FNAL STATISTICS LECTURE

#4

LOUIS LYONS

BAYES v FREQUENTISM

BAYES versus FREQUENTISM

The Return of an Old Controversy

- The ideologies, with examples
- Upper limits
- Systematics

Louis Lyons, Oxford University
and CERN

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 = you are in CERN

B = you are at Workshop

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

$$P(hyothesis; data) \; \alpha \; P(data; hypothesis) \text{ x } P(hypothesis)$$

↑ posterior — ↑ likelihood — ↑ 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?

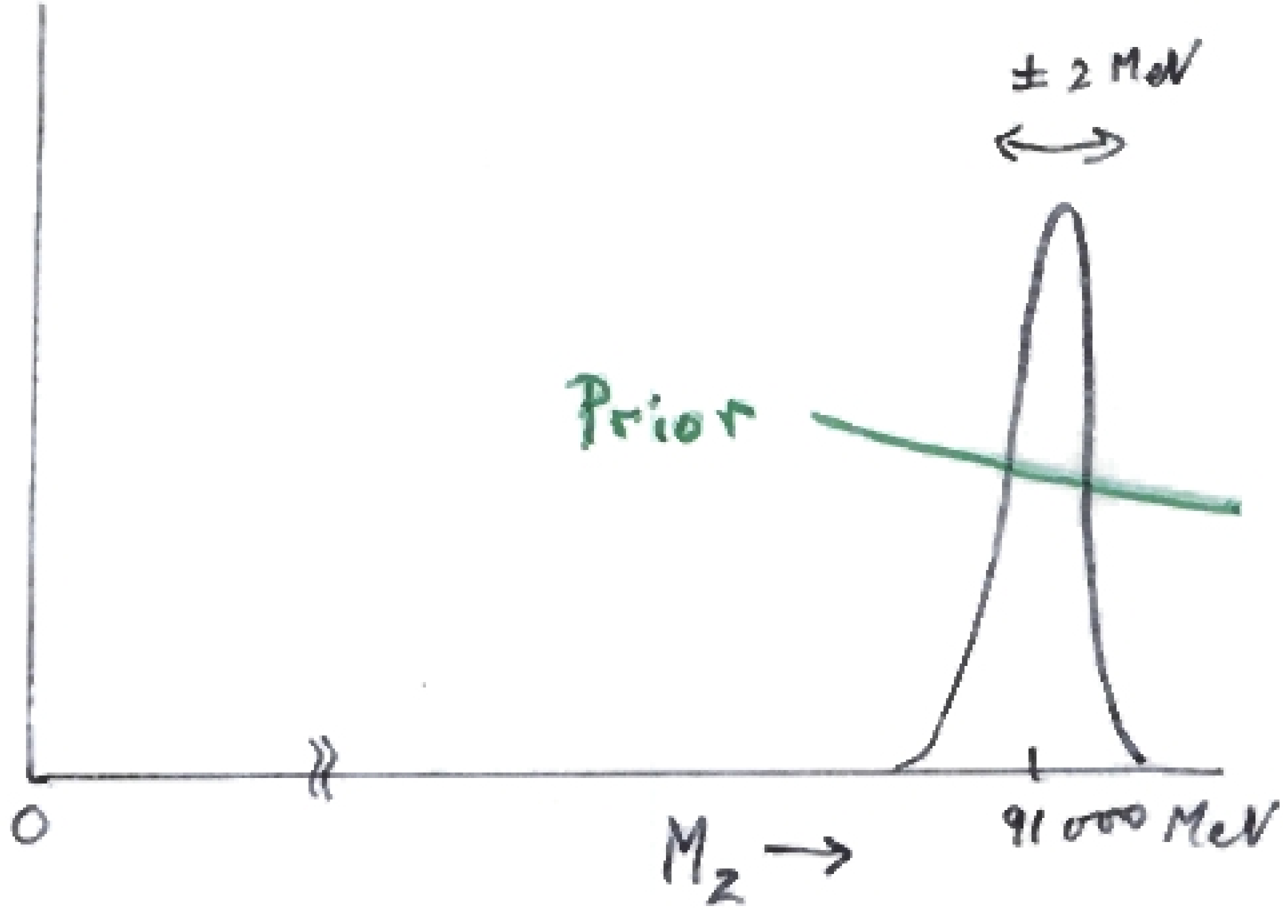
Uninformative prior:

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

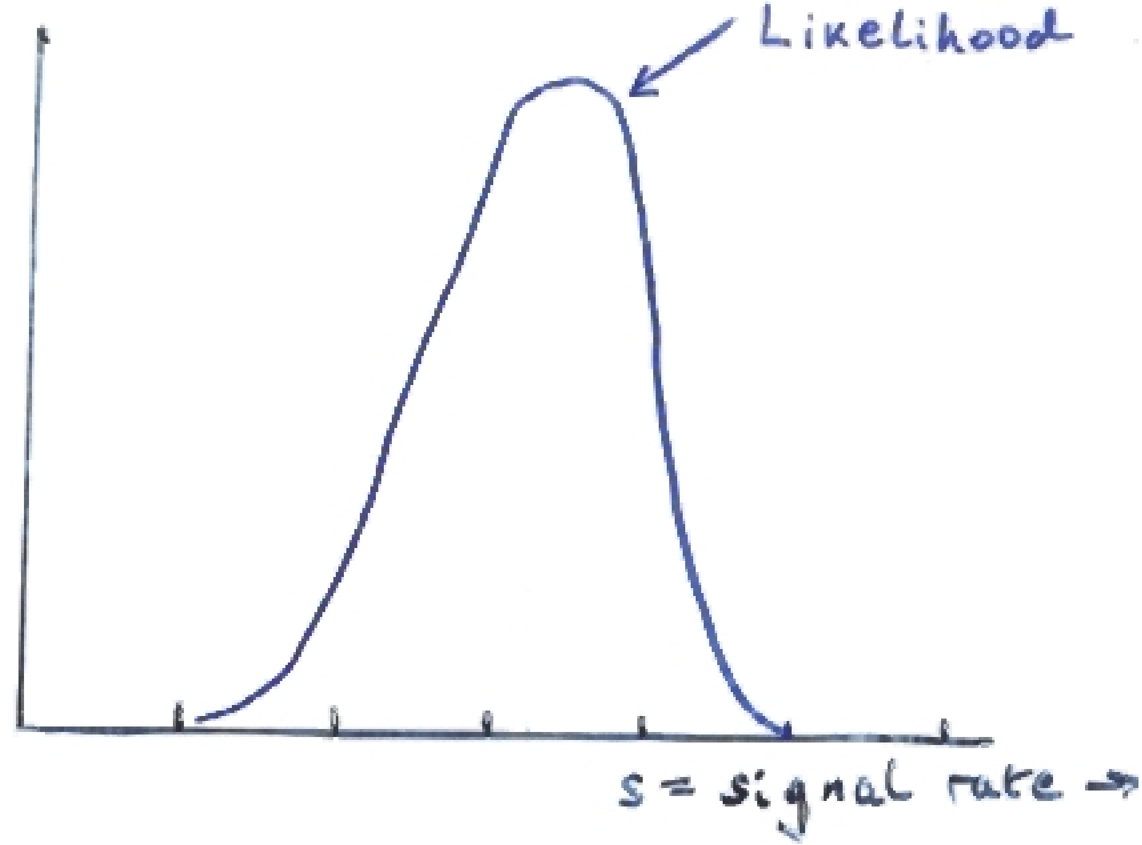
Unimportant if "data overshadows prior"

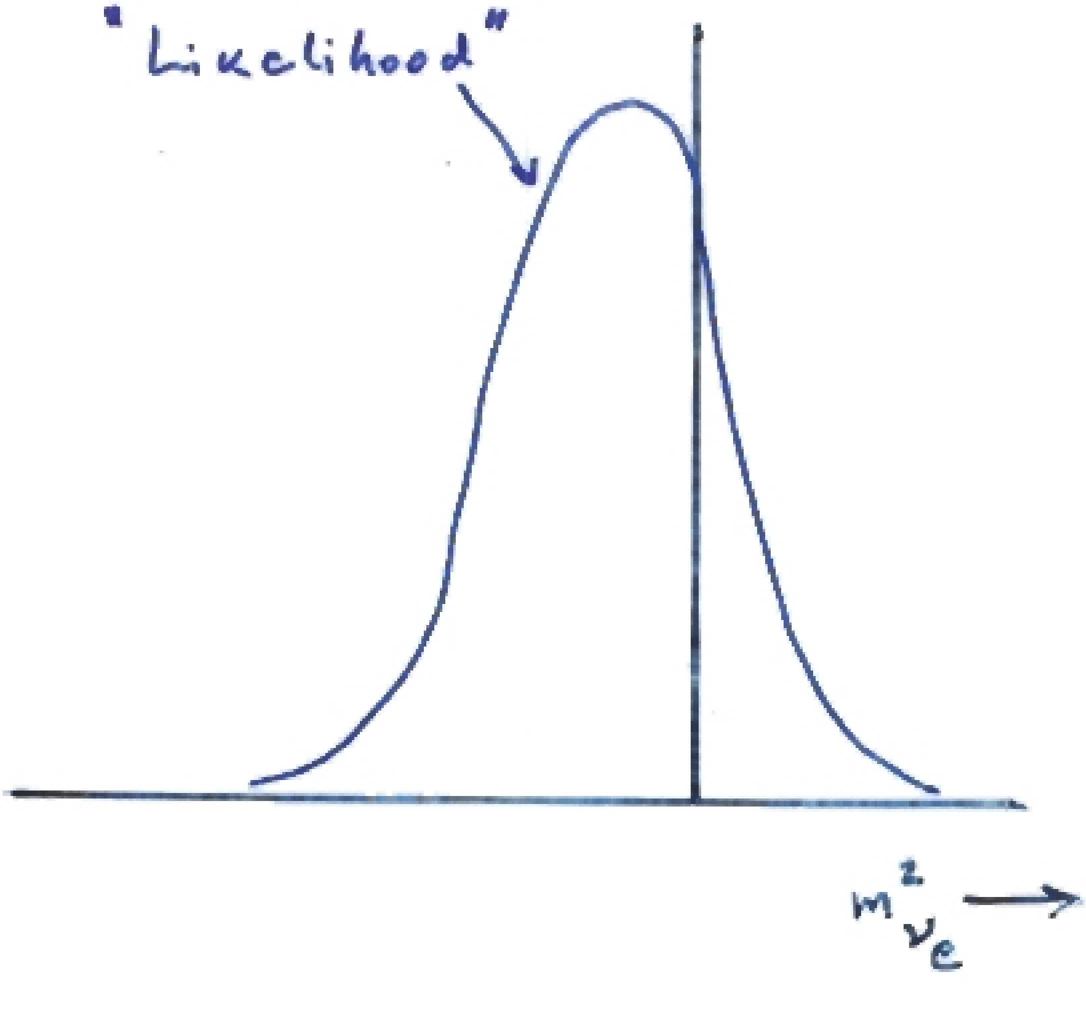
Important for limits

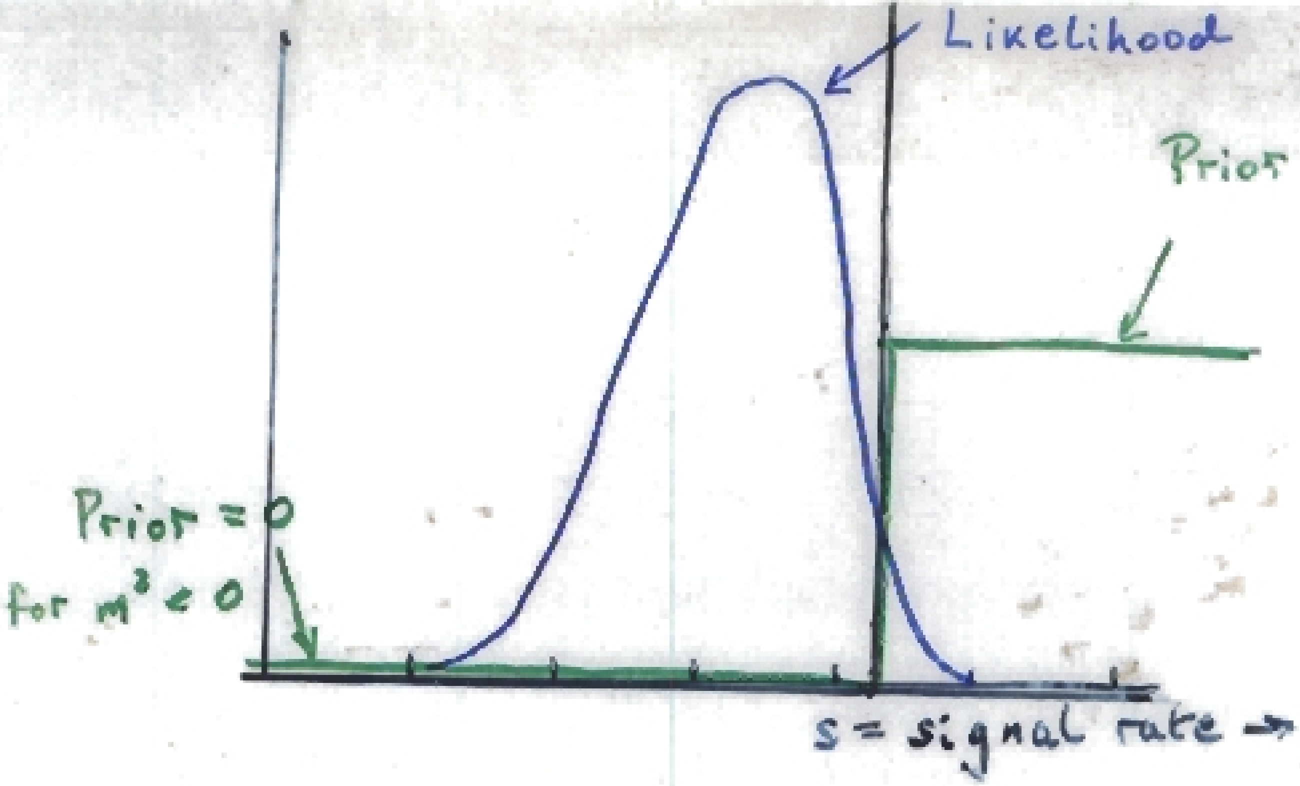
Subjective or Objective prior?

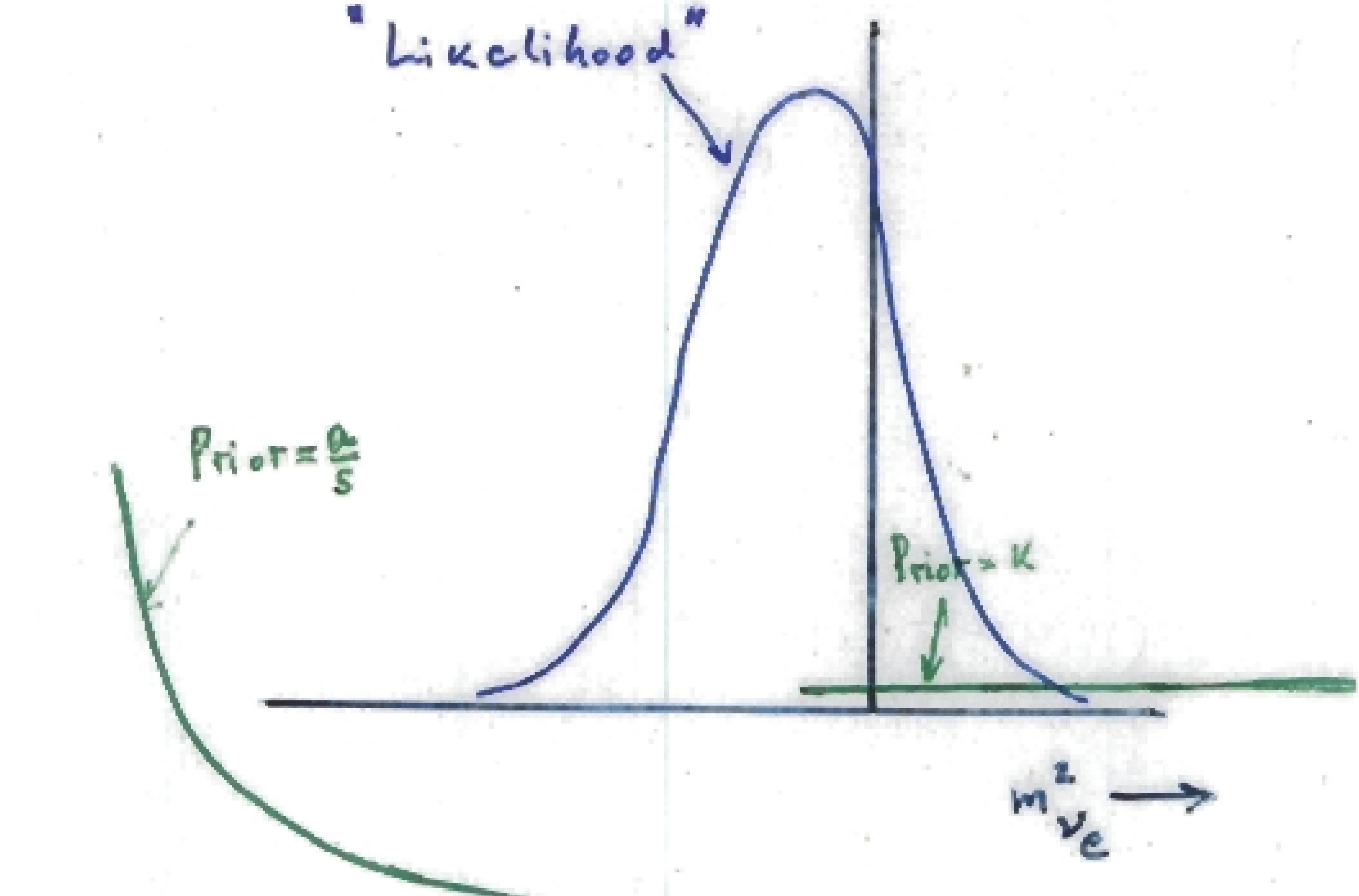


Data overshadows the Prior

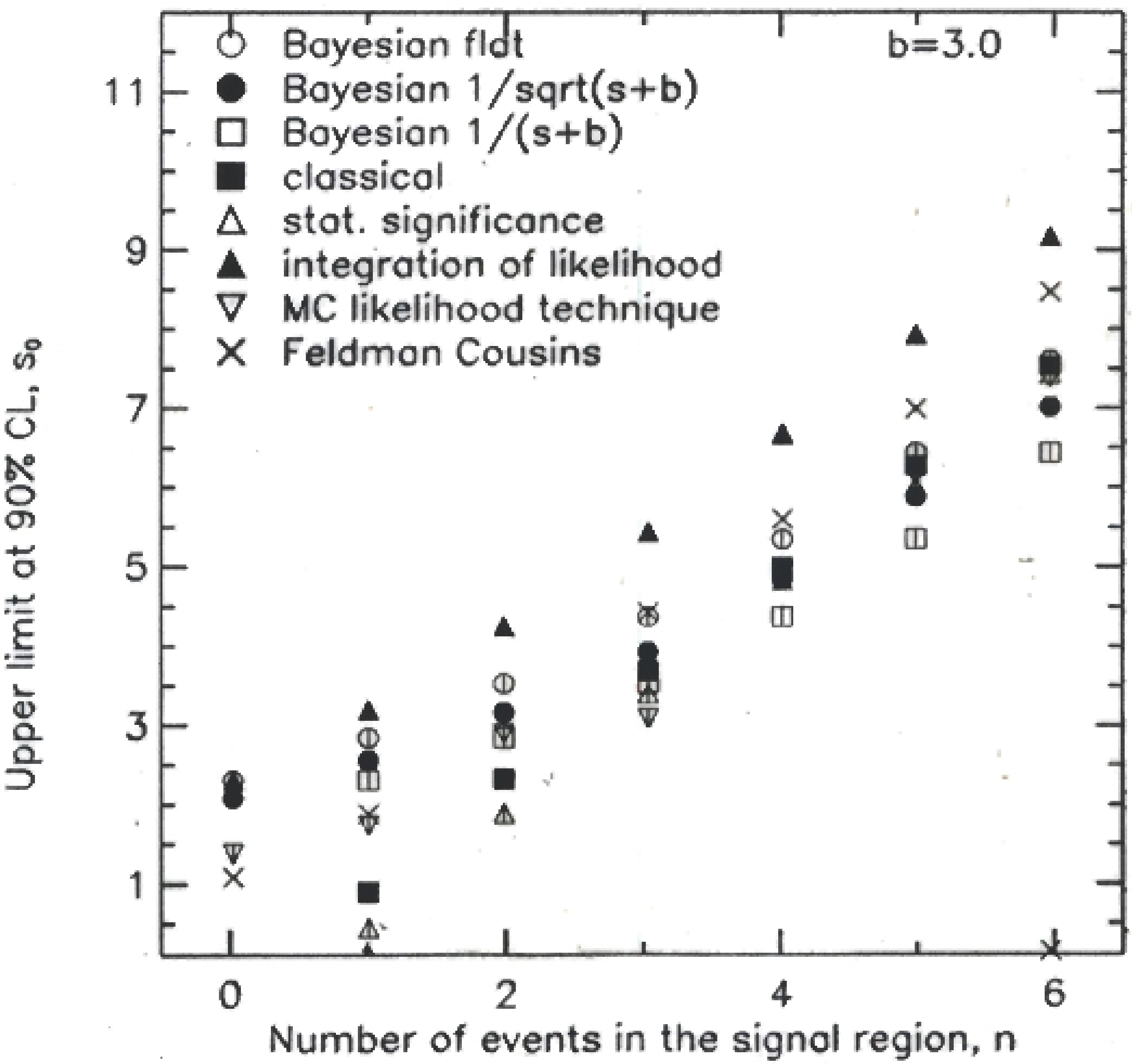

Likelihood
s = signal rate →


"Likelihood"
$m^2_{\nu_e}$ →


Likelihood
Prior
Prior = 0
for m² < 0
s = signal rate →


"Likelihood"
Prior = a/s
Prior = k
m²νe →

ILYA NARSKY
FNAL CL Workshop



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

Theory = male or female

Data = pregnant or not pregnant

P (pregnant ; female) ~ 3%

but

P (female ; pregnant) >>>3%

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"

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 π and protons

probability (p tag;real p) = 0.95

probability (π tag; real p) = 0.05

probability (p tag ; real (π) = 0.10

probability (π tag ; real π) = 0.90

Particle gives proton tag. What is it?

Depends on prior = fraction of protons

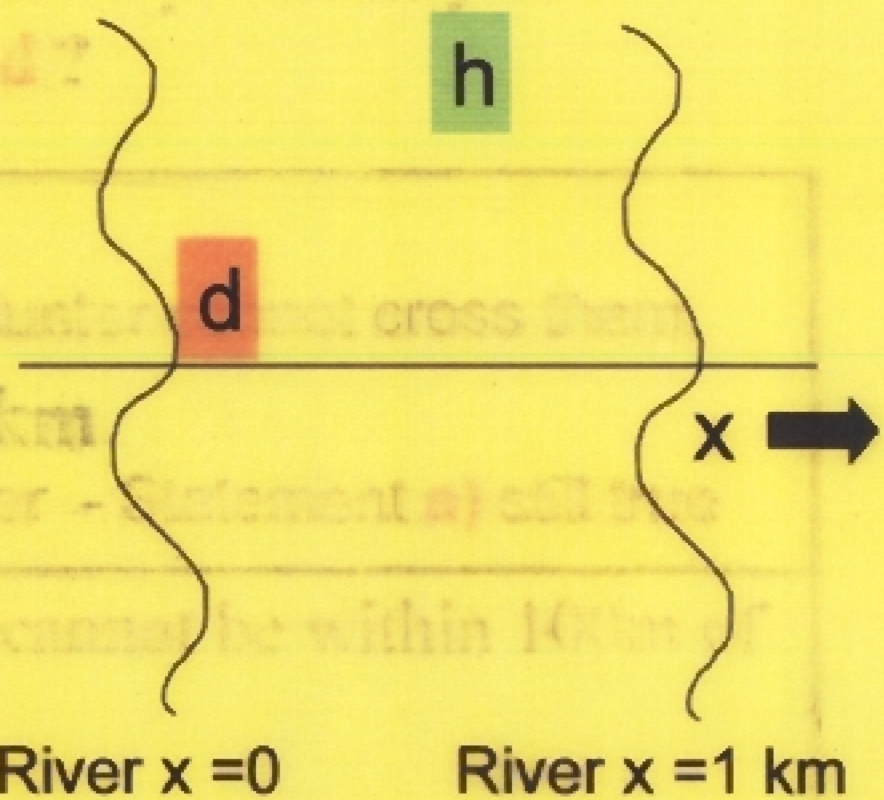
If proton beam, very likely

If general secondary particles, more even

If pure π beam, ~ 0

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



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.

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

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

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

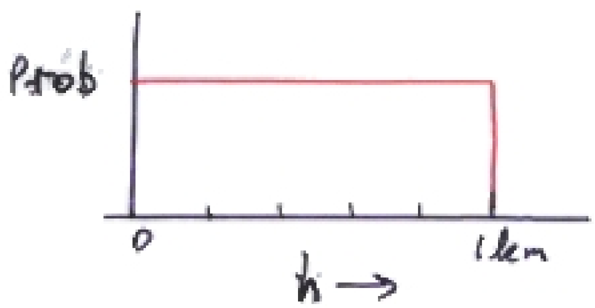
Statement b) untrue

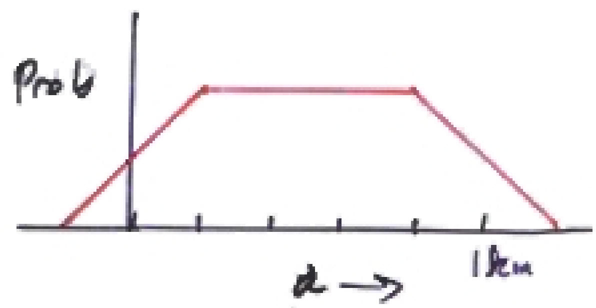
Example:

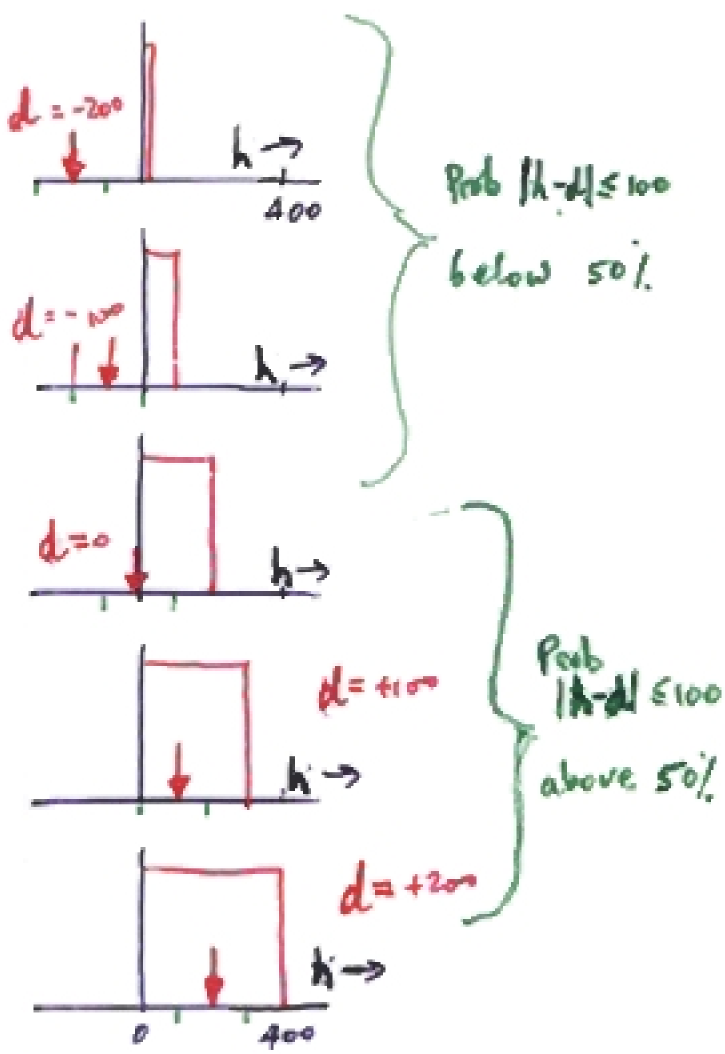
1) More specific on statement ①:

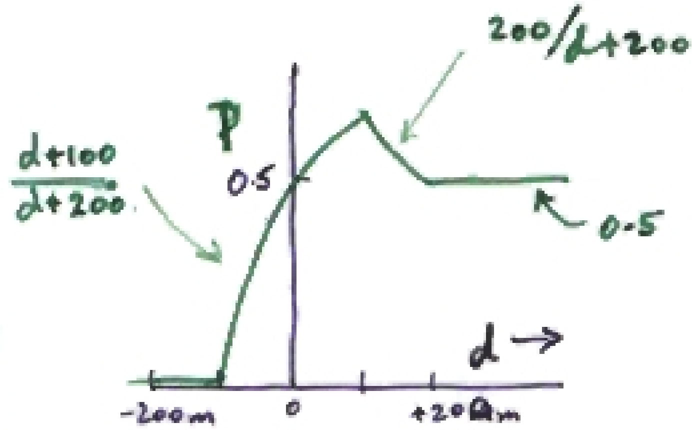
$$\text{Prob}(d - h) = \begin{cases} \text{Const} & \text{for } |d-h| < 200\text{m} \\ 0 & \text{for } |d-h| > 200\text{m} \end{cases} \quad \text{[L'HOOD]}$$

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









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

BF7

Classical Approach

Neyman "confidence interval" avoids pdf for μ

uses only P(x; μ)

Confidence interval $\mu_1 \rightarrow \mu_2$:

P($\mu_1 \rightarrow \mu_2$ contains μ) = α True for any μ

↑ ↑ ↑

Varying intervals from ensemble of experiments — fixed

Gives range of μ for which observed value x_0 was "likely" (α)

Contrast Bayes : Degree of belief = α that μ_t is in $\mu_1 \rightarrow \mu_2$

CLASSICAL (NEYMAN) CONFIDENCE INTERVALS

Uses only P(data | theory)

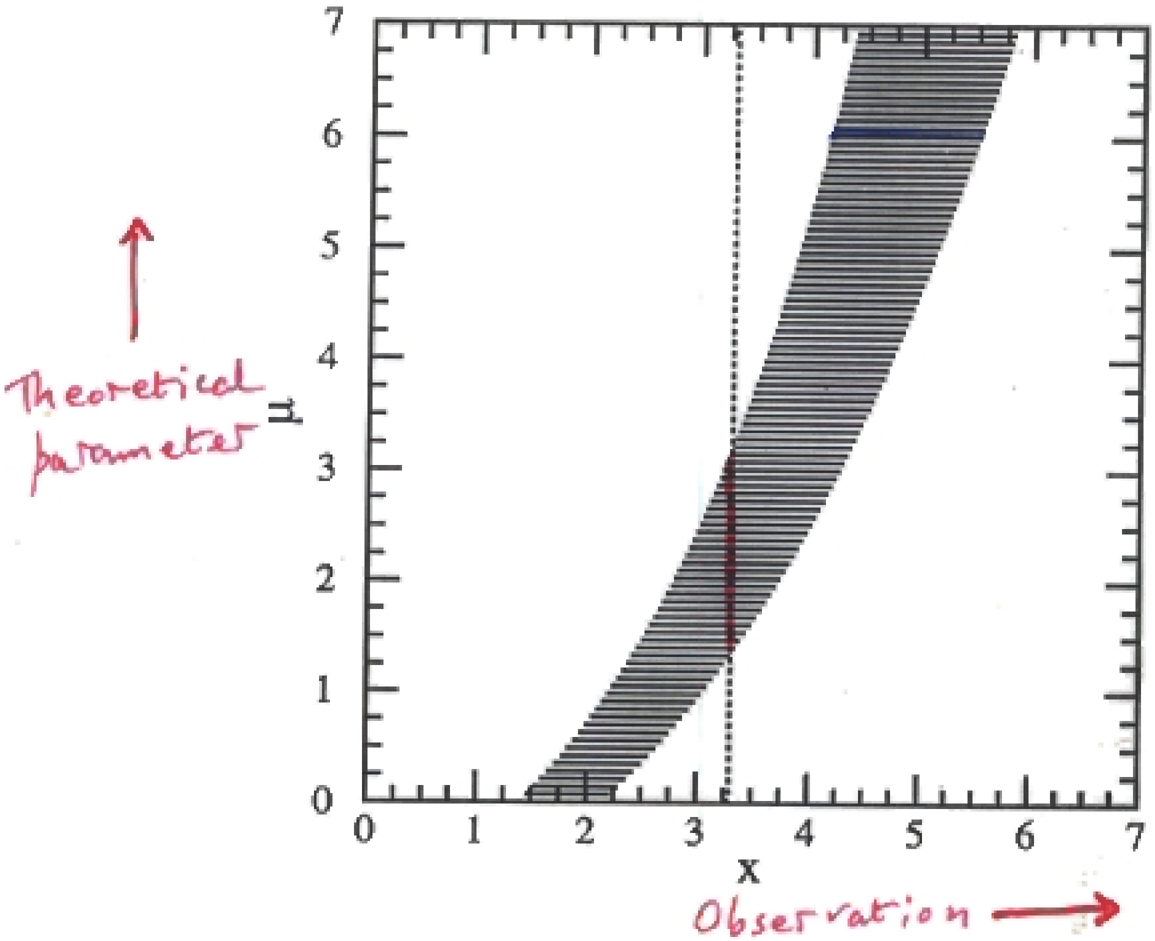


FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] | \mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through x_0.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

90% classical interval for Gaussian

$\sigma = 1$

$\mu \geq 0$

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

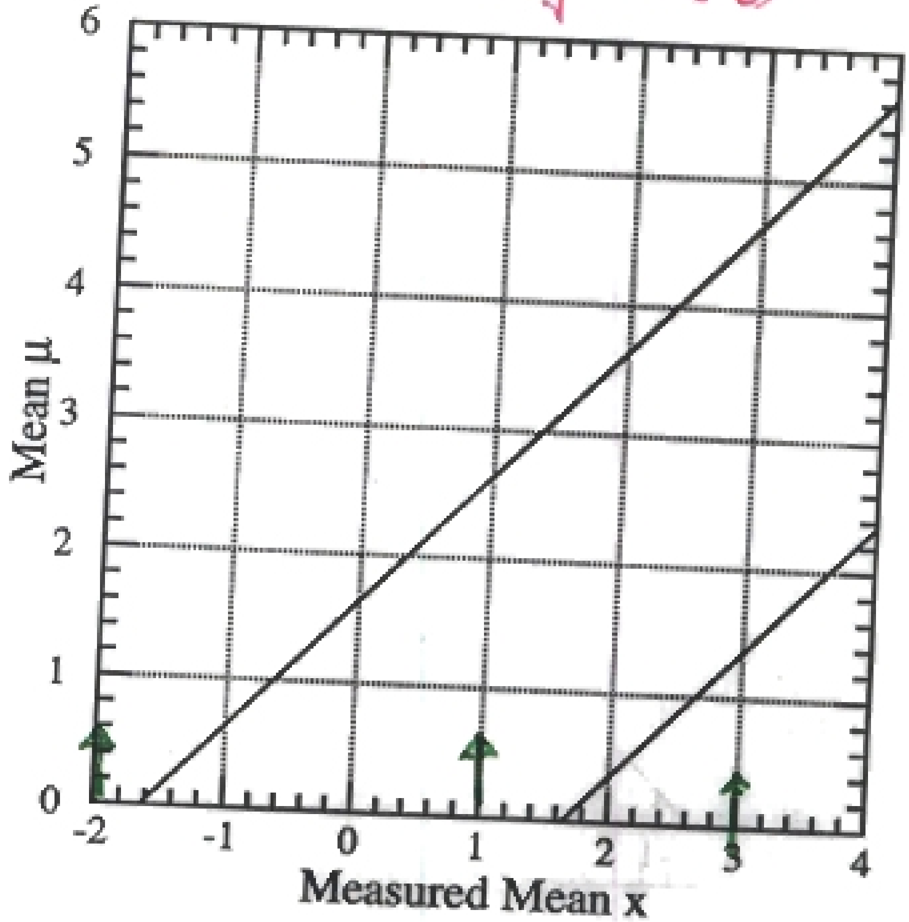


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$x_{obs} = 3$ Two sided limit

$x_{obs} = 1$ Upper limit

$x_{obs} = -2$ No region for μ

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

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

Frequentist

μ_l and μ_u known, but random

μ unknown, but fixed

Probability statement about μ_l and μ_u

Bayesian

μ_l and μ_u known, and fixed

μ unknown, and random

Probability/credible statement about μ

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 ⇒

Feldman-Cousins 90% conf interval

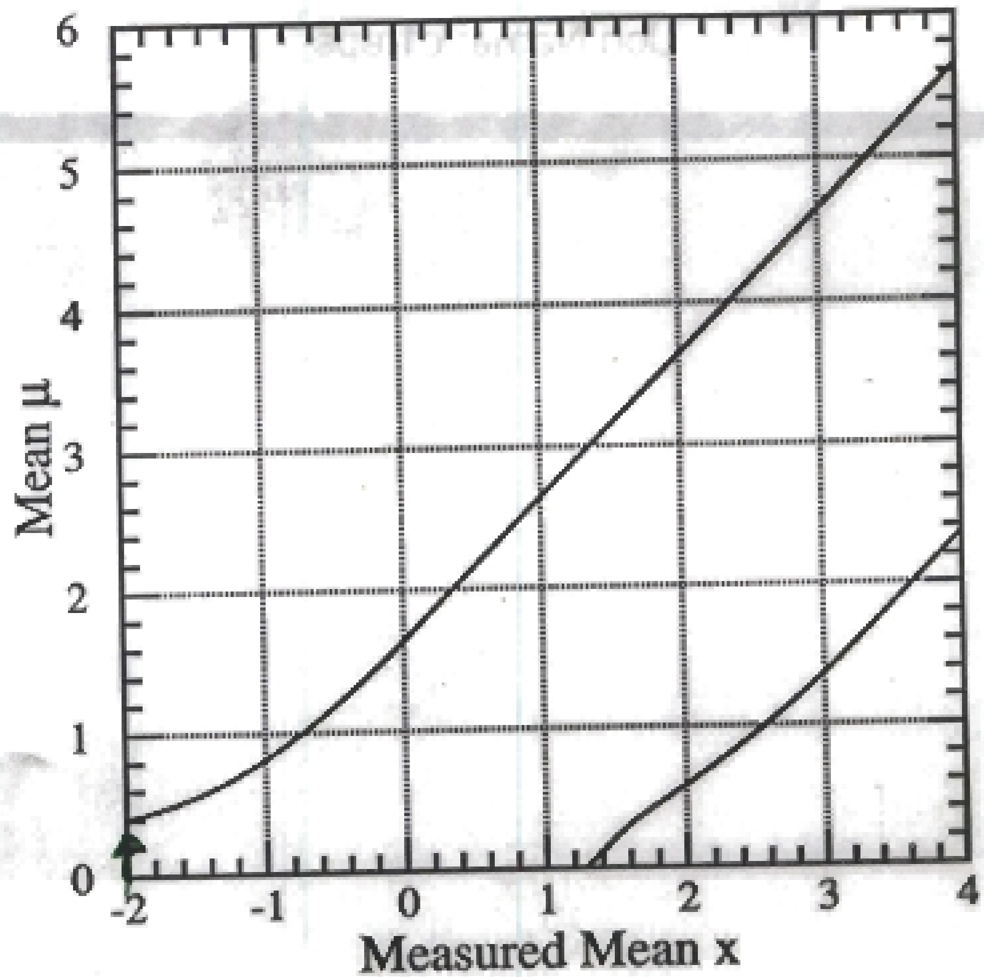


FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

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

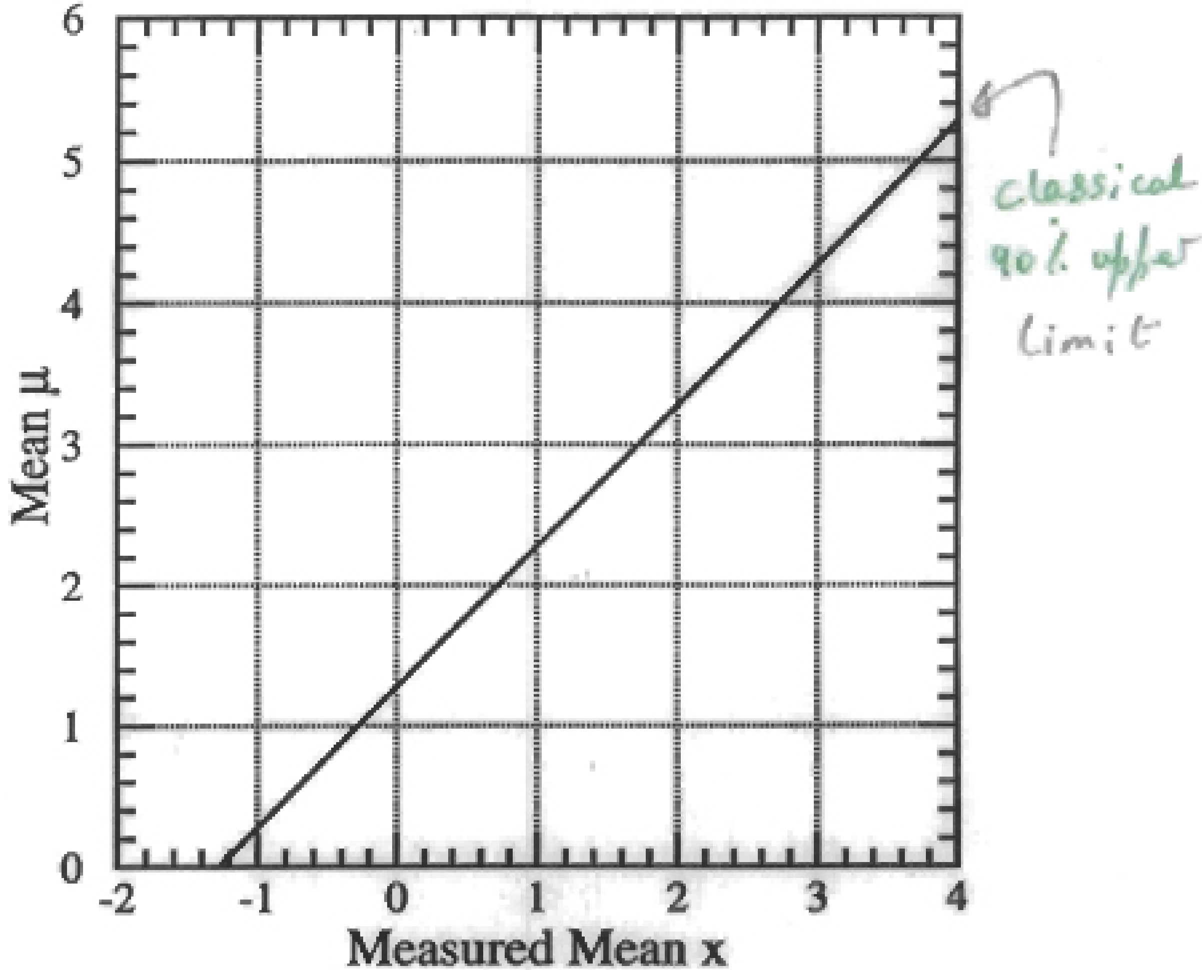


FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, in units of the rms deviation. The second line in the belt is at $x = +\infty$.

FLIP – FLOP

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

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

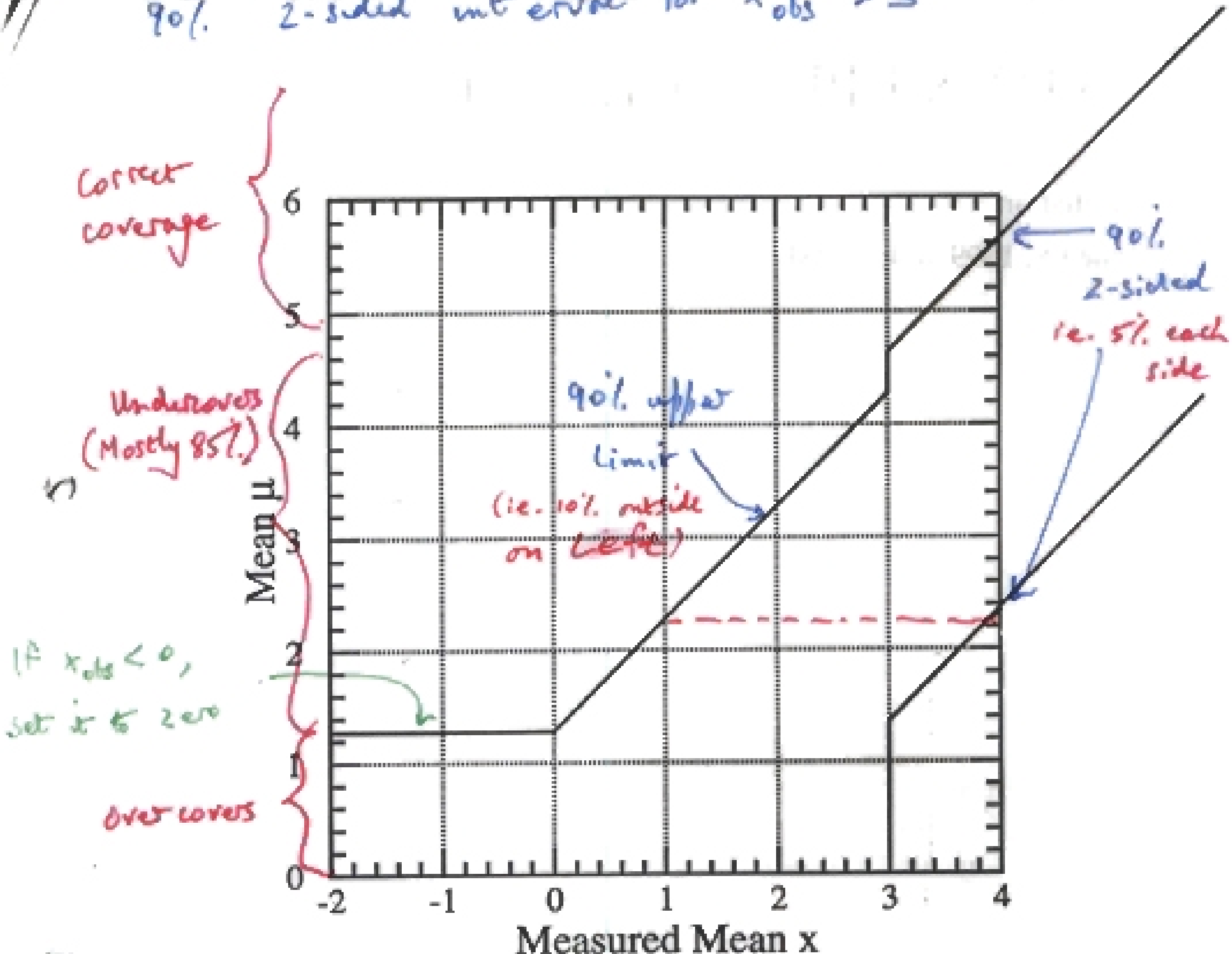


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence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

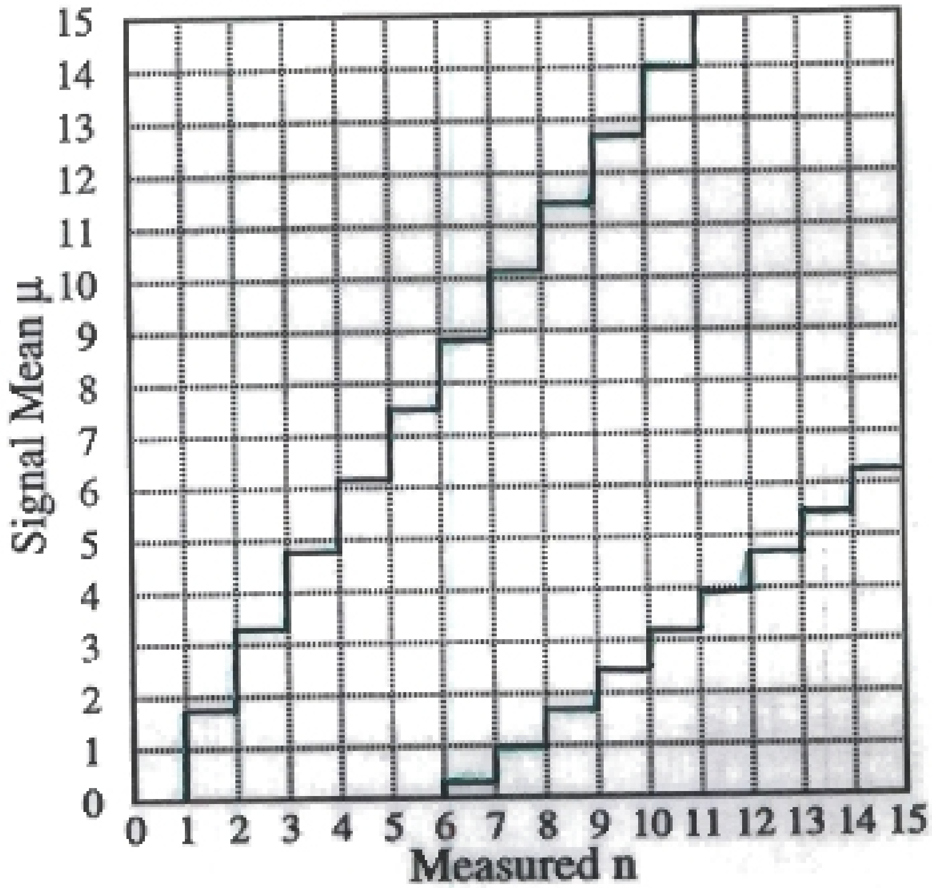


FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$.

Standard Frequentist
for Poisson mean μ

FELDMAN + COUSINS FOR

POISSON MEAN μ

90% Conf

$b = 3.0$

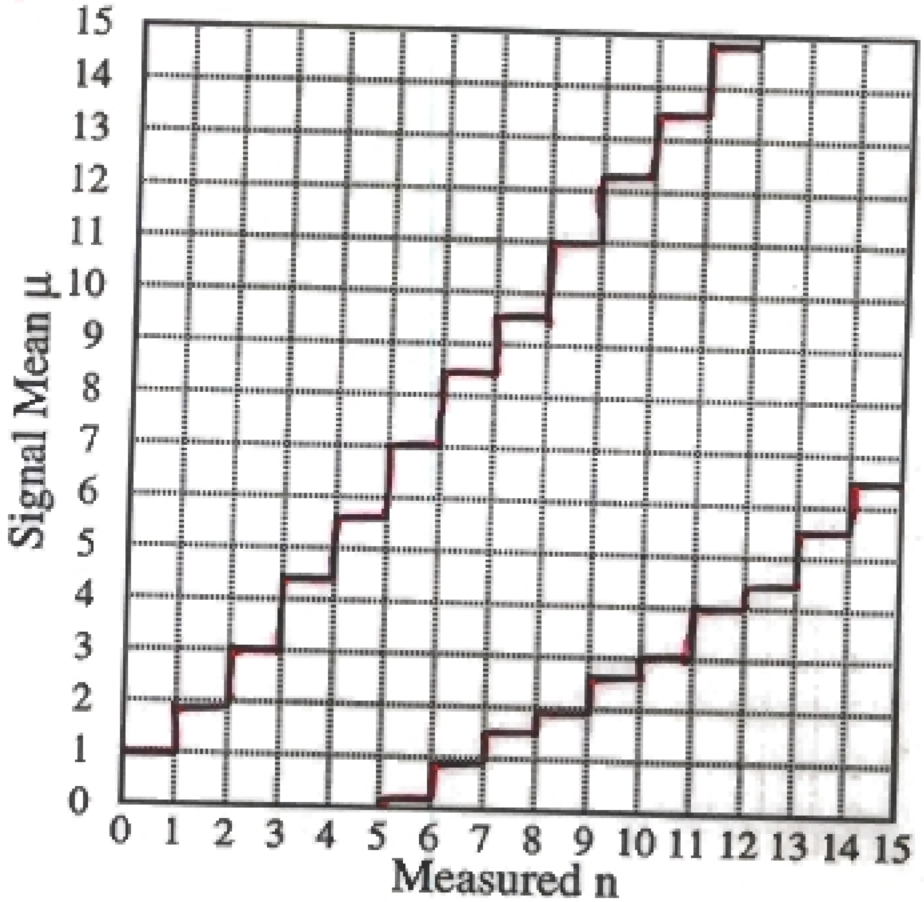


FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$.

22

FREQUENTIST POISSON C.B. CONSTRN.

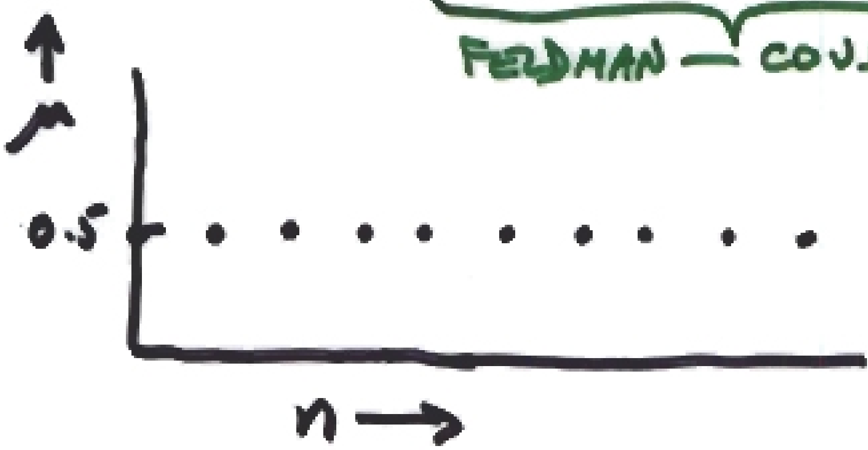
TABLES

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

n	$P(n\|\mu)$	μ_{best}	$P(n\|\mu_{\mathrm{best}})$	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		✓	

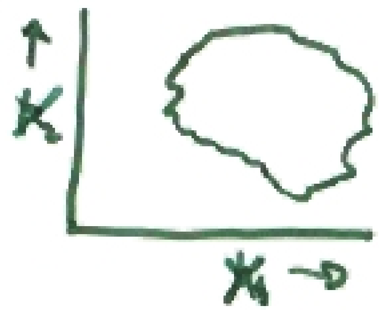
<10%   <5%   <5%   Prob ordering

FELDMAN — COUSINS



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 ?

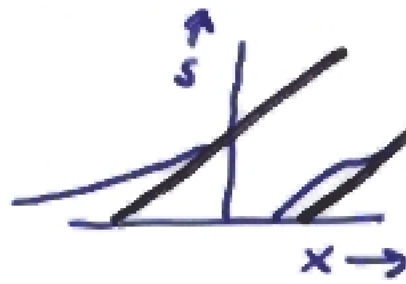
NEYMAN CONSTRUCTION ⇒ CPU-INTENSIVE
(ESP IN SEVERAL DIMENSIONS)

MINOR PATHOLOGIES : DISJOINT 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



Bayesian

Pros:

Easy to understand

Physical Interval

Cons:

Needs prior

Hard to combine

Coverage

Standard Frequentist

Pros:

Coverage

Cons:

Hard to understand

Small or Empty Intervals

Different Upper Limits

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 defintion
Prob of parameters?	Yes	Anathema
Needs prior?	Yes	No
Choice of interval?	Yes	Yes (except F+C)
Data considered	Only data you have	….+ more extreme
Likelihood principle?	Yes	No

Bayesian versus Frequentism

	Bayesian	Frequentist
Ensemble of experiments	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"