CERN CONFIDENCE LIMITS WORKSHO

17 + 18 JAN 200

Louis Lyons
Oxford

What is problem?
What is controversy?
Brief summary of talks, discussions
Where does this leave us?

Introduction to FNAL Workshop program

WHAT IS PROBLEM?

Search for: Higgs

SUSY (R or $\not{R}$)

Compositeness

Dark matter

ν oscillations @ accelerators or reactors

β decay

etc

Need to set limits.

Difficulties when

small or unobserved signal

bgd not well known

measurement near/outside physical bound

observed events < expected bgd

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

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 hrs
Bulletin articles can also be found at
http://Bulletin.cern.ch/News/

Semaine du lundi 17 janvier

no 3/2000

Week Monday 17 January

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 20 Member States of the Organization decide on scientific programmes and financial resources, held its 114th session on 17 December under the chairmanship of Dr. Hans 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 on the state of CERN at the opening of a new millennium. *"CERN has from the beginning delivered outstanding achievements to the scientific, technological and educational world. But now, moving into the next millennium, I believe the best is yet to come. CERN is well prepared to master the challenges of globalisation for research institutions. What will become dominant in the next decades, obviously also in the field of HEP and computing, is global competition and co-operation within strategic alliances. To be successful in that global competition one has to be attractive to the best people, one has to run the best facilities*

135 participants* (upper limit of Council Chamber)

2 Statisticians

Video overflow in Room B (Videos exist)

Shame topic is so popular

Fred. Jones:

"THIS EXPT HAD INCONCLUSIVE, SO WE HAD TO USE STATISTICS"

=>

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

* 90% had read ≥ 50% of required reading material.

BAYES v CLASSICAL

1) BAYES

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

e.g. A = event contains t quark

B = event contains W boson

or A = 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.

$$P(\text{Hypothesis} \mid \text{Data}) \propto P(\text{Data} \mid \text{Hypothesis}) \times P(\text{Hyp.})$$

Posterior — Likelihood — Prior

PROBLEMS:

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

"Degree of belief"

<u>Credible</u> interval

2) 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?

3) Coverage

4) Goodness of fit

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

Theory = male or female

Data = pregnant or not pregnant

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

But

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

Example 1 : Is coin fair ?

Toss coin : 5 consecutive tails

What is

$P(\text{unbiassed} \mid \text{data})$? i.e. $p = \frac{1}{2}$

Depends on Prior (p)

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

stronger in p prior ~ 1 for $0 \leq p \leq 1$

Cost function

Example 2 : Particle identification

Try to separate π and protons

$\text{prob}(p \text{ tag} \mid \text{real } p) = 0.95$ $\text{prob}(\pi \text{ tag} \mid \text{real } p) = .05$

$\text{prob}(p \text{ tag} \mid \text{real } \pi) = 0.10$ $\text{prob}(\pi \text{ tag} \mid \text{real } \pi) = 0.90$

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 π beam, ~ 0.

CLASSICAL (NEYMAN) CONFIDENCE INTERVALS

Uses only P (data | theory)

FIGURES

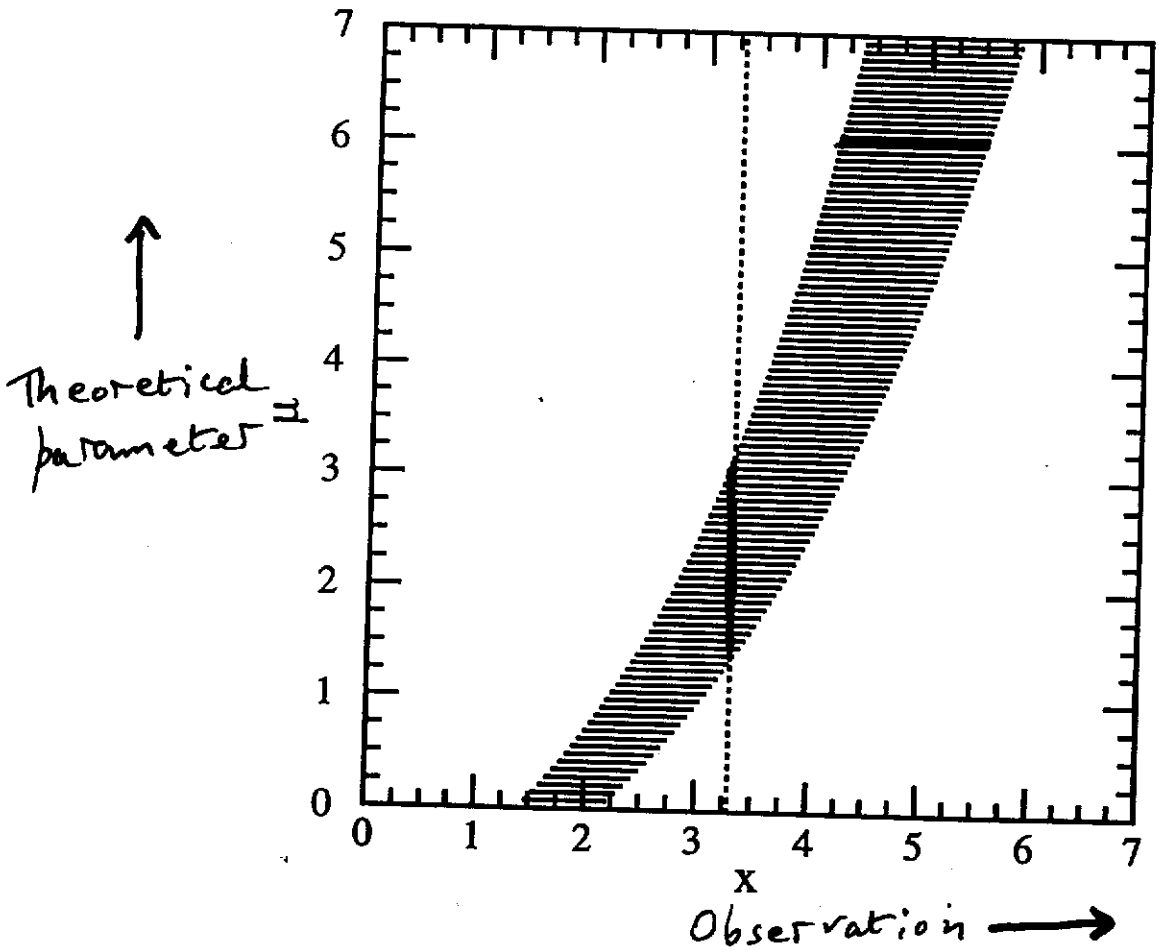
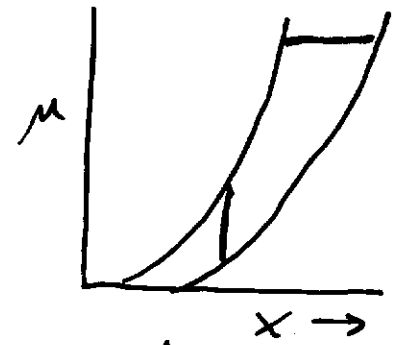


FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] \,|\, \mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through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

NO PRIOR INVOLVED

8

2) CLASSICAL CONFIDENCE INTERVALS



For each possible μ,

Find range of x $\Rightarrow$ 90% conf interval

Generate lines

For observed x, lines give values of μ which have x_{obs} in 90% region

"Values of μ for which x_{obs} is (90%) likely"

"Values of μ not excluded by data"

Statement $\mu_L \leq \mu \leq \mu_U$ $\Rightarrow$ 90% conf

is statement about μ_L & μ_U $\Leftarrow$ Random but known

not about μ $\Leftarrow$ Fixed but unknown

Contrast Bayesian

Difference ~

Classical: $\bar{\mu}$ close to μ

Bayesian: μ close to $\bar{\mu}$

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

HUNTER AND HIS DOG

Given that:

① "DOG y HAS 50% PROB OF BEING WITHIN 100 m OF HUNTER x"

is it true that

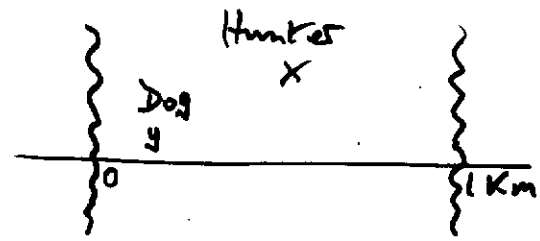
② "HUNTER HAS 50% PROB OF BEING WITHIN 100 m OF DOG" ?

Additional information:

RIVER AT ZERO & 1 Km, HUNTER CANNOT CROSS THEM : $0 \leq x \leq 1$ Km

DOG CAN SWIM ACROSS RIVER

STATEMENT ① STILL TRUE



If dog at −101 m, hunter <u>cannot</u> be within 100 m of dog.

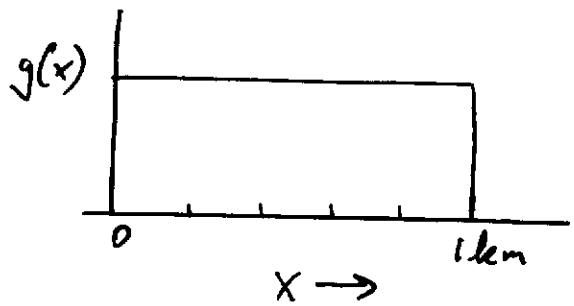
∴ Statement ② untrue.

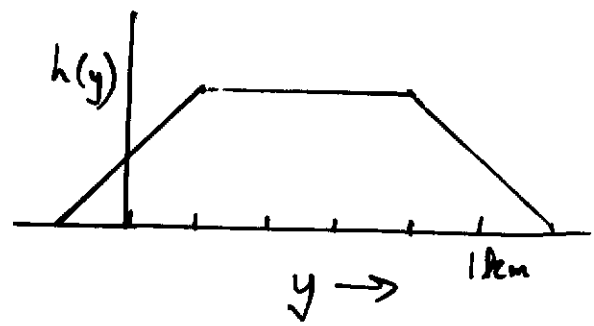
Example:

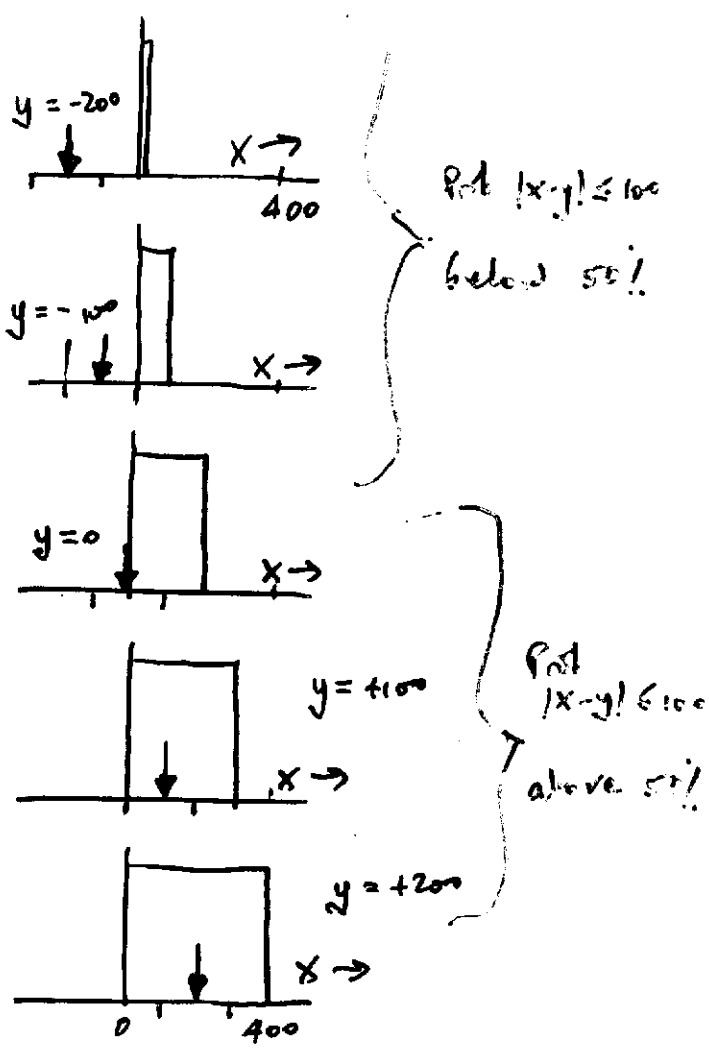
1) More specific on statement ①:

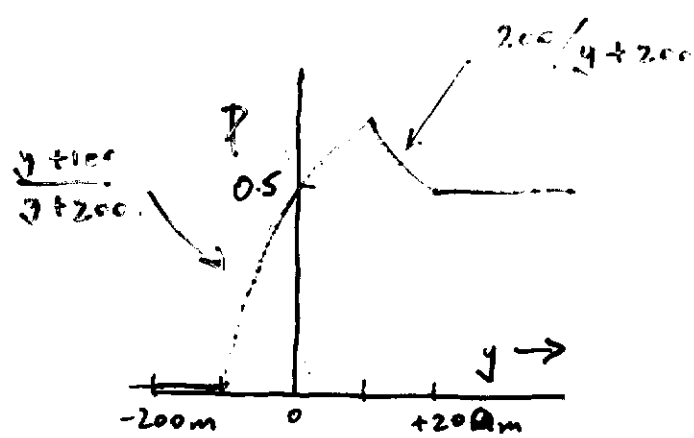
$$\text{Prob}(y-x) = \begin{cases} \text{Const} & \text{for } |y-x| < 200\,m \\ 0 & \text{for } |y-x| > 200\,m \end{cases}$$ [L'Hood]

2) Hunter x uniform in 0 → 1 km [PRIOR]









P = prob |x-y| ≤ 100 m.

CLASSICAL INTERVALS : PROBLEMS

Hard to understand (d'Agostini e-mail)

- 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

PETER CLIFFORD'S COMMENTS

1) NEYMAN : Mass production model.

2) COVERAGE ISN'T EVERYTHING

e.g. Choose number in range $-\infty \to +\infty$

90% coverage by quoting

a) $-\infty \to +\infty$ 90% of time

b) Random number 10% of time

Look for "Uniformly Most Powerful" tests

(Non)-Coverage = Error of 1st kind

Power : Error of 2nd kind

TALKS

BAYESIANS : Prosper, D'Agostini

FELDMAN - COUSINS Cousins

+ F-C' : Giunti, Zech, Woodroofe, Punzi

CL_s : Read, Nielsen, Jin

Specific topics : e.g. CDF

✓ oscillations : Chorus

Karmen ⊕ LSND

DELPHI TGCs (F-C)

B_s mixing (A)

Panel discussion

BAYES TALKS

d'Agostini:

Probability for Bayesians

Barlow's book ⇒

Very critical of Classical approach.

e.g. What does limit on M_H mean?

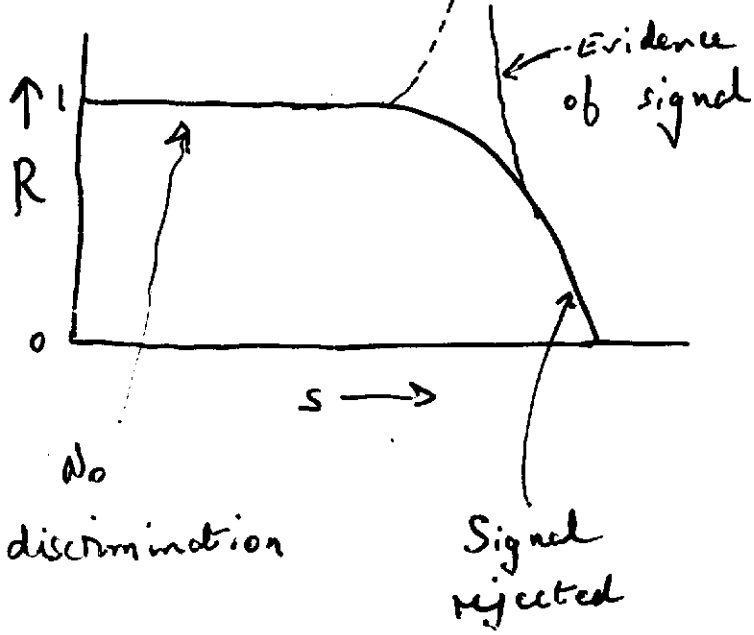
Recently advocates $\mathcal{L}$ ratio for testing hypotheses

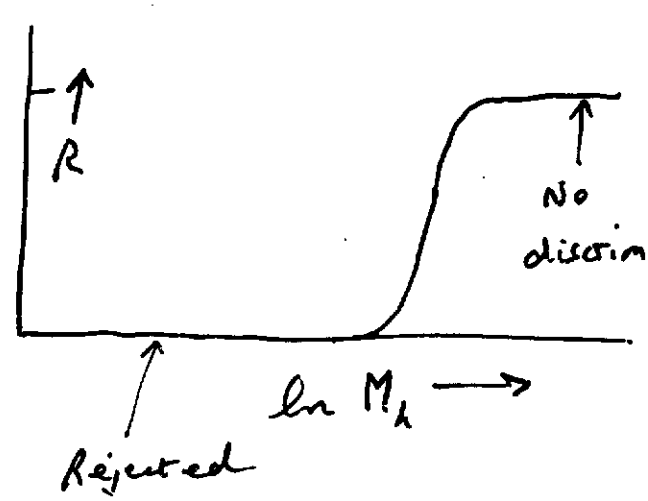
Used in Higgs limit [d'A + Degrassi]

$$\frac{P(s_1 \mid n_{obs}, b)}{P(s_2 \mid n_{obs}, b)} = \underbrace{\frac{P(n_{obs} \mid s_1, b)}{P(n_{obs} \mid s_2, b)}}_{\mathcal{L}\text{ ratio}} \times \frac{\text{Prior}(s_1)}{\text{Prior}(s_2)}$$

Rewrite as

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





$R(s)$ is indep. of prior

Giulio
d'Agostini

14

BAYES TALKS, contd

Harrison Prosper

Bayes needs "in principle" prior

but can make progress without

Examples from the front:—

Do M_t measurement

Solar ν flux

FELDMAN – COUSINS

Want to avoid empty classical intervals ⇒

Use "$\mathcal{L}$ ratio ordering principle" * to resolve ambiguity about "which 90% region?" ⇒

Has no flip-flopping problem.

Properties:

Coverage ✓

Limit ↓ as bgd increases above n_{obs} ✗
(but not unig

e.g. $n_{obs} = 0$ bgd $= 3.2$

cf $n_{obs} = 0$ bgd $= 0.5$

Hard to include systematics consistently ✗

e.g. Poisson Limit ↓ as sensitivity error ↑

Hard to combine ✗

Argument about null regions ?

Adopted by PDG ?

Computing time ✗

---

* Neyman-Pearson say $\mathcal{L}$ ratio best for hypothesis testing

[But this is parameter determination]

90% classical interval for Gaussian

$\sigma = 1$

$\mu \geq 0$

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

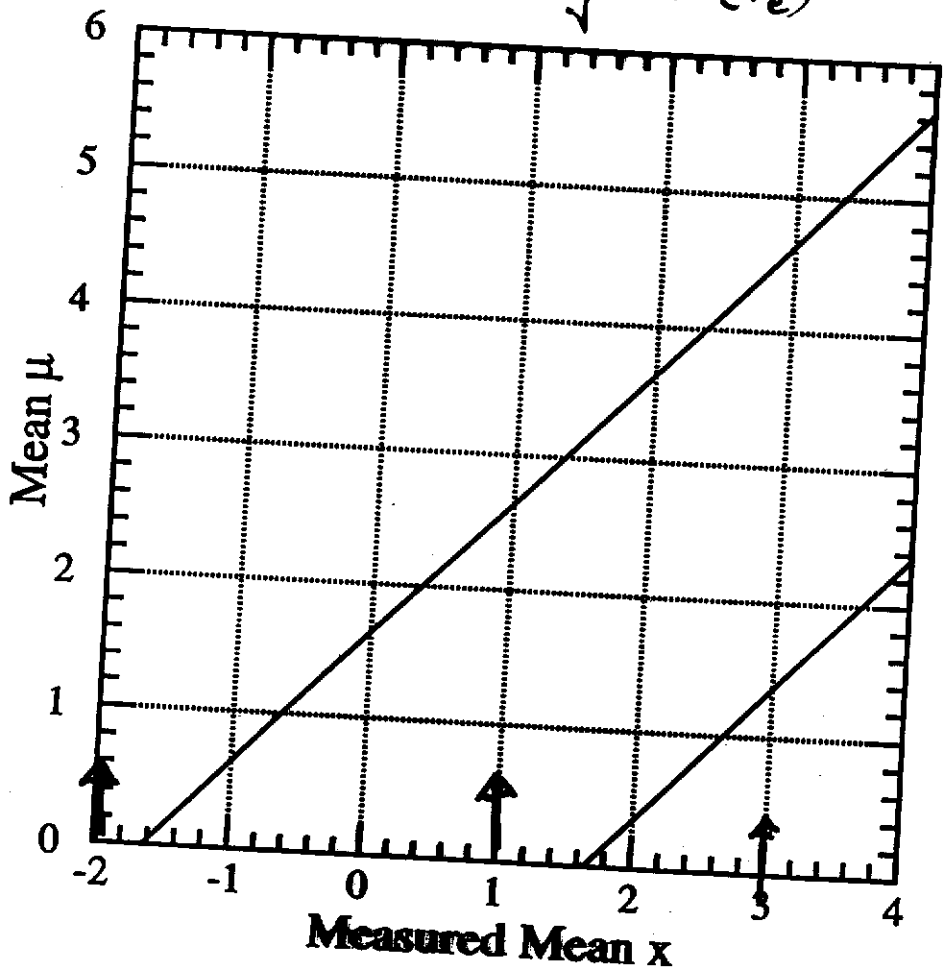


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$x_{obs} = 3$ Two sided limit

$x_{obs} = 1$ Upper limit

$x_{obs} = -2$ No region for μ

Feldman-Cousins 90% conf interval

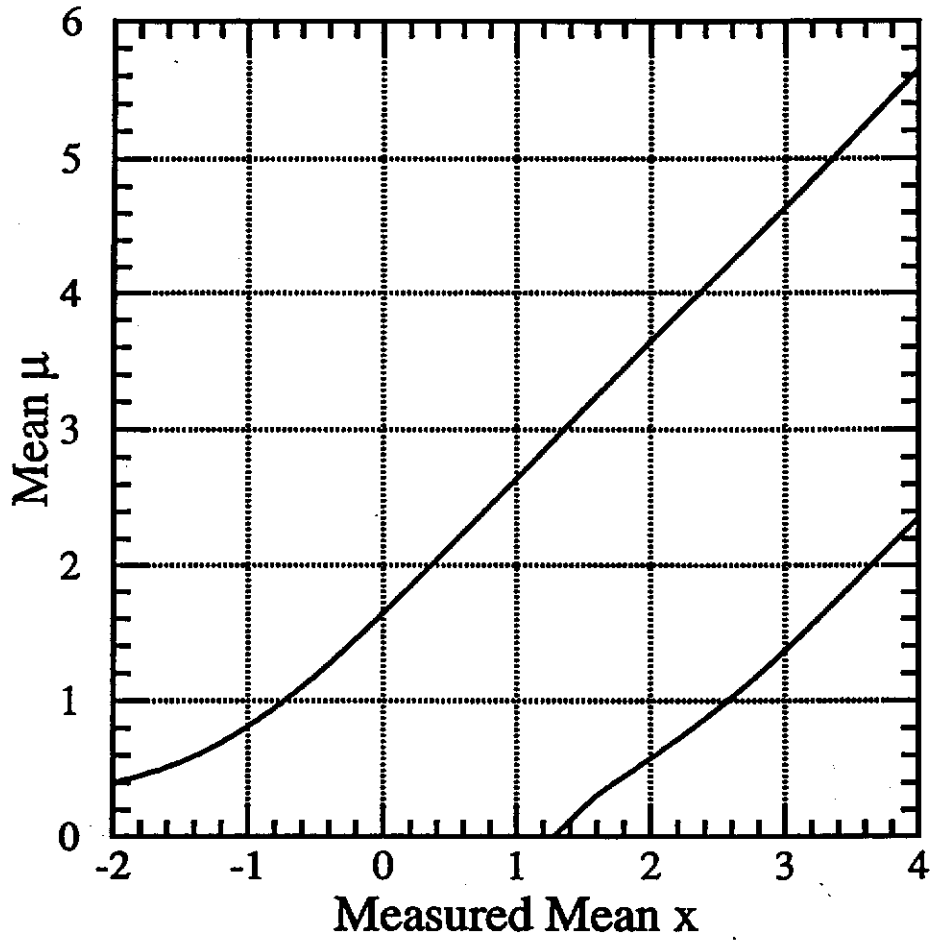


FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

$x_{obs} = -2$ Now gives upper limit

18

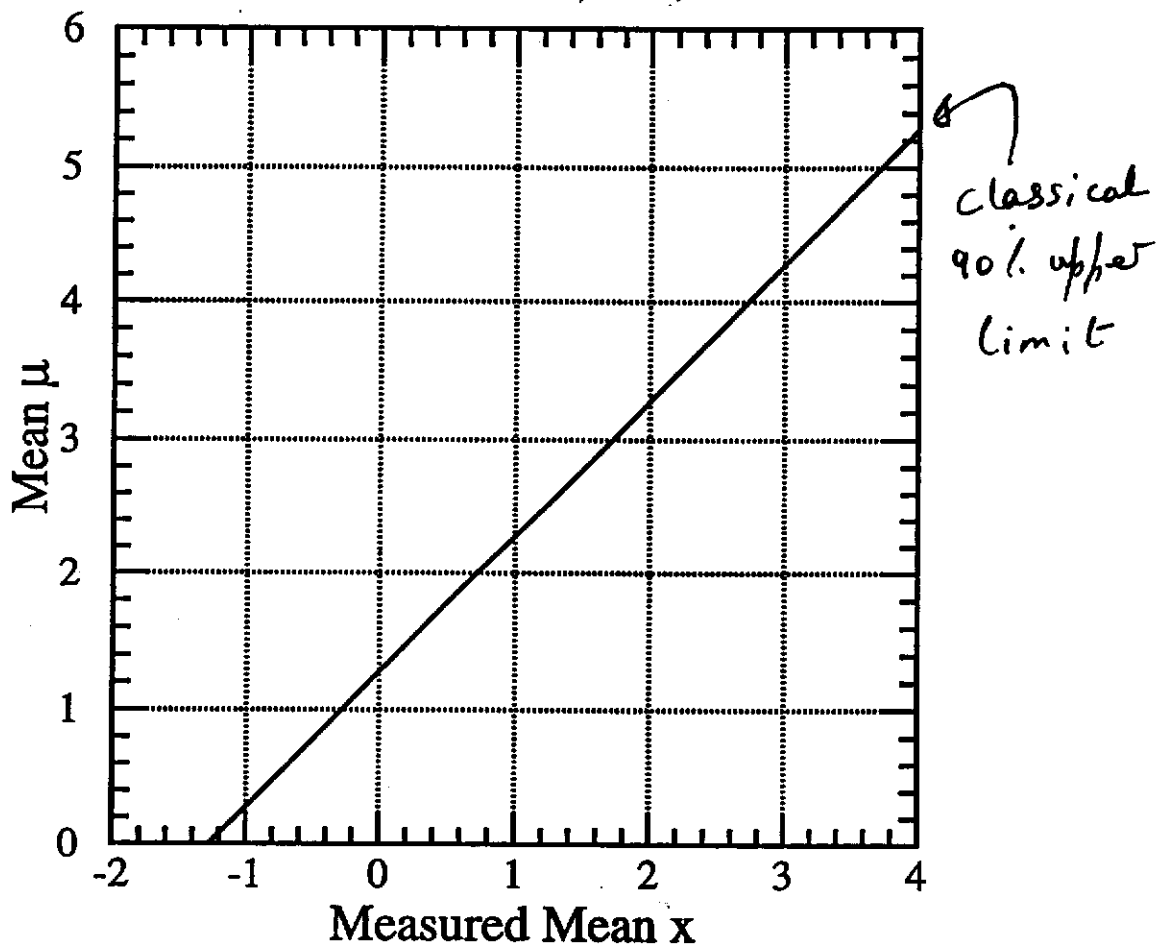


FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, in units of the rms deviation. The second line in the belt is at $x = +\infty$.

FLIP – FLOP

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

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

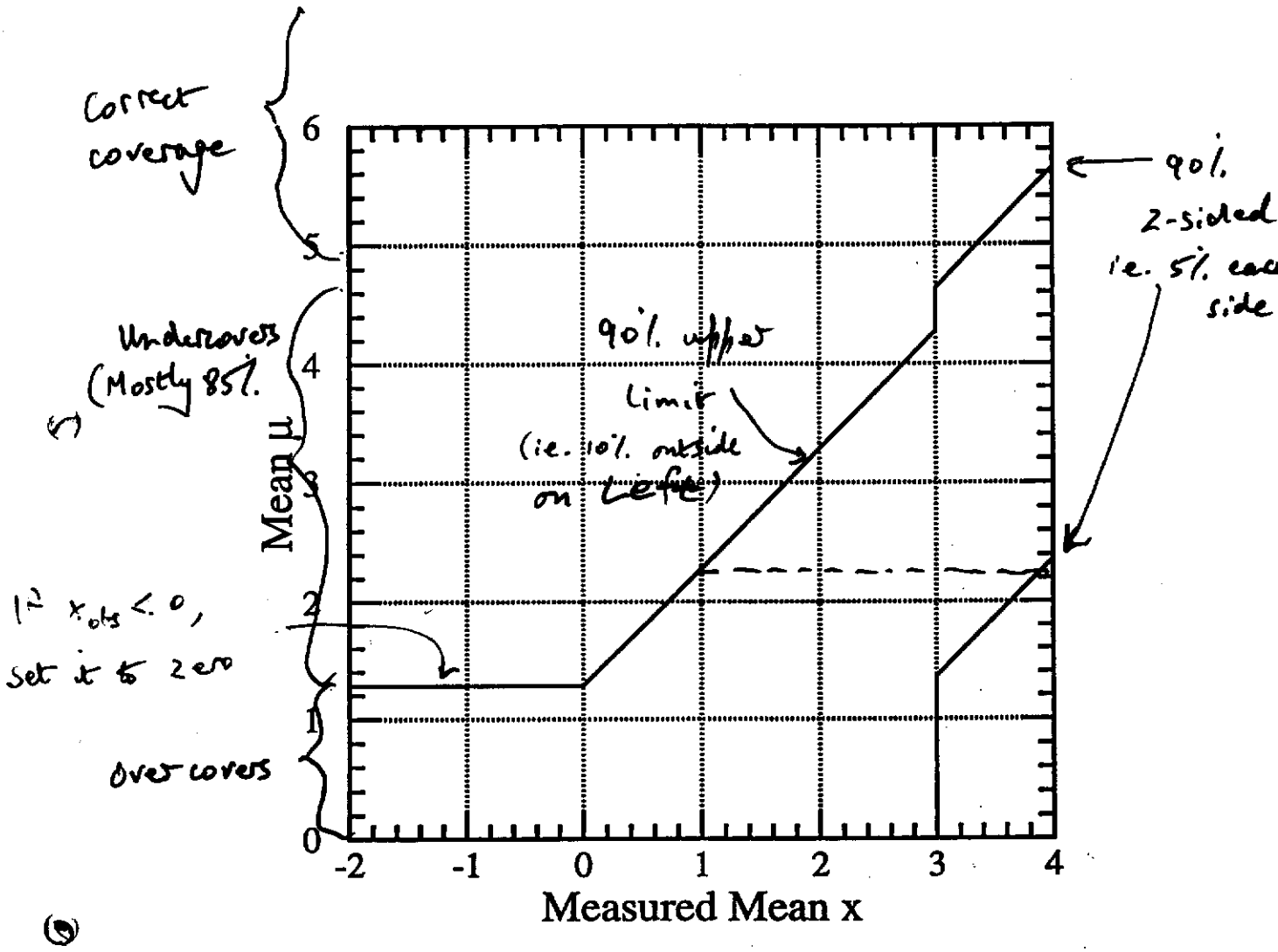


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence 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

F – C'

Try to overcome problem of limit ↓ as bgd > obser

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

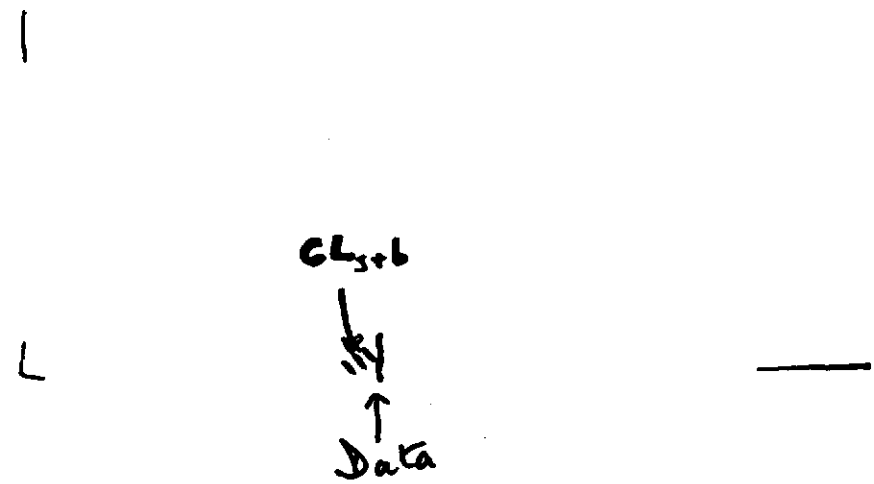
Harrison Prosper

"Freedom to do what you want"

CL_s for Higgs search (Alex Read)

Extract one number ⇒ whether data ~ bgd ⇒ exclude H

Extract one number ⇒ whether data ~ signal + bgd ⇒ discover H



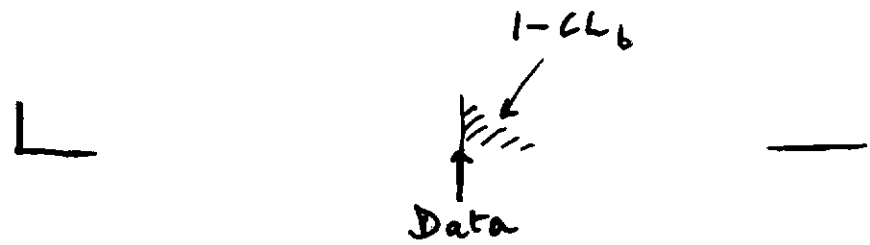
Exclusion :

$$CL_s = \frac{CL_{s+b}}{CL_b} = \text{small} \quad (\text{e.g.} < 5\%)$$

Integrals from $-\infty$

CL_b in denom to avoid exclusion when expected signal very small

N.B. CL_s not classical conf interval.



Discovery

$$\frac{1 - CL_b}{1 - CL_{s+b}} = \frac{\text{integral from RHS for bgd}}{\text{integral from RHS for sig + bgd}}$$

= very very small (e.g. equiv to 5σ

Denom to avoid discovery when no sensitivity

Complaints from F & C:

1) Not frequentist conf interval

2) No 1-sided to 2-sided smooth transition

Q: Non-Gaussian systematics?

A.R: More important for discovery than for exclusion.

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
Systematics } in Classical approach

2) Combining results

e.g. What do you need for F-C?

[Fred James : What do you want it for?]

FNAL : CDF + DØ

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

FNAL C.L.W Mond. March 27th

Tues March 28th

FNAL Program

Monday	am 1	James	Choosing foundatio
	am 2	Prosper, Berger	Bayes
	pm 1	Linnemann	Bayes
		Kruse, Maeshima	Searches
		James	Reply to Bayes
	pm 2	Murray	CLs
		Raja, Narsky	Limits methods
Tuesday	am 1	Cousins, Rolke, Messier, Kafka	F-C
	am 2	Schee, Yellin	CDMS
		Roe	Methodology
	pm	Feldman, Murray, Greenlee	Combining results
			CONCLUSIONS