# BAYES v FREQUENTISM

## Return of an old Controversy

LOUIS LYONS

Oxford

Bayes v Frequentism

Upper Limits

- Workshops
- Comparison of methods
- Feldman + Cousins
- CLs

Where are we now?

HOW CAN TEXT-BOOKS NOT EVEN MENTION BAYES / FREQUENTISM ?

FOR SIMPLEST CASE ($m \pm \sigma$) ← Gaus

WITH NO CONSTRAINT ON $m_{true}$,

$$m - k\sigma < m_{true} < m + k\sigma$$

at some prob

FOR BOTH

(BUT DIFFERENT INTERPRETATIONS)

See :

Bob Cousins "Why isn't every physicist a B... ?" A J P. (3/1995)

NEED TO MAKE STATEMENT ABOUT param, GIVEN data

BASIC DIFFERENCE

BAYES:

Prob(param | data)

Anathema to Frequentist

FREQUENTIST

Prob(data | param)

Likelihood Function

# BAYES v CLASSICAL

## 1) BAYES

P(A and B) = P(A|B) × P(B) = P(B|A) × P

e.g. A = event contains t quark

B = event contains W boson

or A = you are in FNAL

B = you are at Workshop

Completely uncontroversial

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

BAYES :

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

P(Hypothesis | Data) ∝ P(Data | Hypothesis) × P

↑ Posterior ↑ Likelihood Pr

PROBLEMS :

1) P( Hyp ...... ) True or false

"Degree of belief"

<u>Credible</u> interval

2) Prior : What functional form ?

Uninformative prior :

Flat ? In which

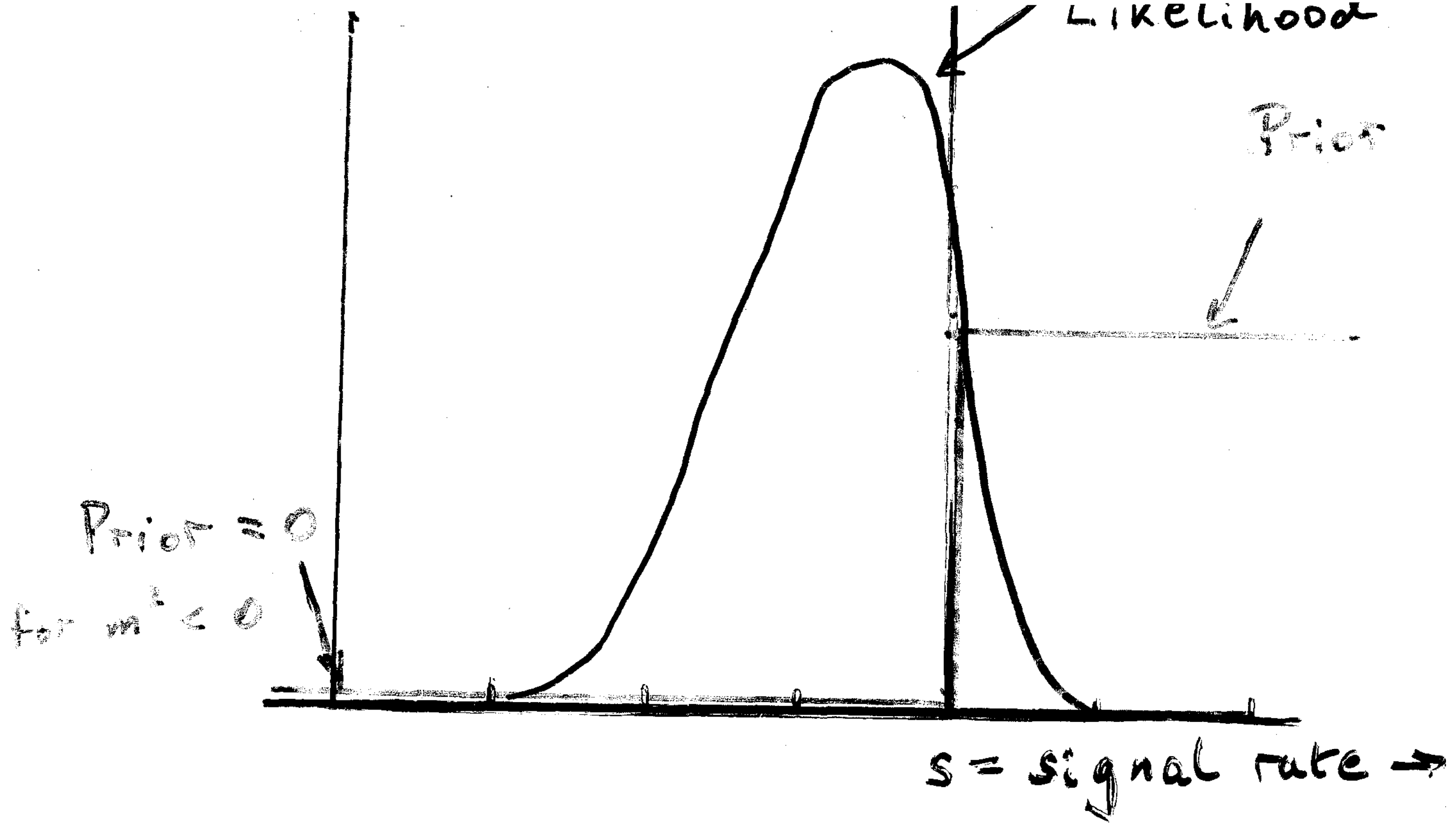
e.g. $m$ , $m^2$, $\ln m$

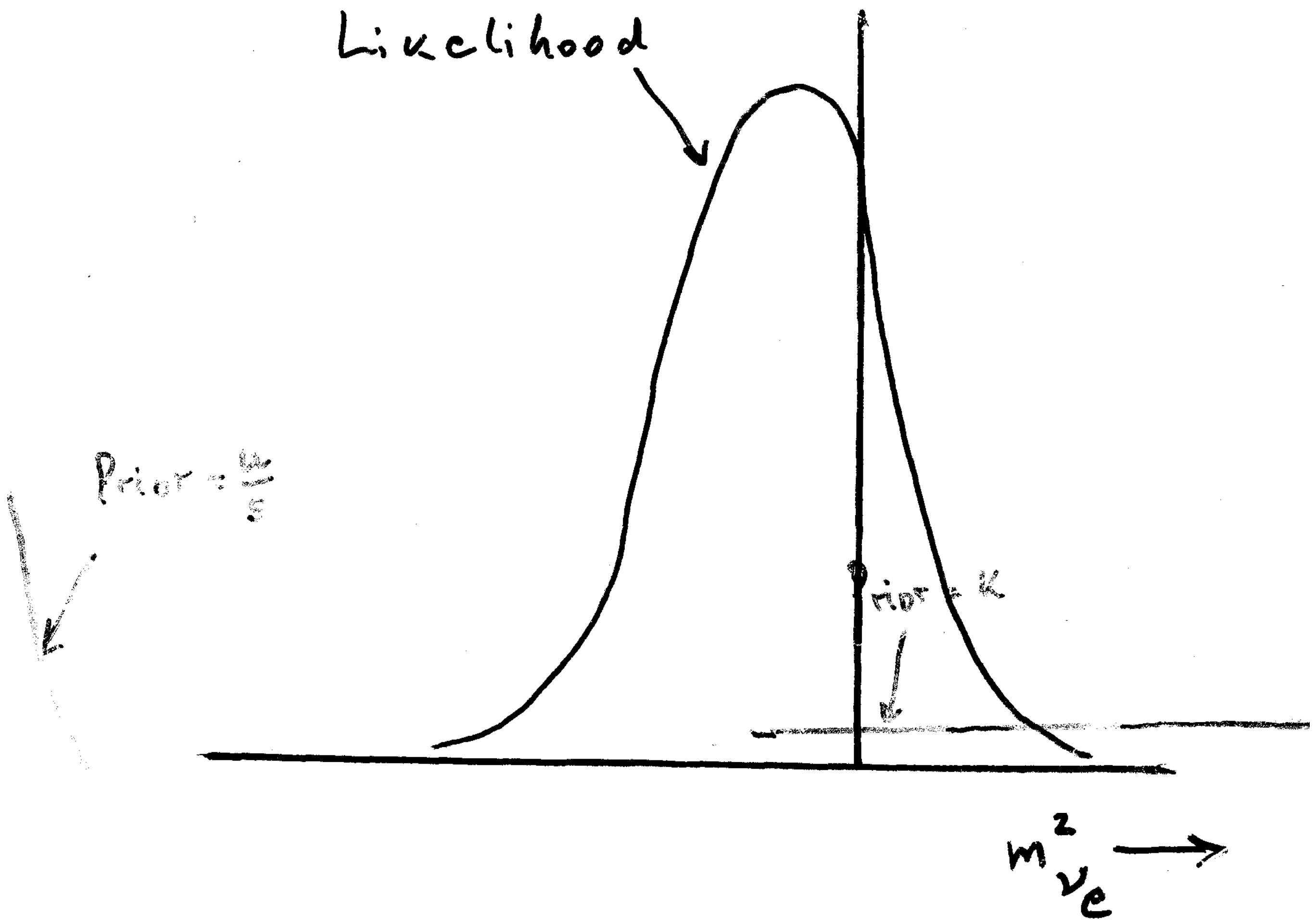
Unimportant if "data overshadows pric

Important for limits

Subjective or objective prior ?

[ 3) Coverage Li


Likelihood
Prior
Prior = 0
for $m^2 < 0$
s = signal rate →


Likelihood
$m^2_{\nu_e}$ →

$$P(\text{data} \mid \text{theory}) \neq P(\text{theory} \mid \text{data})$$

Theory = male or female

Data = pregnant or not pregnant

$$P(\text{pregnant} \mid \text{female}) \approx 2\%$$

BUT

$$P(\text{female} \mid \text{pregnant}) \ggg 2\%$$

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

HIGGS SEARCH AT CERN

Is data consistent with S.M.

or with S.M + Higgs?

End Sept: Data not very consistent with S.

Prob (Data | S.M.) < 1% Frequentist statemen

Turned by Press into

Prob (S.M. | data) < 1%

and hence

Prob (Higgs | data) > 99%

i.e. "IT IS ALMOST CERTAIN THAT THE HIGGS HAS BEEN SEEN"

Example 1 : Is coin fair ?

Toss coin : 5 consecutive tails

What is

$P$ ( unbiassed / data ) ? i.e. $p =$

Depends on Prior ($p$)

If village priest   prior ~ $\delta(\frac{1}{2})$

stranger in pub   prior ! for 0.5

Cost functi

Example 2 : Particle identification

Try to separate $\pi$ and protons

prob ( $p$ tag | real $p$ ) = 0.95   prob ( $\pi$ tag | real $p$ )

prob ( $p$ tag | real $\pi$ ) = 0.10   prob ( $\pi$ tag | real $\pi$ )

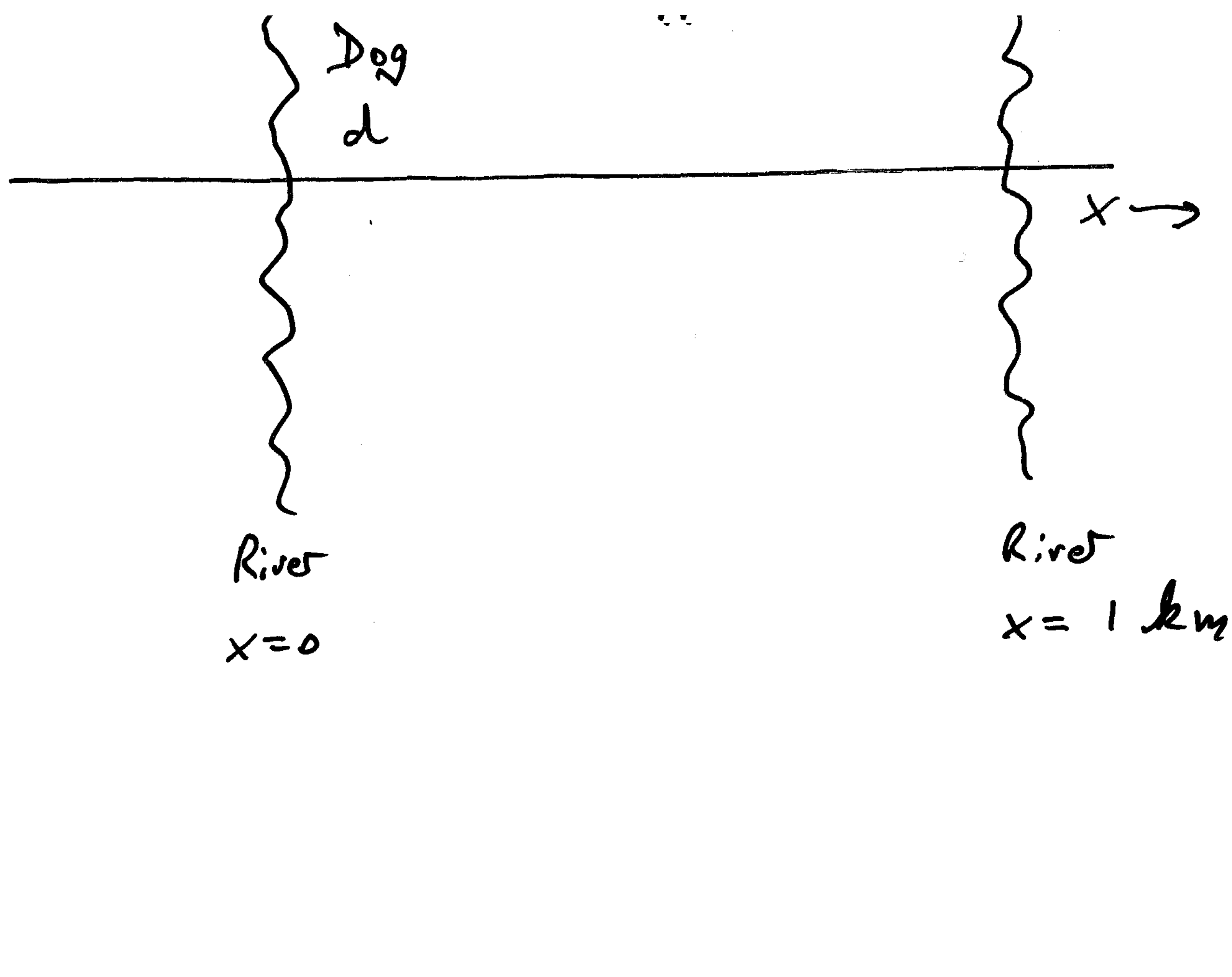
Particle gives proton tag. What is it ?

Depends on prior = fraction of protons.

If proton beam , very likely

If general secondary particles, more even

If pure $\pi$ beam ~ 0.



Dog
d
x
River
x=0
River
x= 1 km

PROB (Data | Theory) $\neq$ Prob (Theory | Data)

## HUNTER AND HIS DOG

Given that:

① "DOG d HAS 50% PROB OF BEING WITHIN 100 m OF HUNTER h"

is it true that

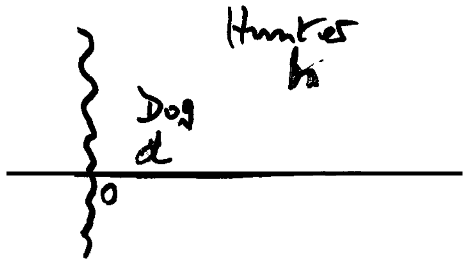
② "HUNTER HAS 50% PROB OF BEING WITHIN 100 m OF DOG" ?

Additional information:

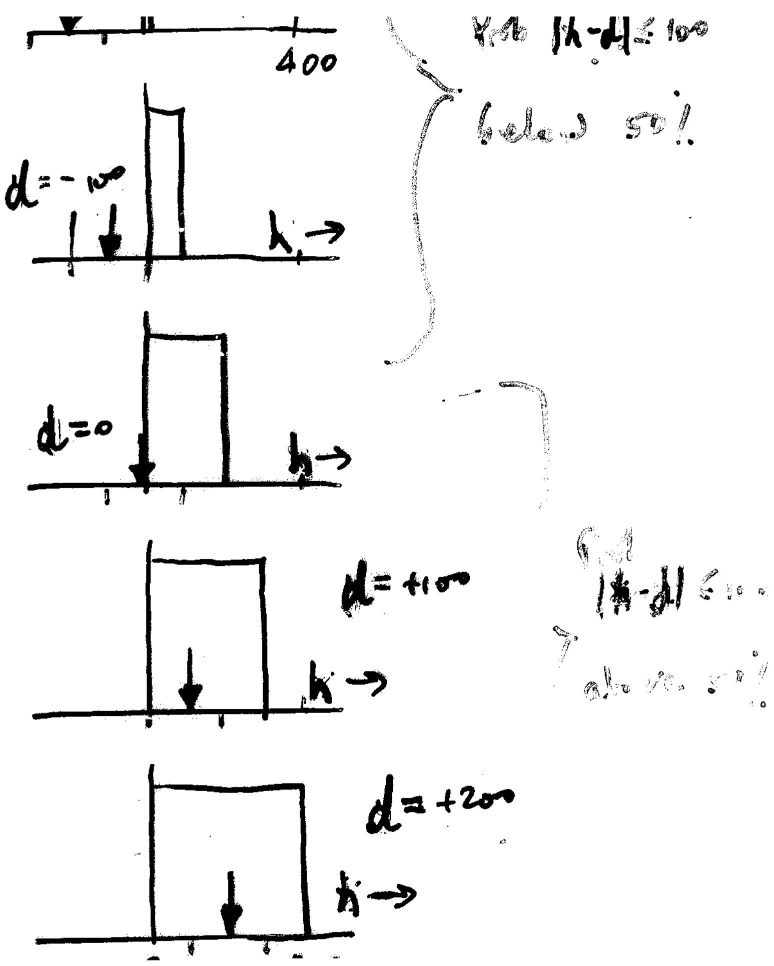
RIVER AT ZERO & 1 KM, HUNTER CANNOT CROSS THEM: $0 \leq h \leq 1$ km

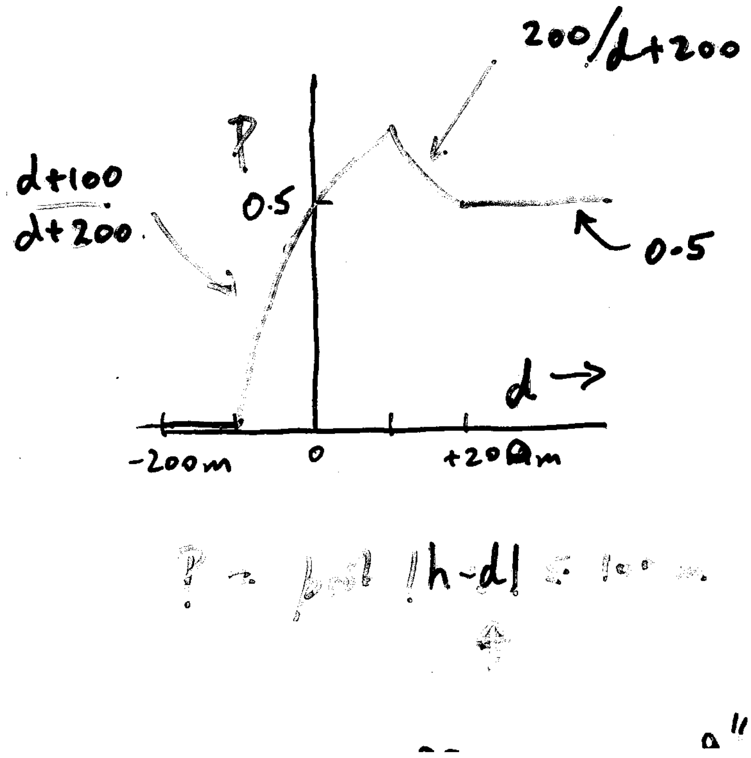
DOG CAN SWIM ACROSS RIVER

STATEMENT (2) STILL TRUE




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


400
|h-d| ≤ 100
below 50%!
d = -100
h →
d = 0
h →
d = +100
h →
d = +200
h →
|h-d| ≤ 100
above 50%!


200/d+200
P
0.5
d+100
d+200
0.5
d →
-200m
0
+200m
P = prob |h-d| ≤ 100 m.

# CLASSICAL

Neyman "confidence interval" avoids pdf

Keeps to $P(x|\mu)$

Confidence interval $\mu_1 \to \mu_2$ :

$P(\mu_1 \to \mu_2 \text{ contains } \mu) = \alpha$ , True for

↑ ↑ (under $\mu_1$, $\mu_2$): **Varying intervals from ensemble of expts**

↑ (under $\mu$): **Fixed**

Gives range of $\mu$ for which observed val $x_0$ was "likely"

[Contrast Bayes : Degree of belief that $\mu$ is in $\mu_1 \to \mu_2$ is $\alpha$ ]

**If true for any $\mu$** — **"correct cove**

**$P<\alpha$ for some $\mu$** — **"undercoverag** SERIOUS

**$P>\alpha$** — **"overcoverage**
**"Conservative**

CLASSICAL INTERVALS : PROBLEMS

Hard to understand (d'Agostini e.n

~ Arbitrary choice of interval

Possibility of empty range

(Overcoverage for integer observation e.g. no. of events

Nuisance parameters (systematic errors)

ADVANTAGES

Widely applicable

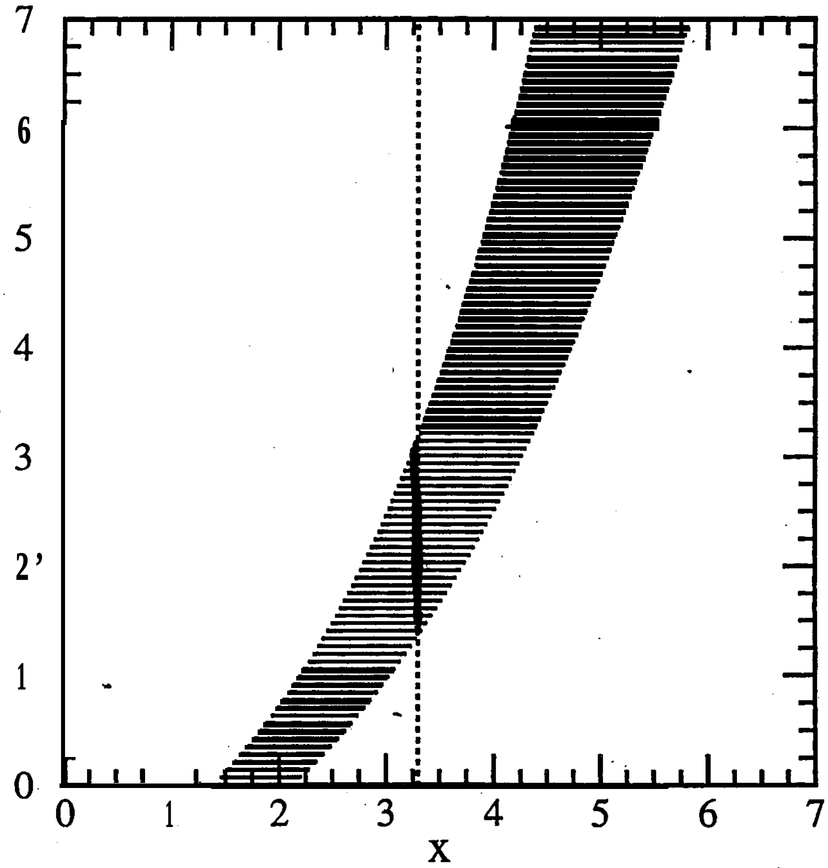
Well defined coverage

CLASSICAL (NEYMAN) CONFIDENCE INTERVALS

Uses only P(data | theory)

FIGURES

Theoretical parameter μ ↑




Observation ⟶

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

Probability / credible statements about $\mu$

(Non) - Coverage = Error of 1st kind

Power : Error of 2nd kind

Where does this leave us?

Difficulties when

small or unobserved signal

bgd not well known

measurement near/outside physical bound

observed events < expected bgd

∴ WORKSHOPS

# CONFIDENCE LIMITS WORKSHOPS

CERN, January 2000

FNAL, March 2000

CERN: Yellow Report 2000-005

+ Website

http://cern.web.cern.ch/CERN/Divisions/EP/Events/CLW/

FNAL: Website

http://conferences.fnal.gov/cl2k/

# Workshop on 'Confidence Limits'

**17-18 January, 2000**
**CERN Council Chamber**

This workshop will be devoted to the problem of setting confidence limits in difficult cases: small or unobserved signal, background larger than signal, background not well known, and measurements near a physical boundary. Among the many examples in HEP are the Higgs mass, accelerator searches for neutrino oscillations, B_s mixing, SUSY, compositeness, neutrino masses, and dark matter. Several different methods are on the market - the CL_s methods used by the LEP Higgs searches; Bayesian; Feldman-Cousins and modifications thereof; etc.

Distinguished invited speakers will present the methods currently being proposed, and all participants are encouraged to submit abstracts for short contributed papers, of which a limited number will be selected for presentation. Proceedings will include summaries of all papers presented, as well as discussions.

| Invited Speakers | Co-convenors |
|---|---|
| Peter Clifford / Oxford | Fred James / CERN |
| Bob Cousins / UCLA | Louis Lyons / Oxford |
| Giulio D'Agostini / Roma | |
| Carlo Giunti / Torino | |
| Fred James / CERN | |
| Harrison Prosper / Florida State | **Local Organization** |
| Alex Read / Oslo | Fred James |
| Michael Woodroofe / Michigan | Yves Perrin |
| Guenter Zech / Siegen | |

Open to all who register. There is no conference fee. All participants will be expected to be familiar with the methods under discussion by having read the publications on the list of 'required reading'. Useful background material will also be presented in the Academic Training lectures of F. James at CERN, 10-14 January, where confidence limits will be discussed from different points of view.

If you plan to participate, please register by filling the registration form on the CL Workshop website at:

http://www.cern.ch/CERN/Divisions/EP/Events/CLW/

**Proceedings**

We will produce proceedings which will contain summaries of all the invited talks and submited papers, as well as the important points brought up during the discussions, to be published as a CERN Yellow Report.

bulletin

*Dernier délai pour soumission des articles : mardi 12.00 h*
*Les articles du Bulletin se trouvent également sous*
*http://Bulletin.cern.ch/News/*

*Deadline for submission of articles : Tuesday 12.00*
*Bulletin articles can also be foun*
*http://Bulletin.cern.ch/Ne*

Semaine du lundi 17 janvier — no 3/2000 — Week Monday 17 Janu

## *Le CERN affronte le nouveau millénaire avec confiance*

Le Conseil du CERN au sein duquel les représentants des 20 Etats membres de l'Organisation décident du programme scientifique et des ressources financières, a tenu sa 114e session le 17 décembre, sous la présidence de M. Hans C. Eschelbacher (DE).

## *CERN confronts the New Millennium with Confidence*

The CERN Council, where the representatives of the Member States of the Organization decide on scient programmes and financial resources, held its 114th sess on 17 December under the chairmanship of Dr. Ha C. Eschelbacher (DE).

**Rapport du Président**

M. Eschelbacher a ouvert la séance en formulant des remarques sur l'état du CERN à l'orée d'un nouveau millénaire. *"Depuis sa création, le CERN a apporté des contributions remarquables au monde de la science, de la technologie et de l'éducation. Aujourd'hui, à l'aube du nouveau millénaire, je crois que le meilleur est encore à venir. Le CERN est bien préparé à relever les défis de la mondialisation des institutions de recherche. Le trait dominant des prochaines décennies, à l'évidence aussi dans les domaines de la physique des hautes énergies et de l'informatique, sera la concurrence et la coopération mondiales au sein d'alliances stratégiques. Pour réussir dans ce contexte, il*

**President's Report**

Dr Eschelbacher opened the meeting with comments the state of CERN at the opening of a new millenniu *"CERN has from the beginning delivered outstanding achie ments to the scientific, technological and educational world. E now, moving into the next millennium, I believe the best is yet come. CERN is well prepared to master the challenges globalisation for research institutions. What will become do nant in the next decades, obviously also in the field of HEP a computing, is global competition and co-operation within stra gic alliances. To be successful in that global competition one h to be attractive to the best people, one has to run the best facilit*

⇒ "THIS EXPT DISCOVERED NOTHING, SO WE WROTE A PAPER ABOUT LIMITS"

* 90% had read $\geq 50\%$ of required reading material.

DELPHI TGCs (F-C)

$B_s$ mixing (A)

Panel discussion

# FNAL Program

| | | | |
|---|---|---|---|
| Monday | am 1 | James | Choosing foundat. |
| | am 2 | Prosper<br>Berger | Bayes |
| | pm 1 | Linnemann | Bayes |
| | | Kruse<br>Maeshima | Searches |
| | | James | Reply to Bayes |
| | pm 2 | Murray | CLs |
| | | Raja<br>Narsky | Limits methods |
| Tuesday | am 1 | Cousins<br>Rolke<br>Messier<br>Kafka | F-C |
| | am 2 | Schee<br>Yellin | CDMS |
| | | Roe | Methodology |
| | pm | Feldman<br>Murray<br>Greenlee | Combining resul |
| | | Cowan | CONCLUSIONS |

3

# BAYES TALKS

d'Agostini

Probability for Bayesians

Barlow's book =

Very critical of Classical approach

e.g. What does limit on $M_H$ mean

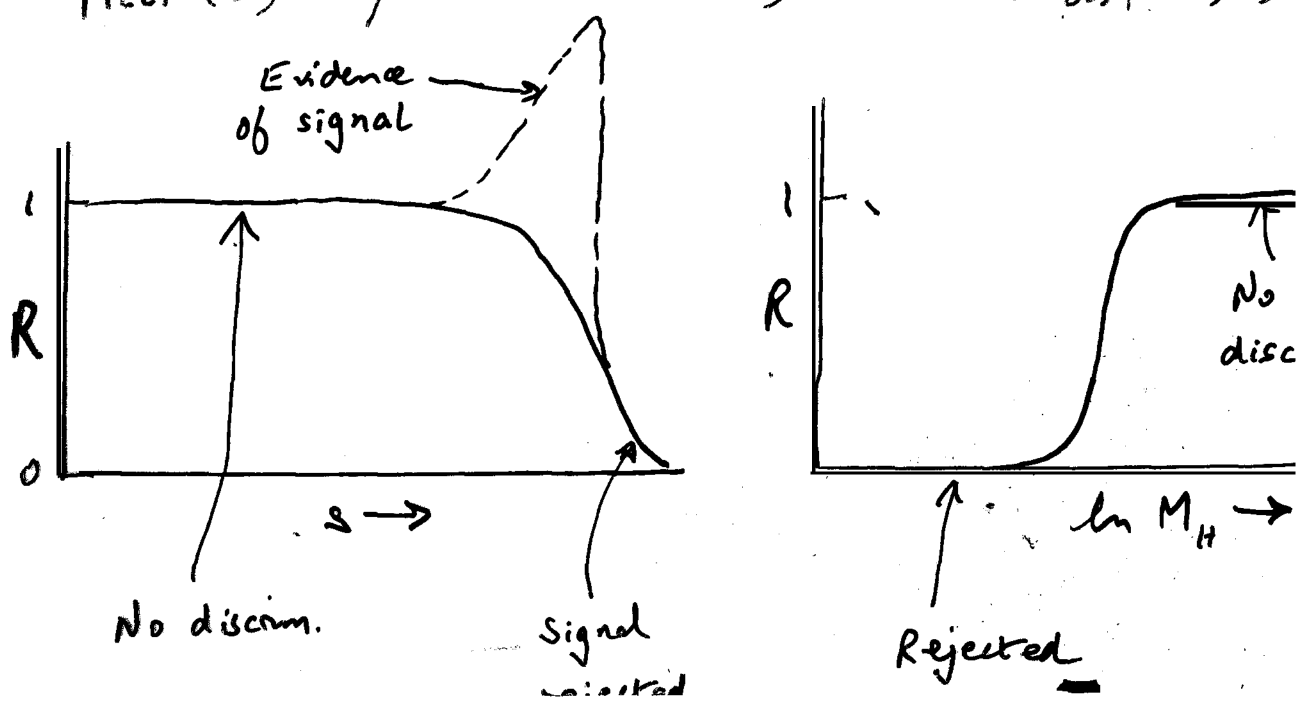
Recently advocates $\mathcal{L}$ ratio for testing hypo

Used for Higgs limit [d'A + Degrassi]

$$\frac{P(s_1 | n_{obs}, b)}{P(s_2 | n_{obs}, b)} = \frac{P(n_{obs} | s_1, b)}{P(n_{obs} | s_2, b)} \times \frac{Prior(\ldots}{Prior(\ldots}$$

Rewrite as

$$\frac{P(s | n_{obs}, b)}{Prior(s)} \Big/ \frac{P(s=0 | n_{obs}, b)}{Prior(s=0)} = \frac{P(n_{obs} | s, b)}{P(n_{obs} | s=0, b)}$$



Incorrect treatment
of Binomial : Wrong errors e.g. $\varepsilon = 97 \pm 10\%$

√

Value = physical

Value ± error = unphysical

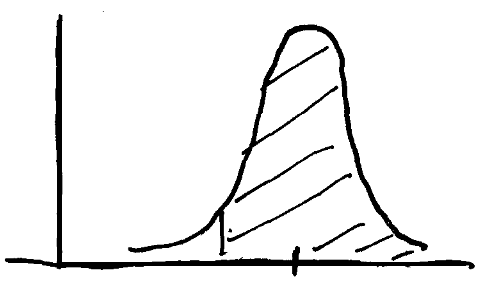
# LIMITS

Observe { zero / very few / < 0 } events ⇒ set limit
rather than give v
(but see Feldma
Cousi

Limit less useful than value ± error
↑ 1 number ⇈ 2 numbers

Cannot combine limits
Can combine estimates

Assume Gaussian dist without physical region prob

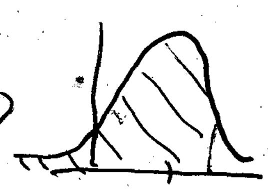

$p_0$

| One sided | 90% | 95% | 99% |
|---|---|---|---|
| | 1.28σ | 1.64σ | 2.33σ |

90% prob that $p_{meas} > p_0 - 1.28$

Bayesian $p_0 < p_{meas} + 1.28$

What to do if { large part / all / some } is unphysical?


$p_{meas}$

UNIFIED APPROACH TO CLASSICAL STATISTICAL ANALYSIS OF SMALL SAM

Gary Feldman & Bob Cousins

HUTP - 97 - A096

Physics 9711021    22 Nov 97

Phys Rev D57 (199

SOLVES

1) PROBLEM OF INCORRECT CONF LE FOR "FLIP-FLOP" APPROACH

2) NON-PHYSICAL CONF INTERVALS / LIM

3) AVOIDS OVERCOVERAGE

EXAMPLES

a) GAUSSIAN WITH PHYSICAL BOUND

b) POISSON COUNTING, WITH BGD

c) $\nu$ OSCILLATIONS

BAYESIAN v CLASSICAL (FREQUENTIST)

CLASSICAL CAN GIVE UNPHYSICAL ANSWER

e.g. $m_\nu^2 < -(20\ e$

MOTIVATED BAYESIAN APPROACH (see e.g. PDG

**HERE, FREEDOM IN CLASSICAL APPRO**

**EXPLOITED ⇒**

**LIMITS ALWAYS PHYSICAL**

**∴ NO NEED FOR BAYES**

**ALSO DECOUPLES { INTERVAL / LIMI**
**GOODNESS OF FI**

90% classical interval for Gaussian

$\sigma = 1$

$\mu \geq 0$

e.g. $m^2(\nu_e)$

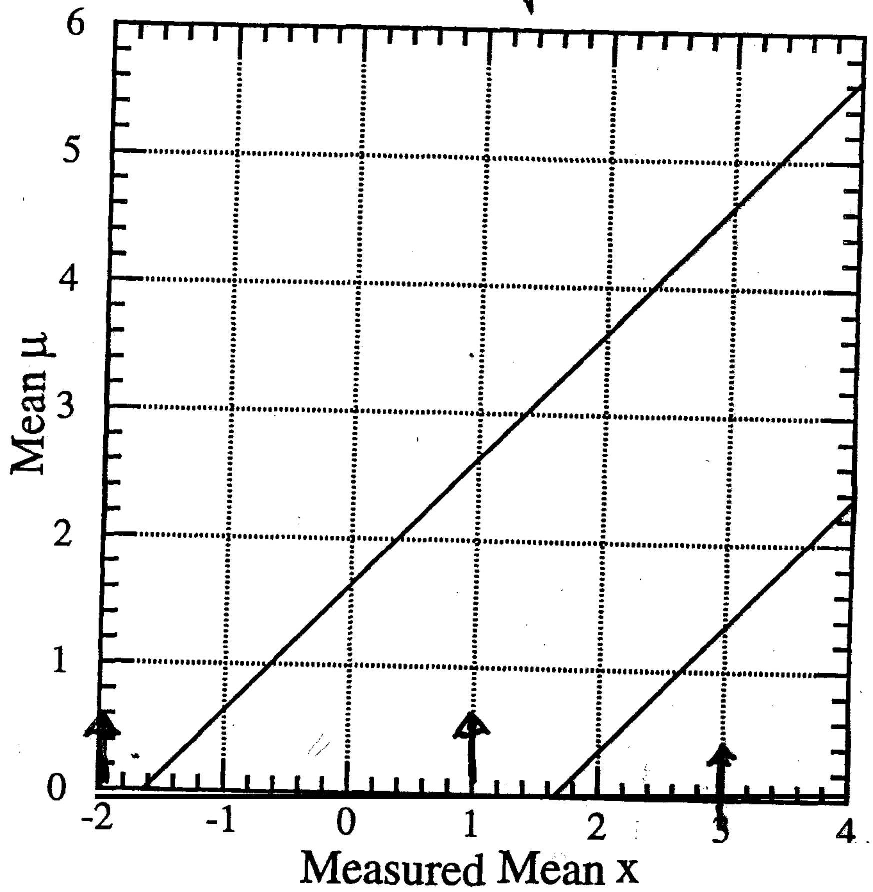



FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$x_{obs} = 3$ Two sided limit

$x_{obs} = 1$ Upper limit

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

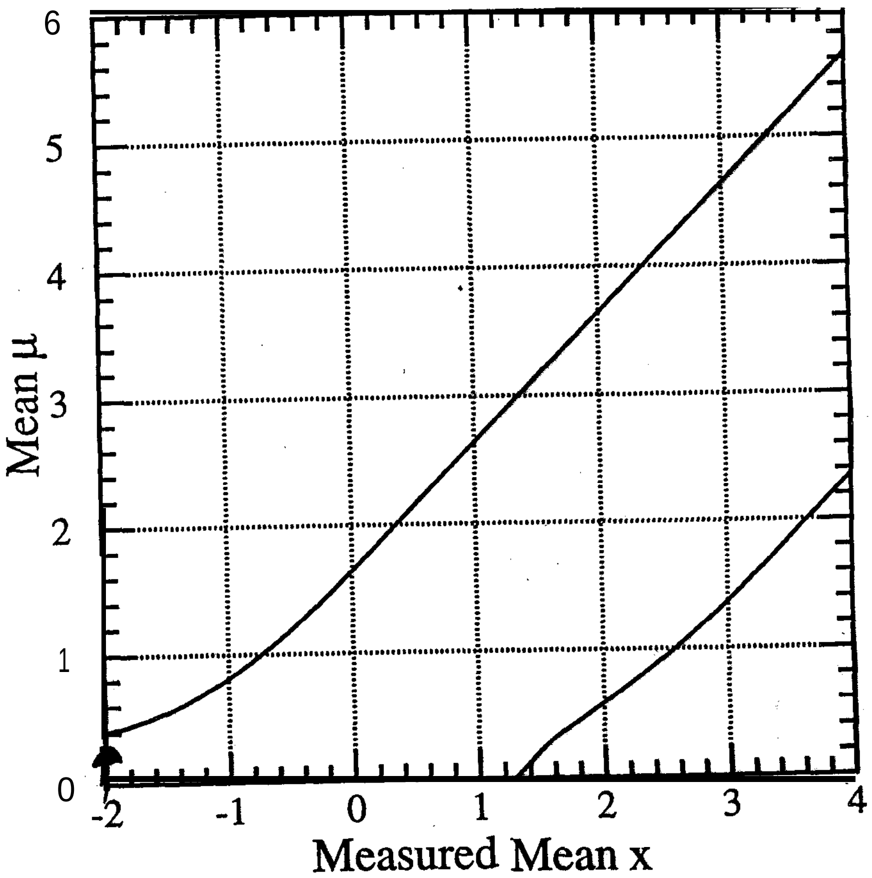



FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian constrained tc non-negative, described in the text.

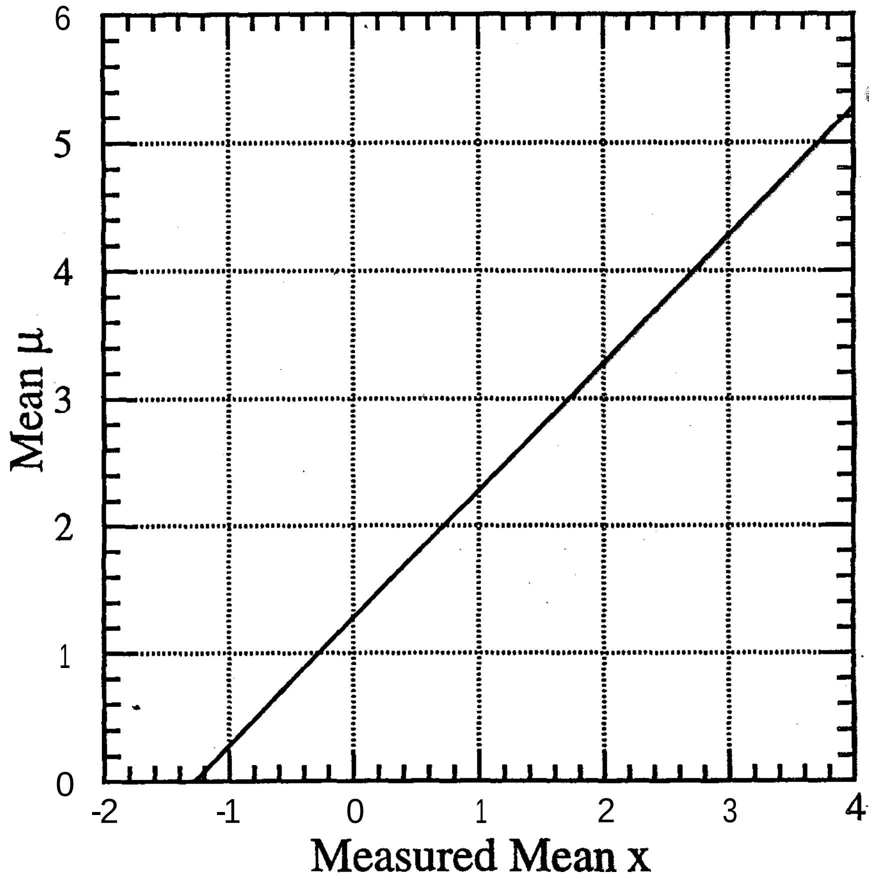



FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, i units of the rms deviation. The second line in the belt is at $x = +\infty$.

FLIP - FLOP

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

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

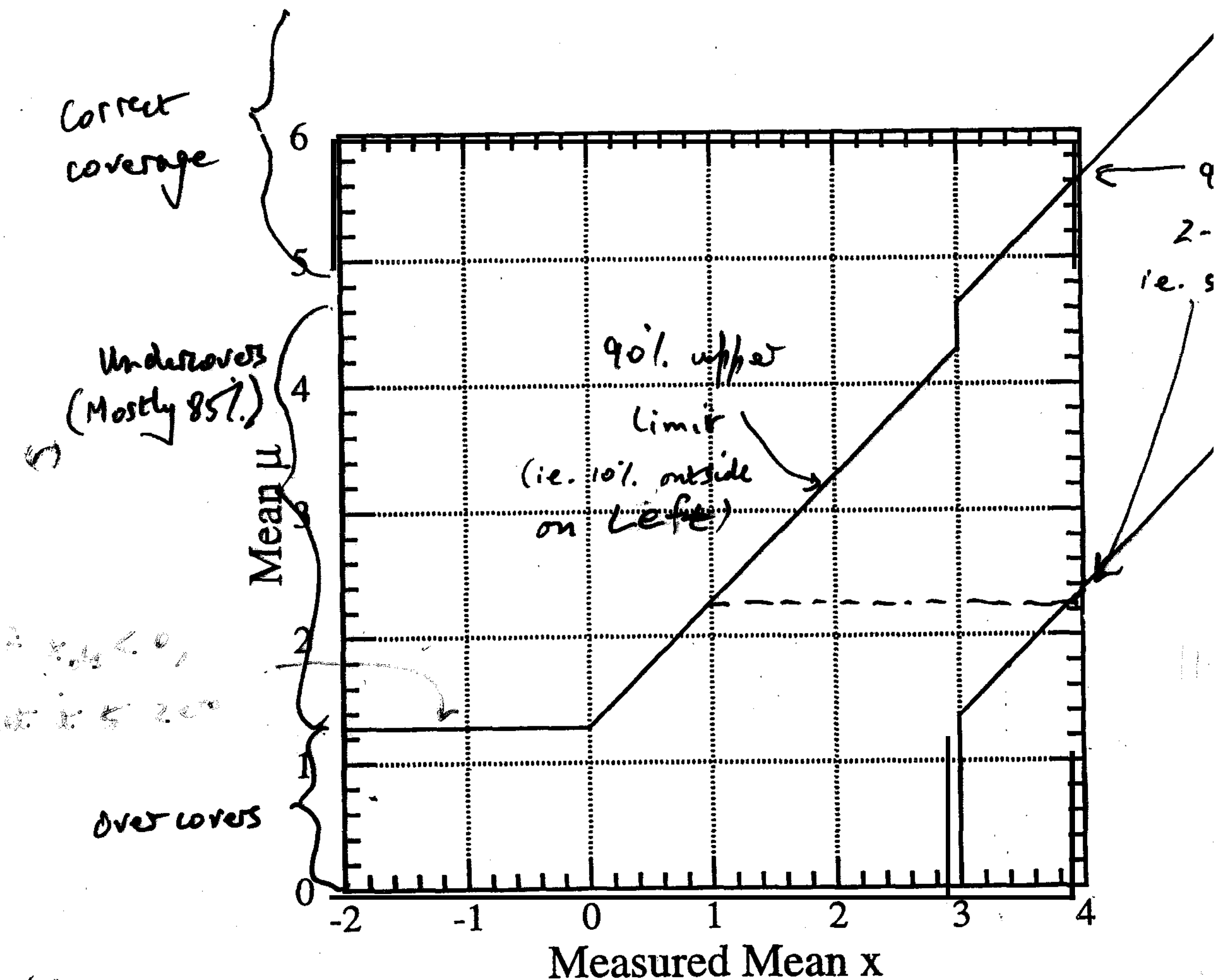



FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (verti intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They : not valid confidence belts, since they can cover the true value at a frequency less than the stai confidence level.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

# ORDERING PRINCIPLE

CHOICE OF HORIZ BAND OF FIG 1 IS ARBI

ORDERING PRINCIPLE

AT ANY $\mu$, CHOSE $x$ VALUES ACCORDIN

ORDER OF
↑
LARGEST

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

← VALUE OF $\mu$ IN RANGE ⇒ MAX P

e.g. GAUSSIAN CASE

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

THEN ADD IN $x$ VALUES (FOR LARGEST $R$) ⇒

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

($x_1, x_2$ numerically)

**Fig 10 : Confidence belt (Cf Fig :**

N.B.

For $x < 1.28$, conf belt ⇒ upper limit

$x = 0$, upper limit is 1.64 (not 1.28)

[Looser limit caused by repairing undercoverage
[Fig. F 2

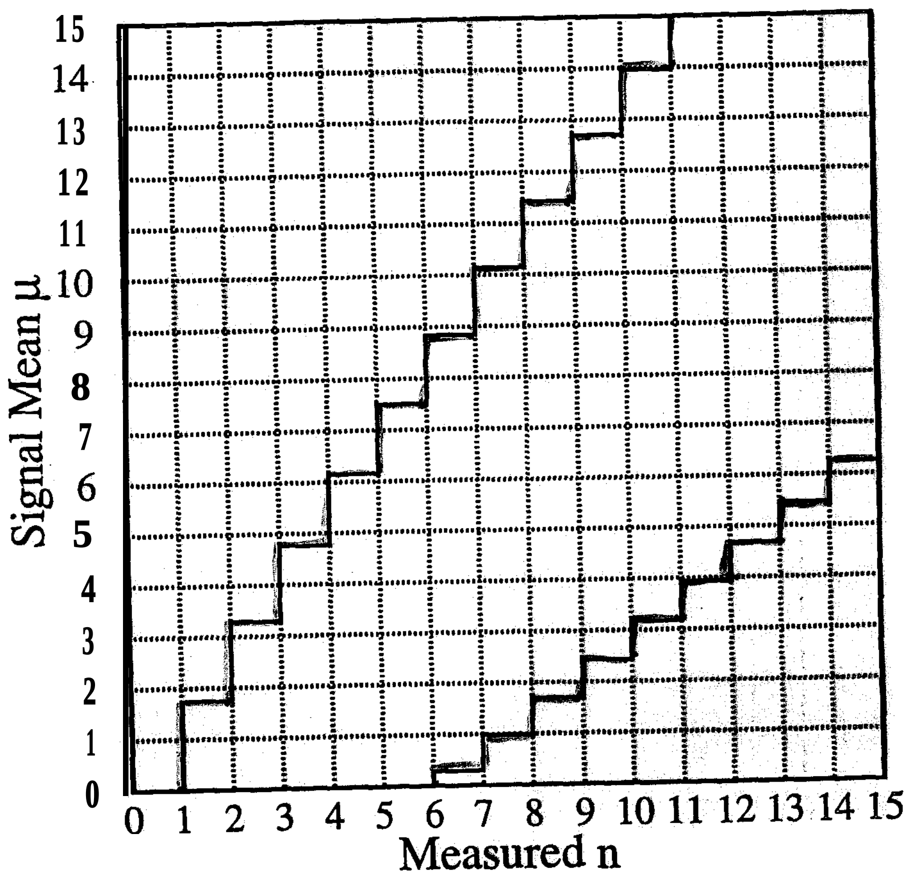



FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poi signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$

Standard frequentist
for Poisson mean $\mu$.

FELDMAN + COUSINS FOR

POISSON MEAN $\mu$

90% Conf

$b = 3.0$

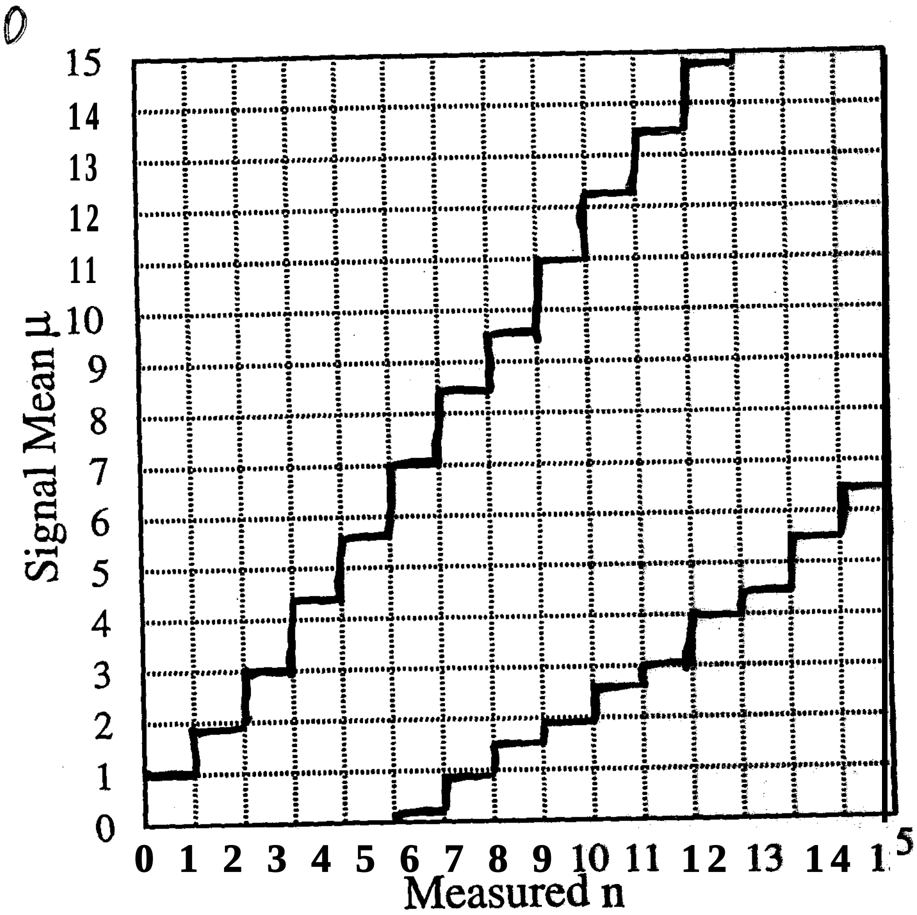



FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence interval unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b$ =

# BAYES

Easy to understand

Physical interval

Needs prior

Hard to combine

Coverage

STANDARD FREQUENTIST

COVERAGE

HARD TO UNDERSTAND

SMALL } INTERVALS
EMPTY }

DIFFERENT UPPER LIMITS

# FELDMAN – COUSINS

UNIFIED

LIFTS OFF FROM ZERO
TOO SOON

NO FLIP-FLOP

COVERAGE

HARD TO UNDERSTAND

DECREASING LIMIT AS BGD > OBS

COMPUTING TIME

# F–C′

Try to overcome problem of limit ↓ as bgd > observ[ed]

F & C were aware, but decided to live with it

Defined "Sensitivity"

Roe & Woodroofe:

Condition on "bgd ≤ $n_{obs}$"

Giunti: Bayes ordering principle

Frequentist intervals

Bouchez: One side Bayes

Other side = frequentist coverage

Mandelkern & Shultz

Punzi

Harrison Prosper

"Freedom to do what you want"

# COMBINING LIMITS

COMBINING **MEASUREMENTS** IS "EASY"

$$x_i \pm \sigma_i \quad \Rightarrow \quad \bar{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

$$\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$$

WHY IS IT IMPOSSIBLE TO COMBINE LIMITS $L_i$?

1) $x_i \pm \sigma_i$ = 2 numbers

$L_i$ = 1 number : Not enough information

2.) Measurement : $\sigma_i^2$ = variance

Limit : No information about distribution

# SUMMARY

Comments : Too polite
Too theoretical

Statisticians : Don't reinvent the wheel!
Quote books after Kendall & Stuart
Useful to explain what we are doing

Lots of detailed improved understanding
Dispel misconceptions
Cousin's 10 points ⇒

Experts found it useful to interact
Going away to think
Problems to attack

1) Nuisance params } in Classical approach
Systematics

2) Combining results
e.g. What do you need for F-C?
[Fred James : What do you want it for?]

Bob Cousins' 10 Commandments

## IX. What Might We Agree On?

Maybe **quite** a **bit!**

1. **First of all, civility. Bohr and Einstein make better role models than Lindley and others in the statistics community.**
2. P(hypothesis | data) cannot be calculated without a prior.
3. **The likelihood function is** not a pdf in the unknown parameters: "**Integrating the** likelihood function" **is not a concept in either Bayesian or classical** statistics. Uniform priors **should be explicitly stated, not hidden.**
4. Answers based on "uniform priors" depend on the metric in which the prior is uniform.
5. The problem of upper limits **cannot be considered solved without having a consistent** method for two-sided intervals.
6. Bayesian intervals typically. do not have frequentist coverage. [FIG]
7. Publishing enough info to reconstruct an approximate **likelihood function should be** strongly **encouraged.**
8. Our usual gof tests do not exist in Bayesian statistics: must be reformulated as extending the space of P.
9. The confidence interval construction does not use a prior.
   It uses P(data | theory), which requires the ensemble to be specified.
   **Priors enter when** going from P(data | theory) to P(theory | data), which confidence intervals do not do.
10. Regardless of your opinion about priors, a subjective utility function is needed to make a. decision, so any argument for totally objective decisions is highly suspicious.

Clarification of many specific points

No "Grand Unified Method" yet

Review for Rev Mod Phys