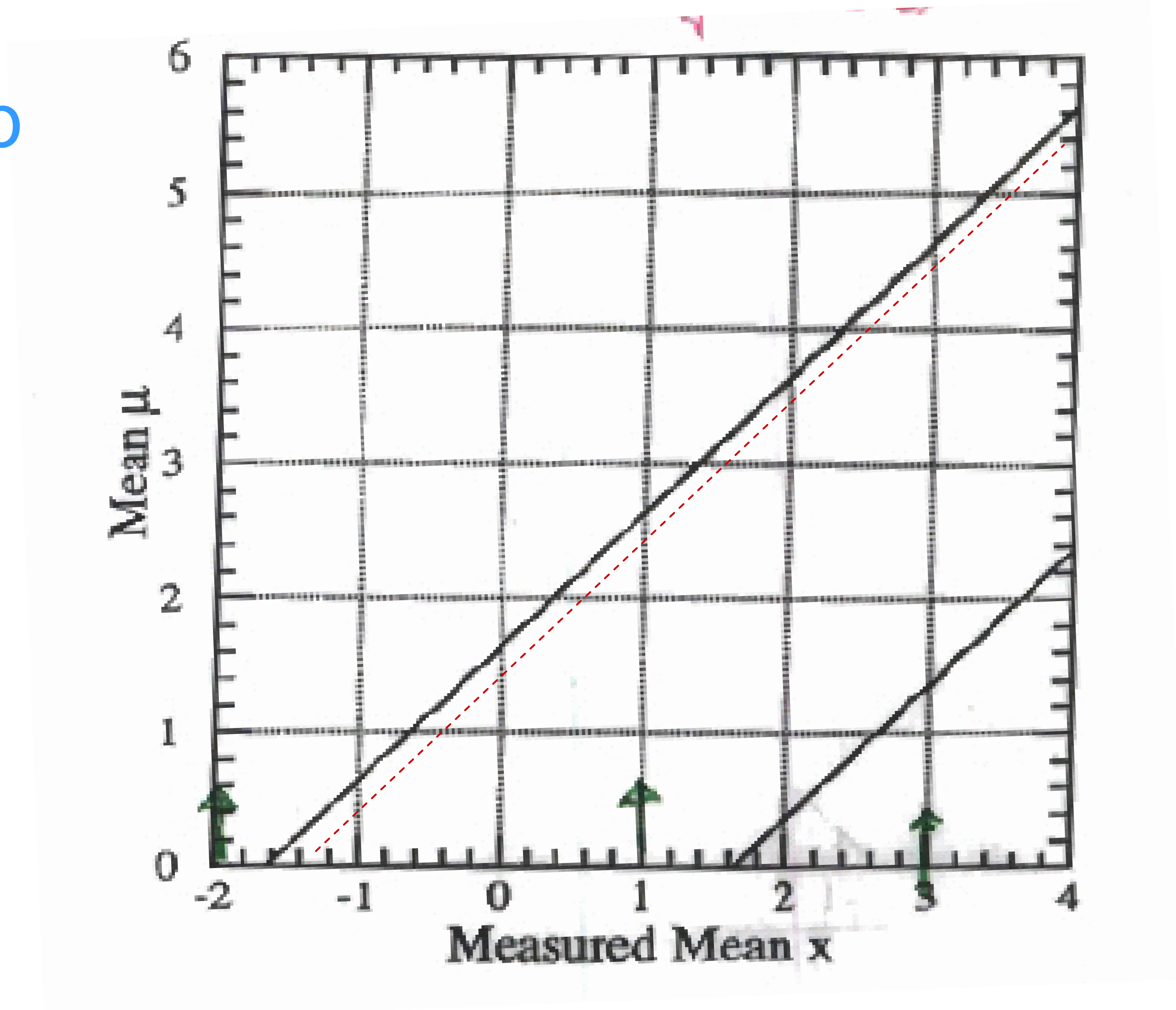



FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

$X_{obs}$ = -2 now gives upper limit

# Flip-flop




Black lines Classical 90% central interval

Red dashed: Classical 90% upper limit

FLIP – FLOP

90% upper limit for $x_{obs} \leq 3$

90% 2-sided interval for $x_{obs} > 3$

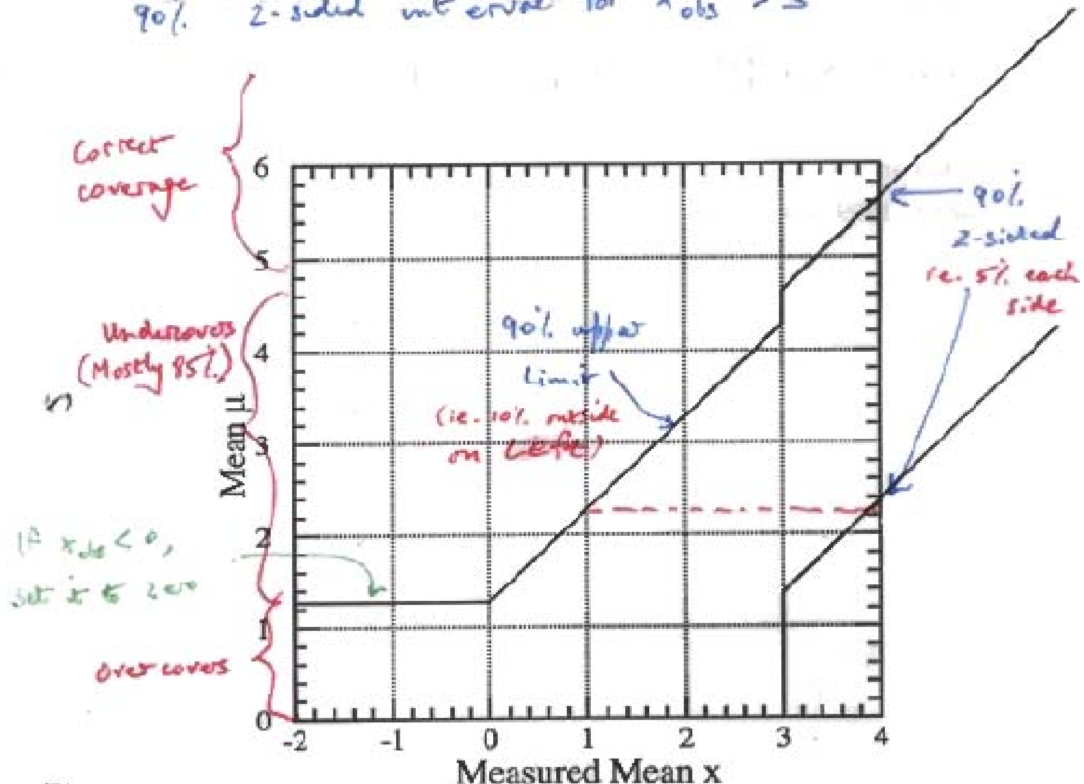



FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence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

Not good to let $x_{obs}$ determine how result will be presented

F-C goes smoothly from 1-sided → 2-sided

# Poisson confidence intervals.    Background = 3

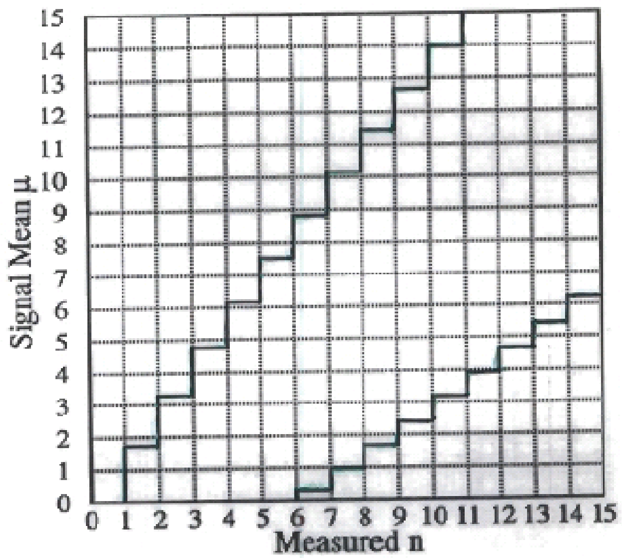



FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$.

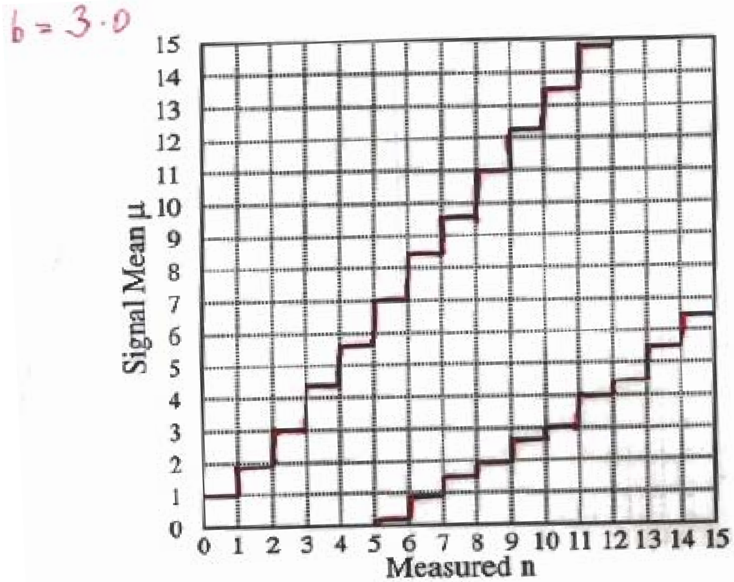



FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$.

Standard Frequentist

Feldman- Cousins

FREQUENTIST POISSON C.B. CONSTRN.

< 10%

< 5%

Prob ordering

TABLES

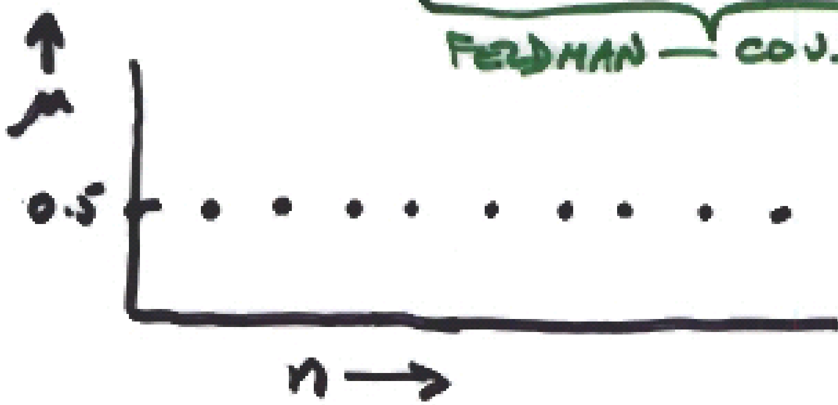
TABLE I. Illustrative calculations in the confidence belt construction for signal mean $\mu$ in the presence of known mean background $b = 3.0$. Here we find the acceptance interval for $\mu = 0.5$.

| $n$ | $P(n\|\mu)$ | $\mu_{best}$ | $P(n\|\mu_{best})$ | $R$ | rank | U.L. | central |
|---|---|---|---|---|---|---|---|
| 0 | 0.030 | 0. | 0.050 | 0.607 | 6 | | |
| 1 | 0.106 | 0. | 0.149 | 0.708 | 5 | ✓ | ✓ |
| 2 | 0.185 | 0. | 0.224 | 0.826 | 3 | ✓ | ✓ |
| 3 | 0.216 | 0. | 0.224 | 0.963 | 2 | ✓ | ✓ |
| 4 | 0.189 | 1. | 0.195 | 0.966 | 1 | ✓ | ✓ |
| 5 | 0.132 | 2. | 0.175 | 0.753 | 4 | ✓ | ✓ |
| 6 | 0.077 | 3. | 0.161 | 0.480 | 7 | ✓ | ✓ |
| 7 | 0.039 | 4. | 0.149 | 0.259 | | ✓ | ✓ |
| 8 | 0.017 | 5. | 0.140 | 0.121 | | ✓ | |
| 9 | 0.007 | 6. | 0.132 | 0.050 | | ✓ | |
| 10 | 0.002 | 7. | 0.125 | 0.018 | | ✓ | |
| 11 | 0.001 | 8. | 0.119 | 0.006 | | ✓ | |

5 3 1 2 4 6 7

FELDMAN — COUSINS

< 5%

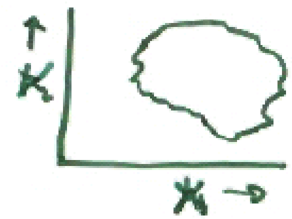


# FEATURES OF F+C

REDUCES EMPTY INTERVALS

UNIFIED 1-SIDED + 2-SIDED INTERVALS
ELIMINATES FLIP-FLOP
NO ARBITRARINESS OF INTERVAL

"READILY" EXTENDS TO SEVERAL DIMENSIONS




LESS OVERCOVERAGE THAN "5% AT ENDS"

MAX PROB DENSITY
5% AT ENDS ?

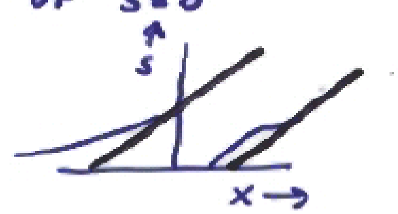
NEYMAN CONSTRUCTION ⇒ CPU-INTENSIVE
(ESP IN SEVERAL DIMENSIONS)

MINOR PATHOLOGIES : DISTANT INTERVALS
WRONG BEHAVIOUR WRT BGD

TIGHT LIMITS FOR $b > n_{obs}$ e.g.

| $n_{obs}$ | bgd | 90% Limit |
|---|---|---|
| 0 | 3.0 | 1.08 |
| 0 | 0 | 2.44 |

UNIFIED ⇒ QUICKER EXCLUSION OF s=0



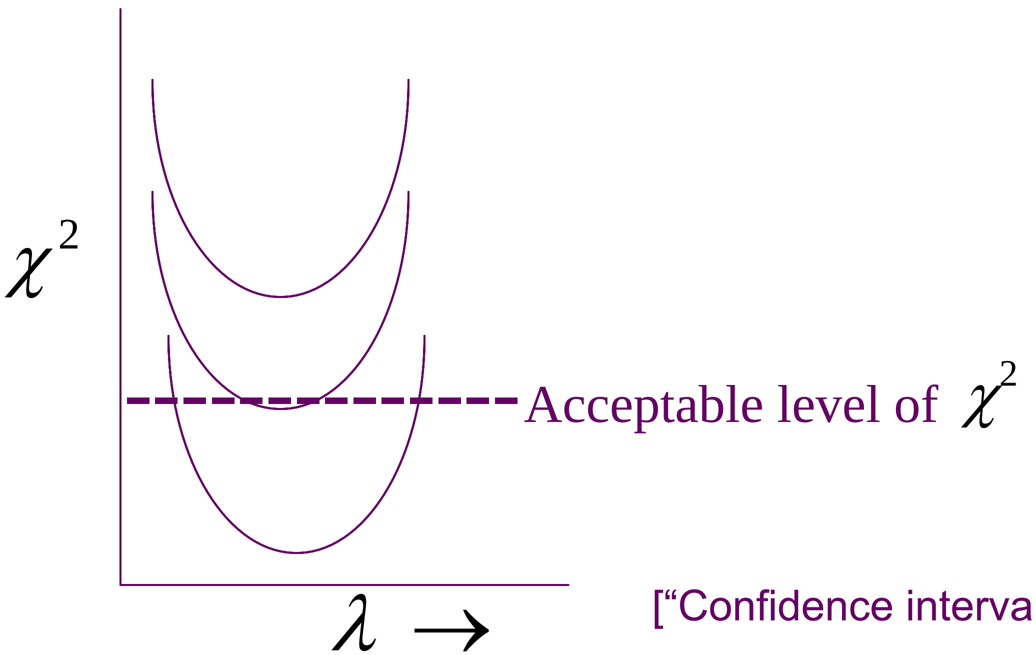
# Importance of Ordering Rule

Neyman construction in 1 parameter $\mu$

2 measurements $\mathbf{X}_1$ $\mathbf{X}_2$

$$p\,(x\,;\mu) = G\,(x - \mu, 1)$$

An aside: Determination of single parameter p via $\chi^2$




Range of parameters given by

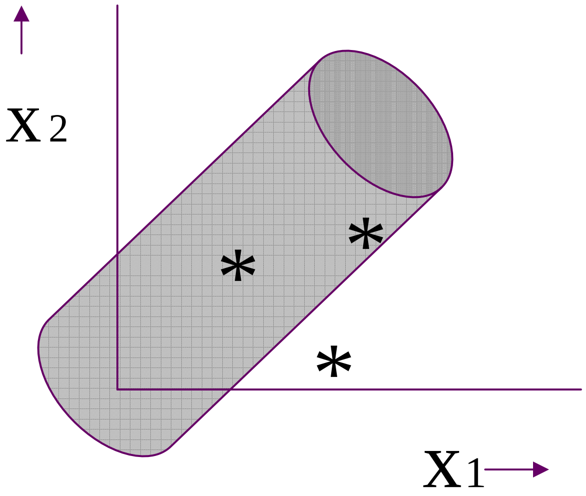
1) Values of $\lambda$ for which data is likely i.e. p($\chi^2$) is acceptable or

2) $\chi^2(\lambda) < \chi^2_{\min}(\lambda) + 1$

2) is good

1) Range depends on $\chi^2_{\min}$

[“Confidence interval coupled to goodness of fit”]




# Neyman Construction

For given $\mu$, acceptable $(x_1, x_2)$ satisfy

$$\chi^2 = (x_1 - \mu)^2 + (x_2 - \mu)^2 \leq Ccut$$

Defines cylinder in $(\mu, x_1, x_2)$ space

Experiment gives $(x_1, x_2) \rightarrow \mu$ interval

Range depends on $|x_1 - x_2|$

$$\mu = \frac{x_1 + x_2}{2} \pm \sqrt{2 - (x_1 - x_2)^2} / 2$$

Range and goodness of fit are coupled

That was using Probability Ordering

Now change to Likelihood Ratio Ordering

For $x_1 \neq x_2$ ,no value of $\mu$ gives very good fit

For Neyman Construction at fixed $\mu$ , compare:

$$(x_1-\mu)^2+(x_2-\mu)^2 \quad \text{with} \quad (x_1-\mu_{\text{best}})^2+(x_2-\mu_{best})^2$$

where $\mu_{\text{best}} = (x_1 + x_2)/2$

giving
$$2\left[\mu^2-\mu(x_1+x_2)+\frac{1}{4}(x_1+x_2)^2\right] = 2\left[\mu-\frac{1}{2}(x_1+x_2)\right]^2$$

Cutting on Likelihood Ratio Ordering gives:

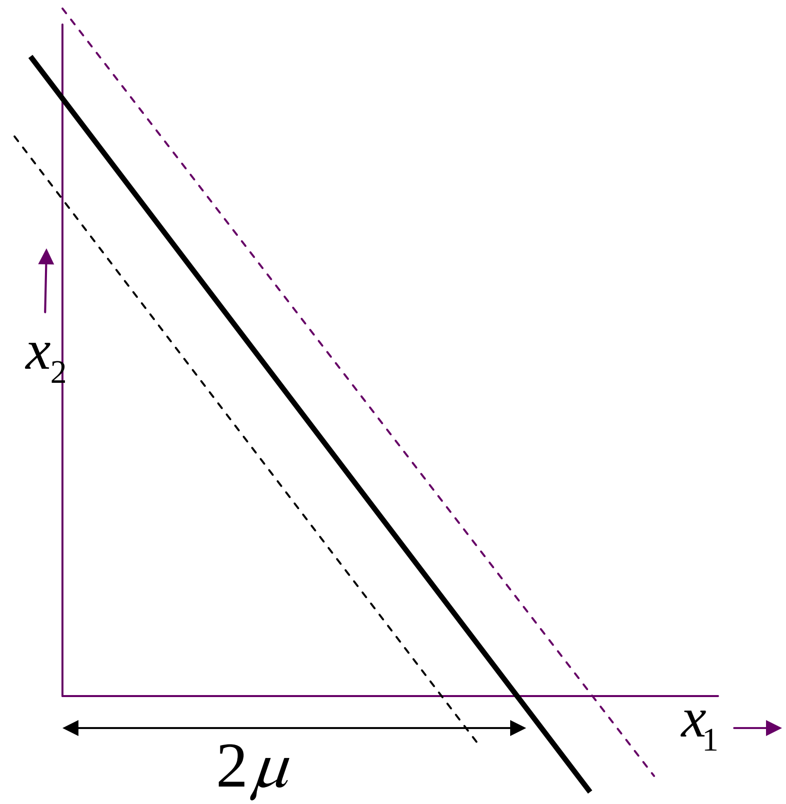
$$\mu = \frac{x_1+x_2}{2} \pm \sqrt{\frac{C}{2}}$$

$$\mu = \frac{x_1 + x_2}{2} \pm \sqrt{\frac{C}{2}}$$

Therefore, range of $\mu$ is

Constant Width

Independent of $x_1 - x_2$




Confidence Range and Goodness of Fit are completely decoupled

# Bayesian

<u>Pros:</u>

Easy to understand

Physical Interval

<u>Cons:</u>

Needs prior

Hard to combine

Coverage

# Standard Frequentist

Pros:

Coverage

Cons:

Hard to understand

Small or Empty Intervals

Different Upper Limits

# SYSTEMATICS

For example

$$N_{events} = \sigma \, LA + b$$

$\uparrow$ Observed $\uparrow$ $N \pm \sqrt{N}$ for statistical errors

$\uparrow$ Physics parameter

$\uparrow$ $\uparrow$ we need to know these, probably from other measurements (and/or theory)

Uncertainties →error in $\sigma$

Some are arguably statistical errors

$$LA = LA_0 \pm \sigma_{LA}$$
$$b = b_0 \pm \sigma_b$$

Shift Central Value

Bayesian

Frequentist

Mixed

## Shift Nuisance Parameters

$$N_{events} = \sigma \ LA + b$$

Simplest Method

Evaluate $\sigma_0$ using $LA_0$ and $b_0$

Move nuisance parameters (one at a time) by their errors → $\delta\sigma_{LA}$ & $\delta\sigma_{b}$

If nuisance parameters are uncorrelated,

combine these contributions in quadrature

→ total systematic

# Bayesian

Without systematics

$$p(\sigma; N) \propto p(N; \sigma)\Pi(\sigma)$$

prior

With systematics

$$p(\sigma, LA, b; N) \propto p(N; \sigma, LA, b)\Pi(\sigma, LA, b)$$

$$\sim \Pi_1(\sigma)\Pi_2(LA)\Pi_3(b)$$

Then integrate over LA and b

$$p(\sigma; N) = \iint p(\sigma, LA, b; N)\, dLA\, db$$

$$p(\sigma; N) = \iint p(\sigma, LA, b; N)\, dLA\ db$$

If $\Pi_1(\sigma)$ = constant and $\Pi_2(\mathrm{LA})$ = truncated Gaussian TROUBLE!

Upper limit on $\sigma$ from $\int p(\sigma; N)\, \mathrm{d}\sigma$

Significance from likelihood ratio for $\sigma = 0$ and $\sigma_{\max}$

# Frequentist

Full Method

Imagine just 2 parameters $\sigma$ and LA

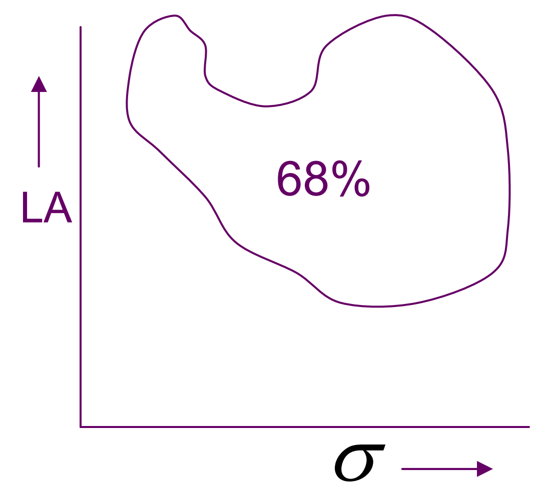
and 2 measurements N and M

$\uparrow$ Physics $\uparrow$ Nuisance

Do Neyman construction in 4-D

Use observed N and M, to give

Confidence Region for LA and $\sigma$



Then project onto $\sigma$ axis

This results in OVERCOVERAGE

Aim to get better shaped region, by suitable choice of ordering rule

Example: Profile likelihood ordering

$$\frac{\mathrm{L}\left(N_0 M_0; \sigma, LA_{best}(\sigma)\right)}{\mathrm{L}\left(N_0 M_0; \sigma_{best}, LA_{best}(\sigma)\right)}$$

## Full frequentist method hard to apply in several dimensions

Used in $\leq$ 3 parameters

For example: Neutrino oscillations (CHOOZ)

$$\sin^2 2\theta \, , \Delta \mathrm{m}^2$$

Normalisation of data

Use approximate frequentist methods that reduce dimensions to just physics parameters

e.g. Profile pdf

i.e. $pdf_{profile}\left(N;\sigma\right) = pdf\left(N, M_0; \sigma, LA_{best}\right)$

Contrast Bayes marginalisation

Distinguish "profile ordering"



See Giovanni Punzi, PHYSTAT05 page 88

Talks at FNAL CONFIDENCE LIMITS WORKSHOP

(March 2000) by:

Gary Feldman

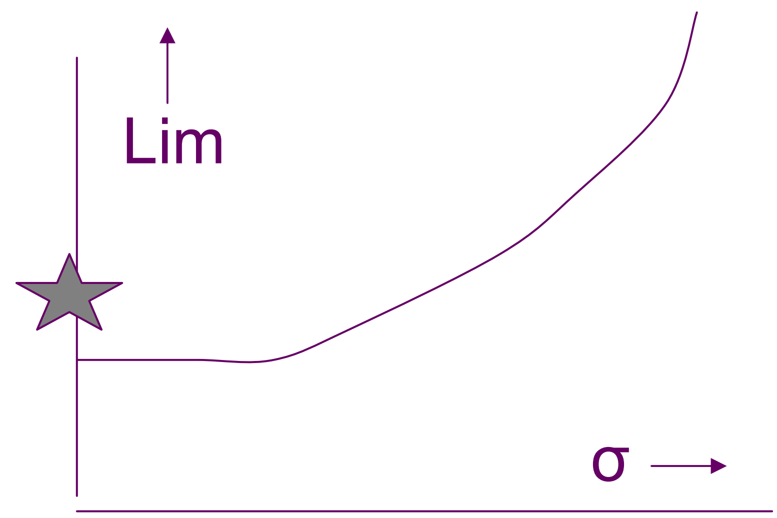
Wolfgang Rolke hep-ph/0005187 version 2

Acceptance uncertainty worse than Background uncertainty

Limit of C. Lim. as $\sigma \rightarrow 0$

$\neq$ C.L. for $\sigma = 0$

Need to check Coverage



## Method: Mixed Frequentist - Bayesian

Bayesian for nuisance parameters and

Frequentist to extract range

Philosophical/aesthetic problems?

Highland and Cousins

(Motivation was paradoxical behaviour of Poisson limit when LA not known exactly)

# Bayesian versus Frequentism

| | Bayesian | Frequentist |
|---|---|---|
| Basis of method | Bayes Theorem --> Posterior probability distribution | Uses pdf for data, for fixed parameters |
| Meaning of probability | Degree of belief | Frequentist definition |
| Prob of parameters? | Yes | Anathema |
| Needs prior? | Yes | No |
| Choice of interval? | Yes | Yes (except F+C) |
| Data considered | Only data you have | ….+ other possible data |
| Likelihood principle? | Yes | No |

# Bayesian versus Frequentism

| | Bayesian | Frequentist |
|---|---|---|
| Ensemble of experiment | No | Yes (but often not explicit) |
| Final statement | Posterior probability distribution | Parameter values → Data is likely |
| Unphysical/ empty ranges | Excluded by prior | Can occur |
| Systematics | Integrate over prior | Extend dimensionality of frequentist construction |
| Coverage | Unimportant | Built-in |
| Decision making | Yes (uses cost function) | Not useful |

# Bayesianism versus Frequentism

"Bayesians address the question everyone is interested in, by using assumptions no-one believes"

"Frequentists use impeccable logic to deal with an issue of no interest to anyone"

Next time : Discovery and p-values

Hope: LHC moves us from era of ‘Upper Limits’ to that of

DISCOVERY