Language: 英語/English

# KEK 物理セミナー Physics Seminar

10月14日 (火) October 14 (Tue)
16:00 ~ 17:00

4号館 345号室 Room 345 in the 4th Building

## Bayes versus Frequentism: the return of an old controversy

**Louis Lyons**
**(Oxford/KEK)**

- Abstract:
It is possible to spend a lifetime analysing data without realising that there are these two very contrasting approaches to the subject. This talk explains the differences, and gives examples from Physics and from every-day life to illustrate cases where they give different results.

Past presentations are archived at our home page, http://seminar.kek.jp/physics/ .

Organizers: T. Komatsubara and I. Adachi

2003.10.02

# BAYES versus FREQUENTISM

## The Return of an Old Controversy

- The ideologies, with examples
- Upper limits
- Systematics

Louis Lyons, Oxford University

and CERN



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 2/ 54

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

Bayesianism and Frequentism.



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 3/ 54

# 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

03-10-15;17:58 ;BELLE GROUP ;scan for file ;+81-298-64-5340 # 4/ 54

# 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)



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 5/ 54

# PROBABILITY

## MATHEMATICAL

Formal

Based on Axioms

## FREQUENTIST

Ratio of frequencies as n→ 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"



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 scan for file # 6/ 54

# 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....

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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 7/ 54

Bayesian

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

Bayes Theorem

$$P(hyothesis;data) \propto P(data;hypothesis) \times P(hypothesis)$$

posterior ↑ likelihood ↑ prior ↑

Problems: P(hyp..) true or false

"Degree of belief"

Prior What functional form?

Coverage

Goodness of fit



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 8/ 54

P(hypothesis.....) True or False

"Degree of Belief"

credible interval

Prior: What functional form?

Uninformative prior:

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

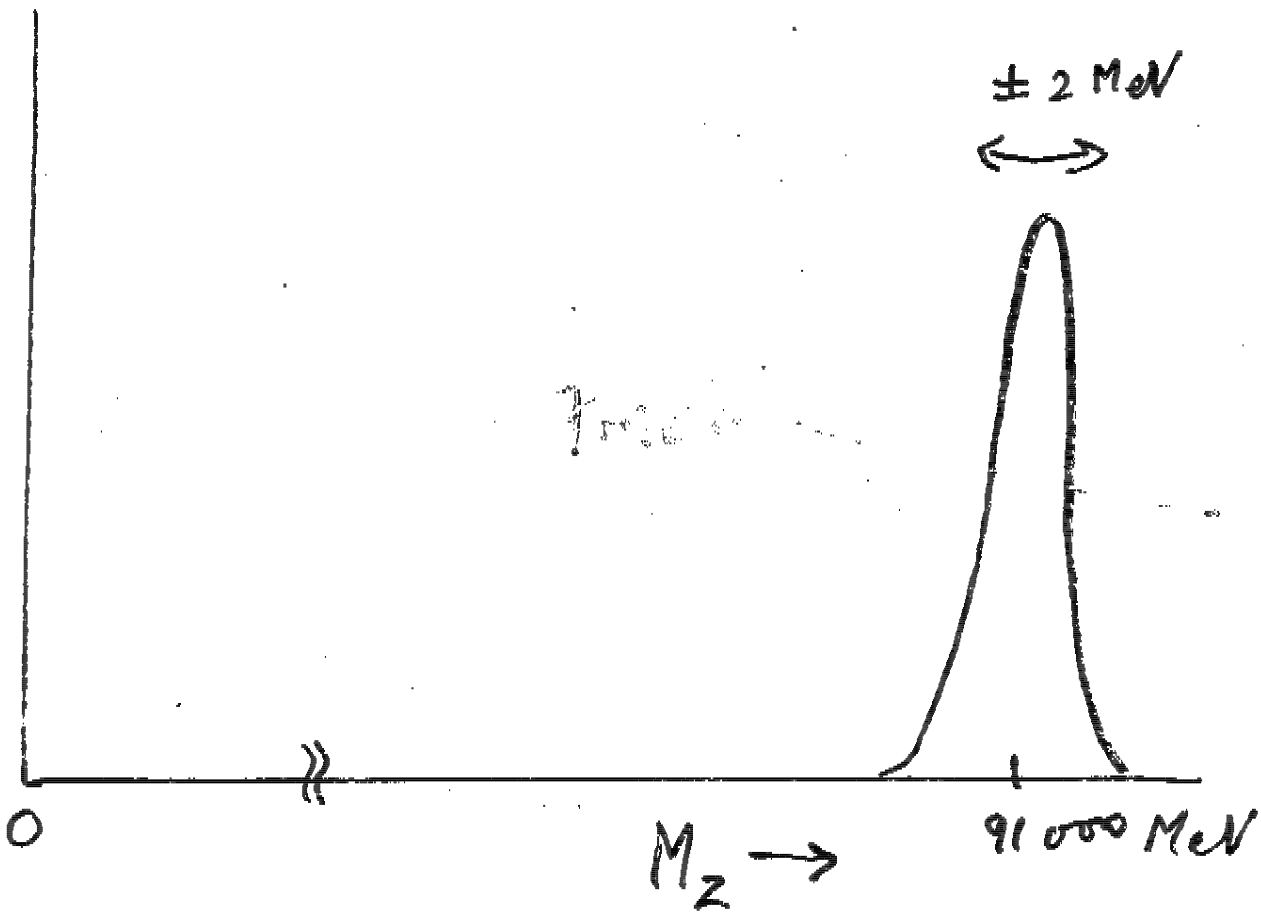
Unimportant if "data overshadows prior"

Important for limits

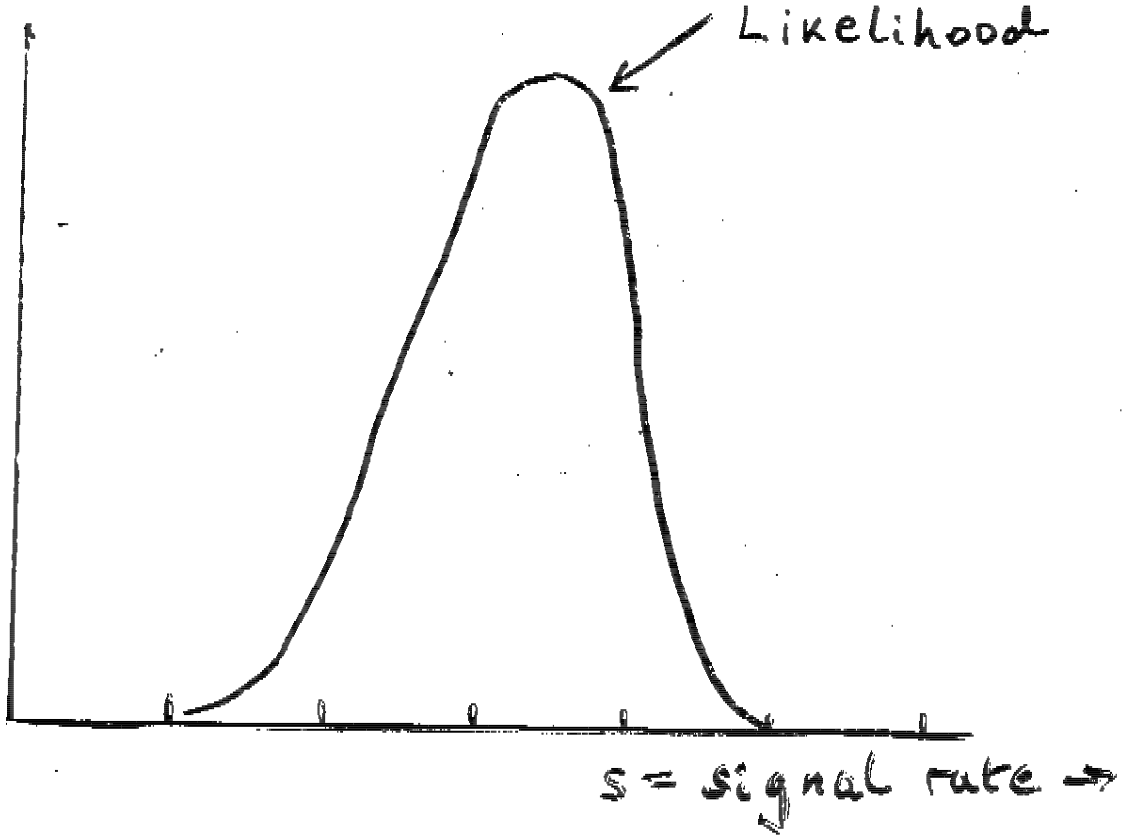
Subjective or Objective prior?

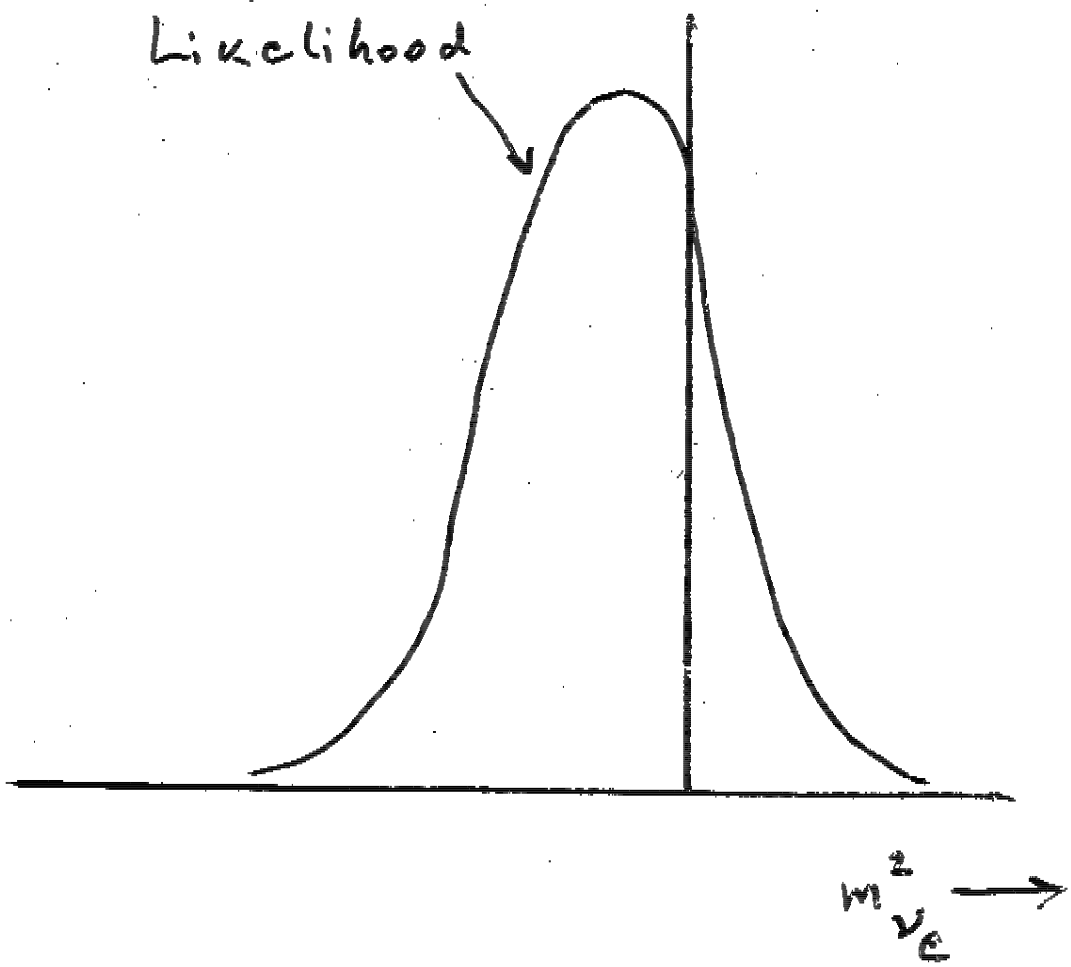


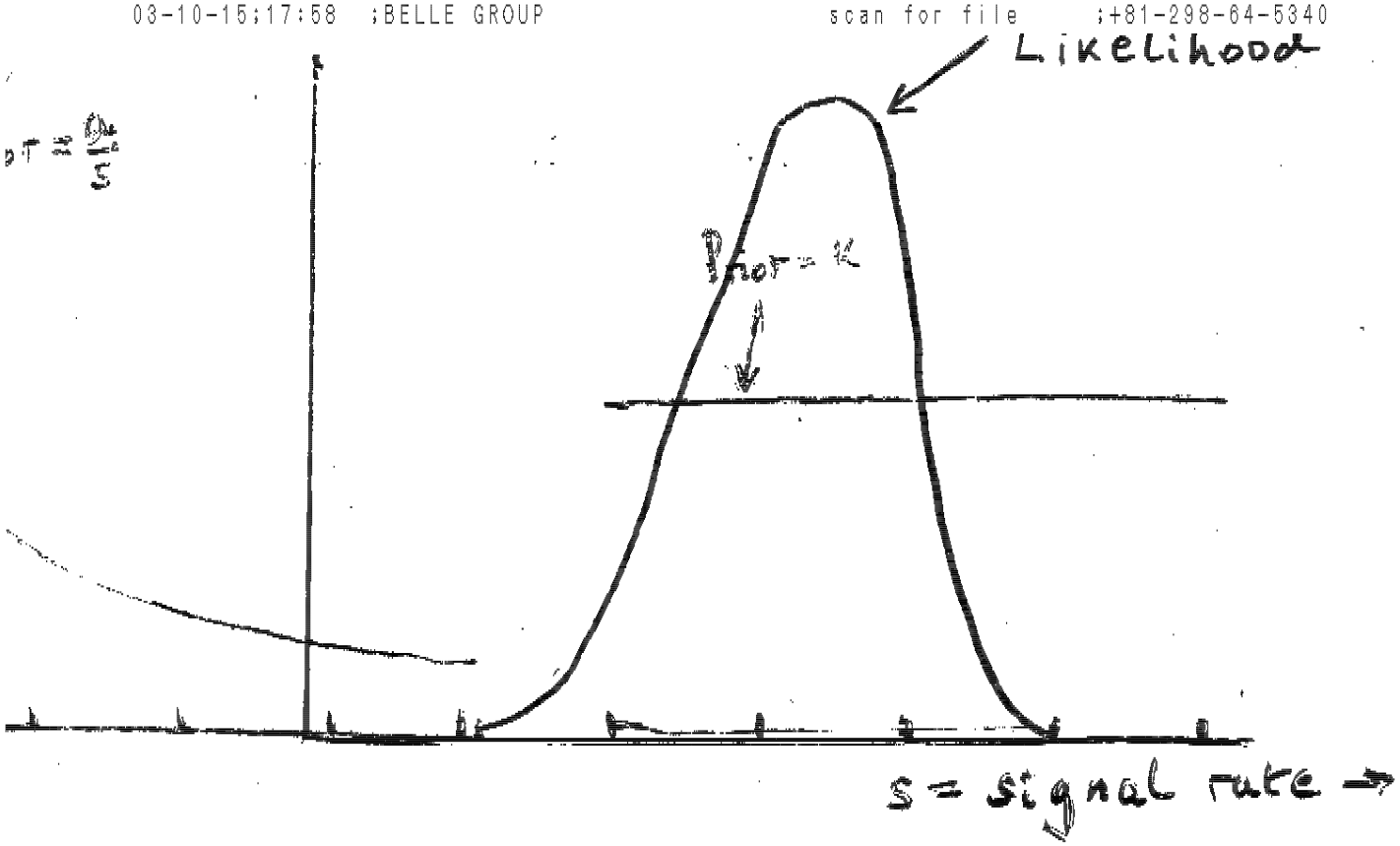
03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 9/ 54

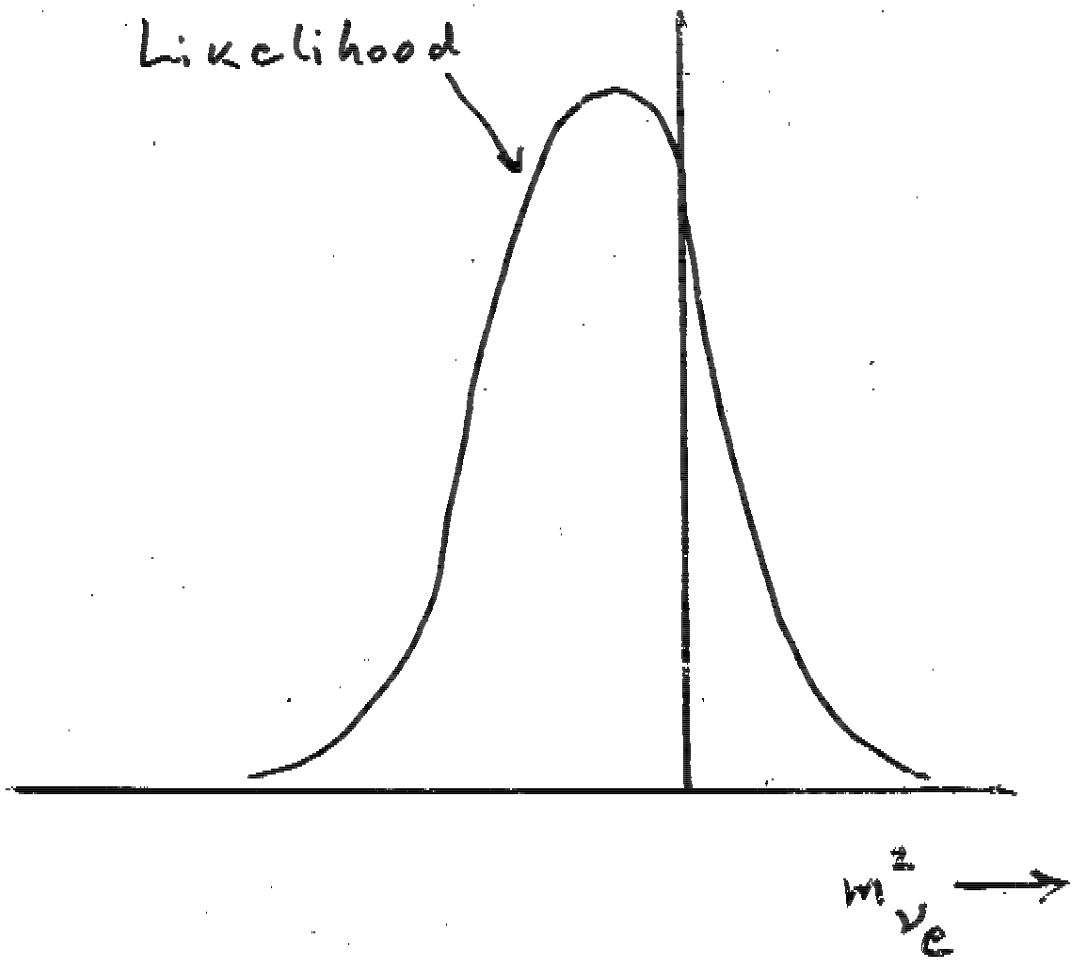


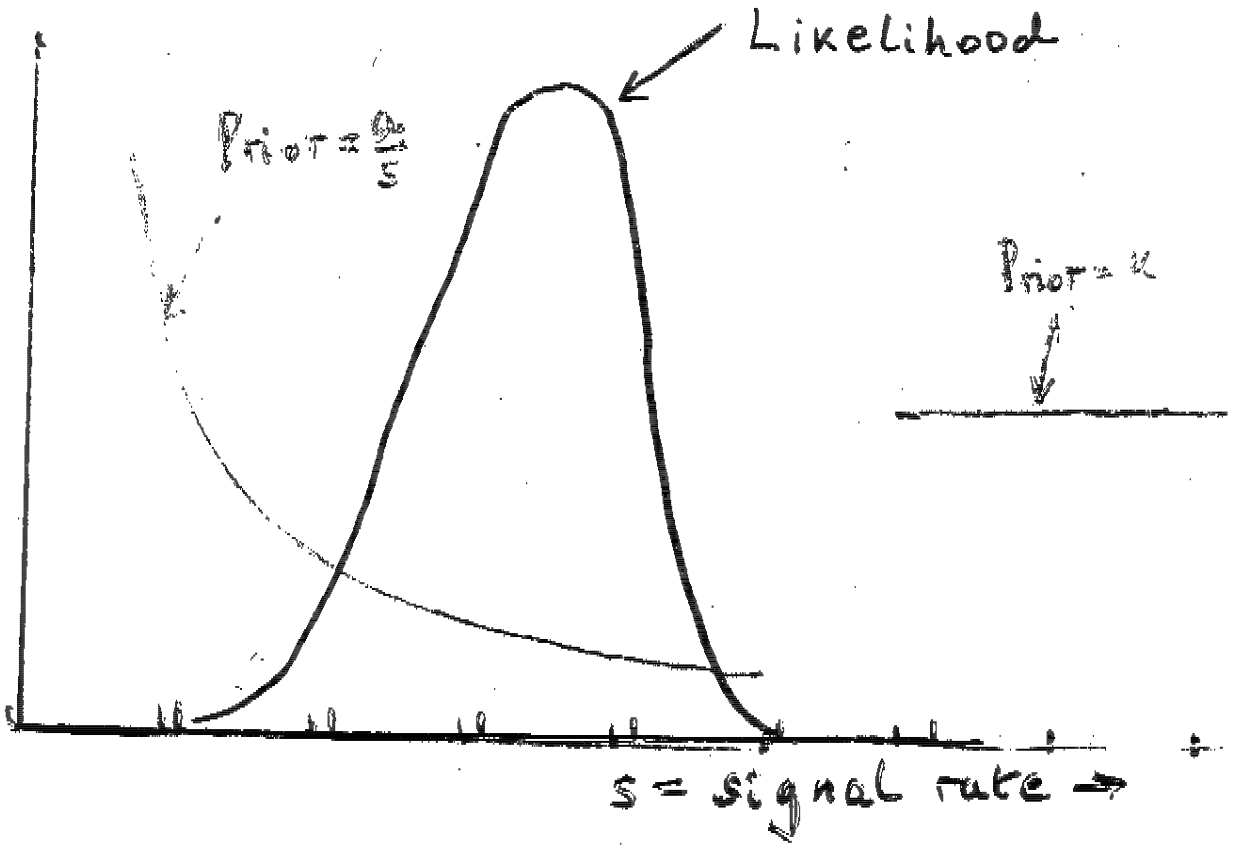

Data overshadows the Prior

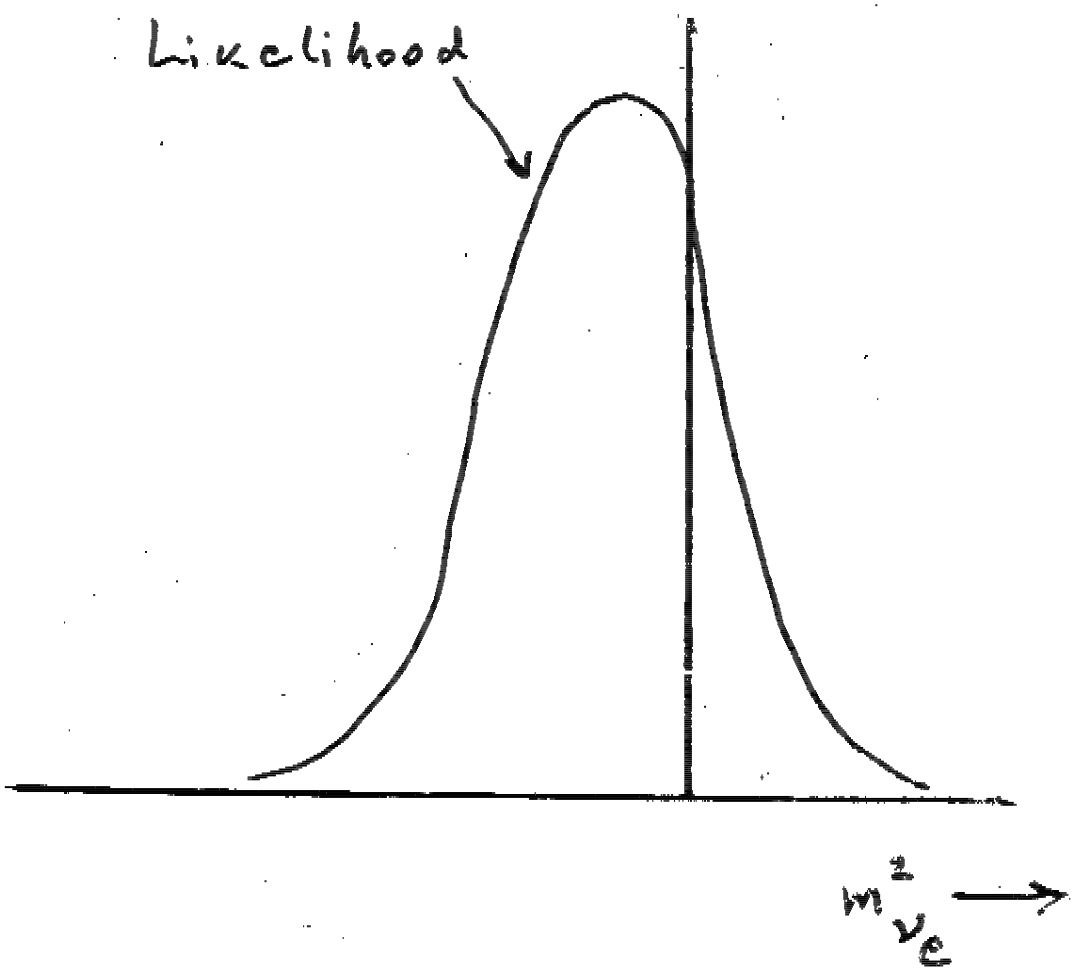

Likelihood
s = signal rate →

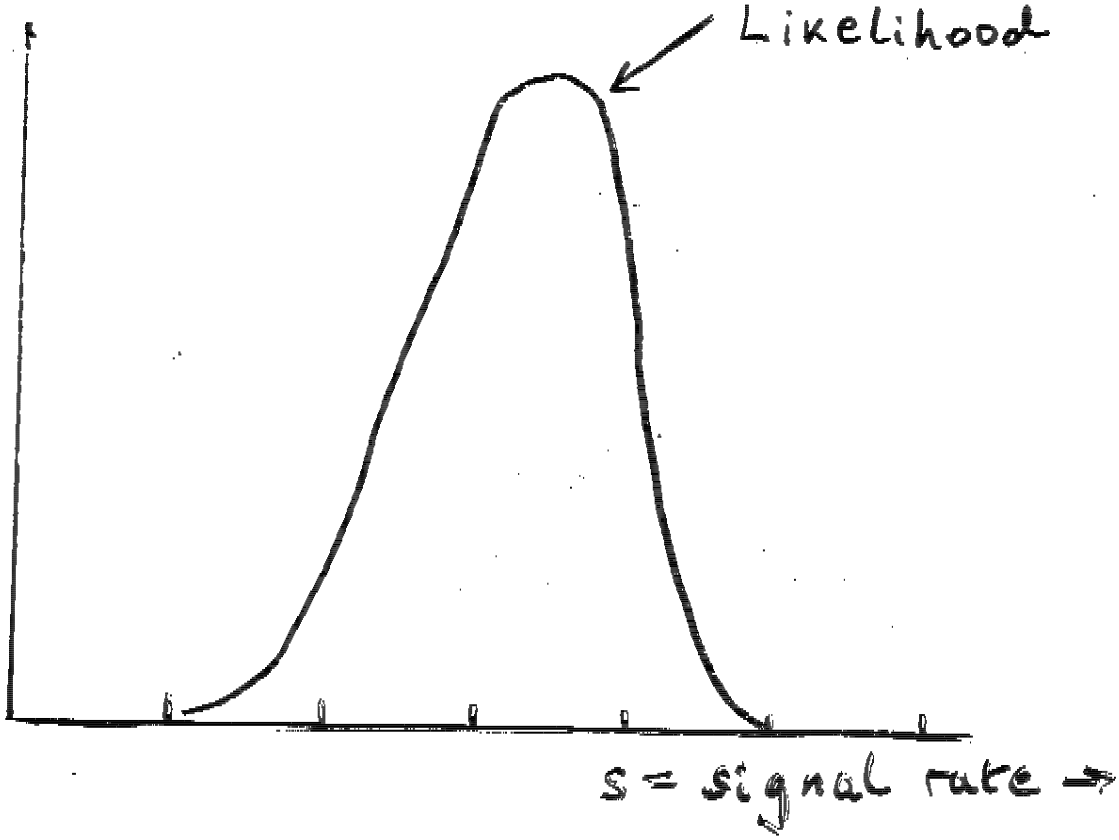

Likelihood
$m^2_{\nu_e}$ →

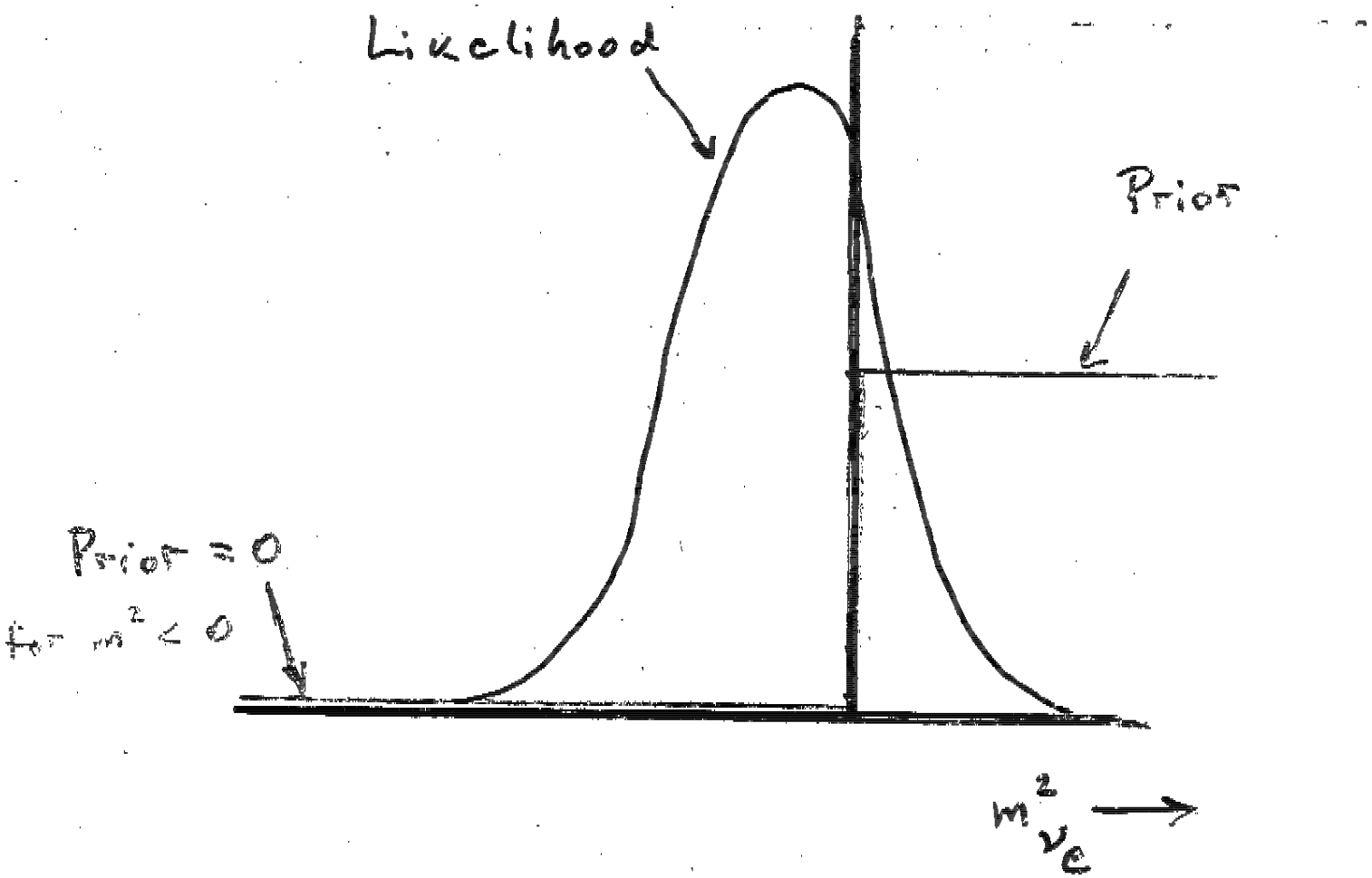

Likelihood
Prior = K
s = signal rate →


Likelihood
$m^2_{\nu_e}$ →

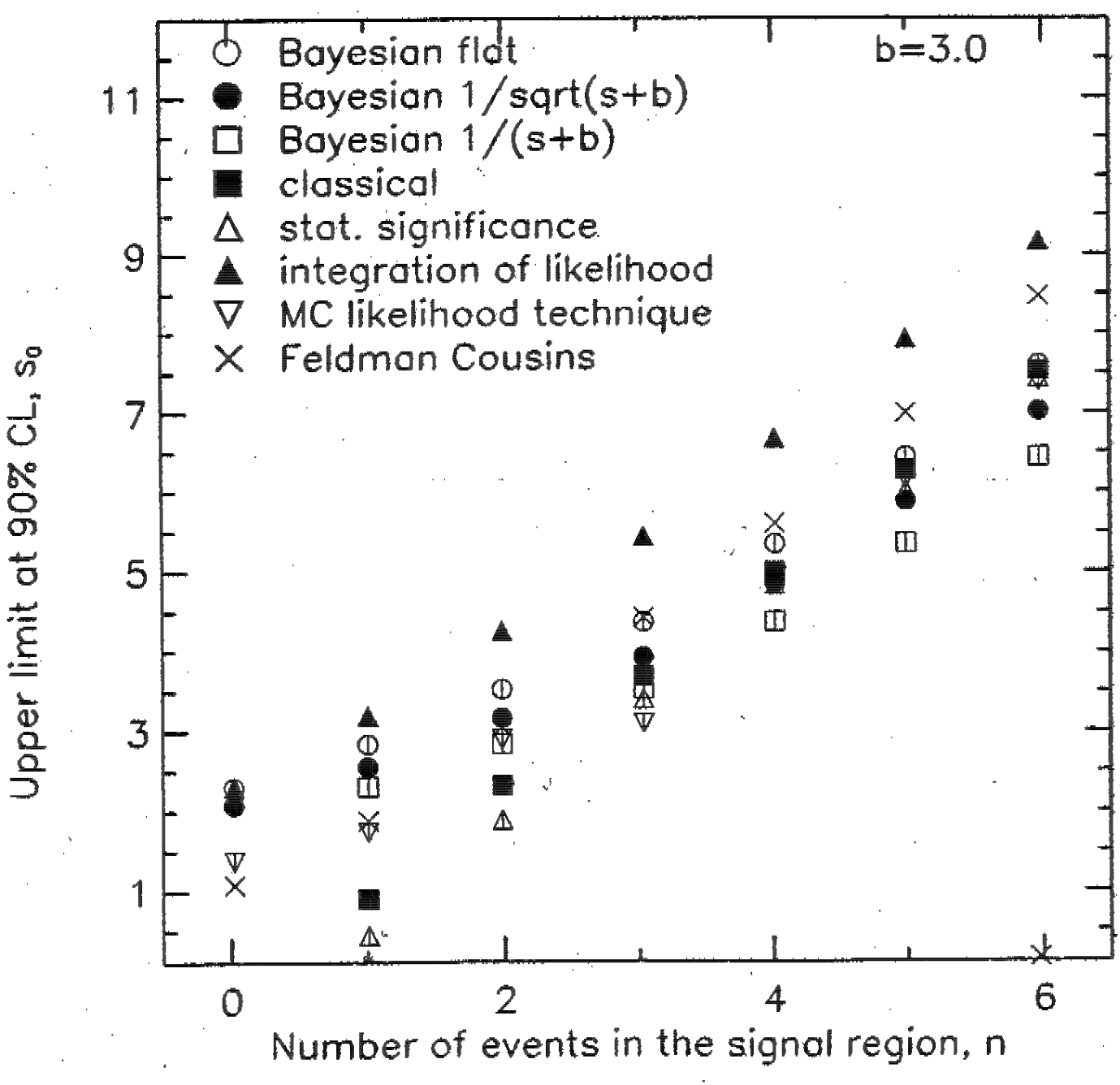

Likelihood
Prior = $\frac{1}{s}$
Prior = 1
s = signal rate →


Likelihood
$m^2_{\nu_e}$ →


Likelihood
s = signal rate →


Likelihood
Prior
Prior = 0
for $m^2 < 0$
$m^2_{\nu_e}$ →

ILYA NARSKY
FNAL CL Workshop



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

# 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"



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 16/ 54

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

Theory = male or female

Data = pregnant or not pregnant

$P(\text{pregnant} ; \text{female}) \sim 3\%$

but

$P(\text{female} ; \text{pregnant}) >>> 3\%$



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 17/ 54

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)



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 18/ 54

# 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

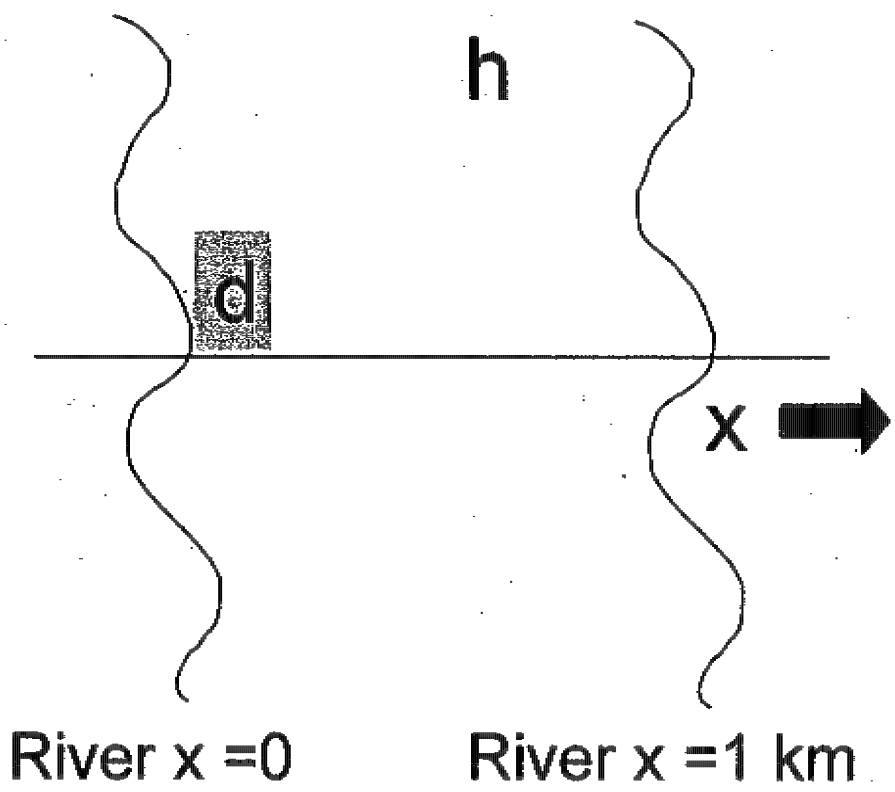
If pure $\pi$ beam, ~ 0



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 19/ 54

# 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






03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 20/ 54

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



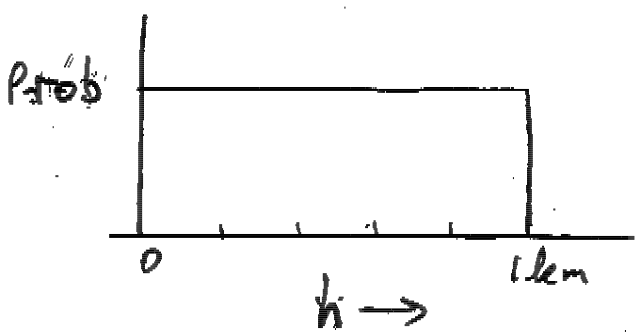
03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 21/ 54

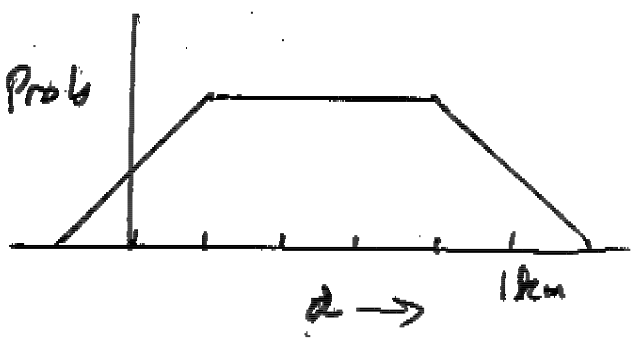
Example:

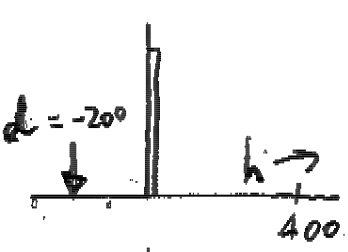
1) More specific on statement (1):

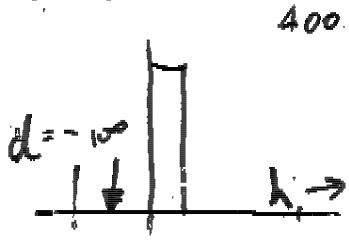
$$\text{Prob}(d - h) = \begin{cases} \text{Const.} & \text{for } |d-h| < 200\,m \\ 0 & \text{for } |d-h| > 200\,m \end{cases} \quad [L' \text{Hood}]$$

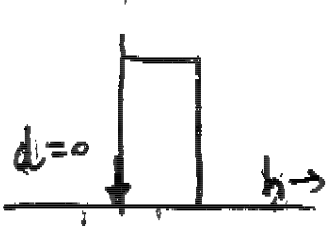
2) Hunter $h$ uniform in $0 \to 1\,km$ [PRIOR]



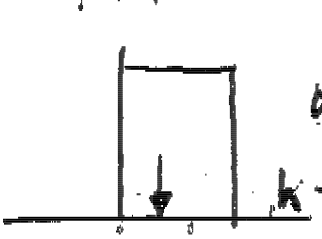



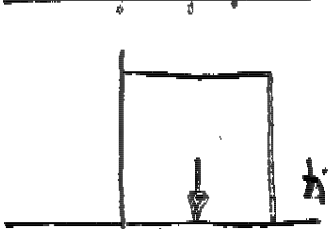



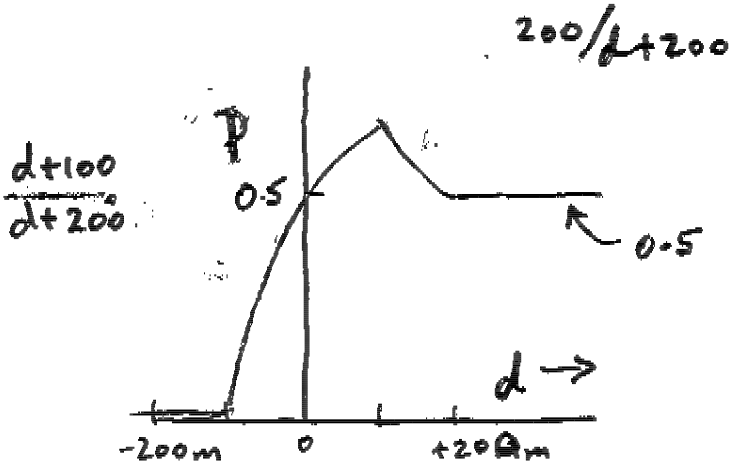




Prob $|h-d| \le 100$ low 50%




Prob $|h-d| \le 100$ above 50%




$P$ = prob $|h-d| \le 100\,m$

# 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 $\mu_1 \rightarrow \mu_2$



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 # 23/ 54
scan for file

CLASSICAL (NEYMAN) CONFIDENCE INTERVALS

? (data ! theory)

FIGURES

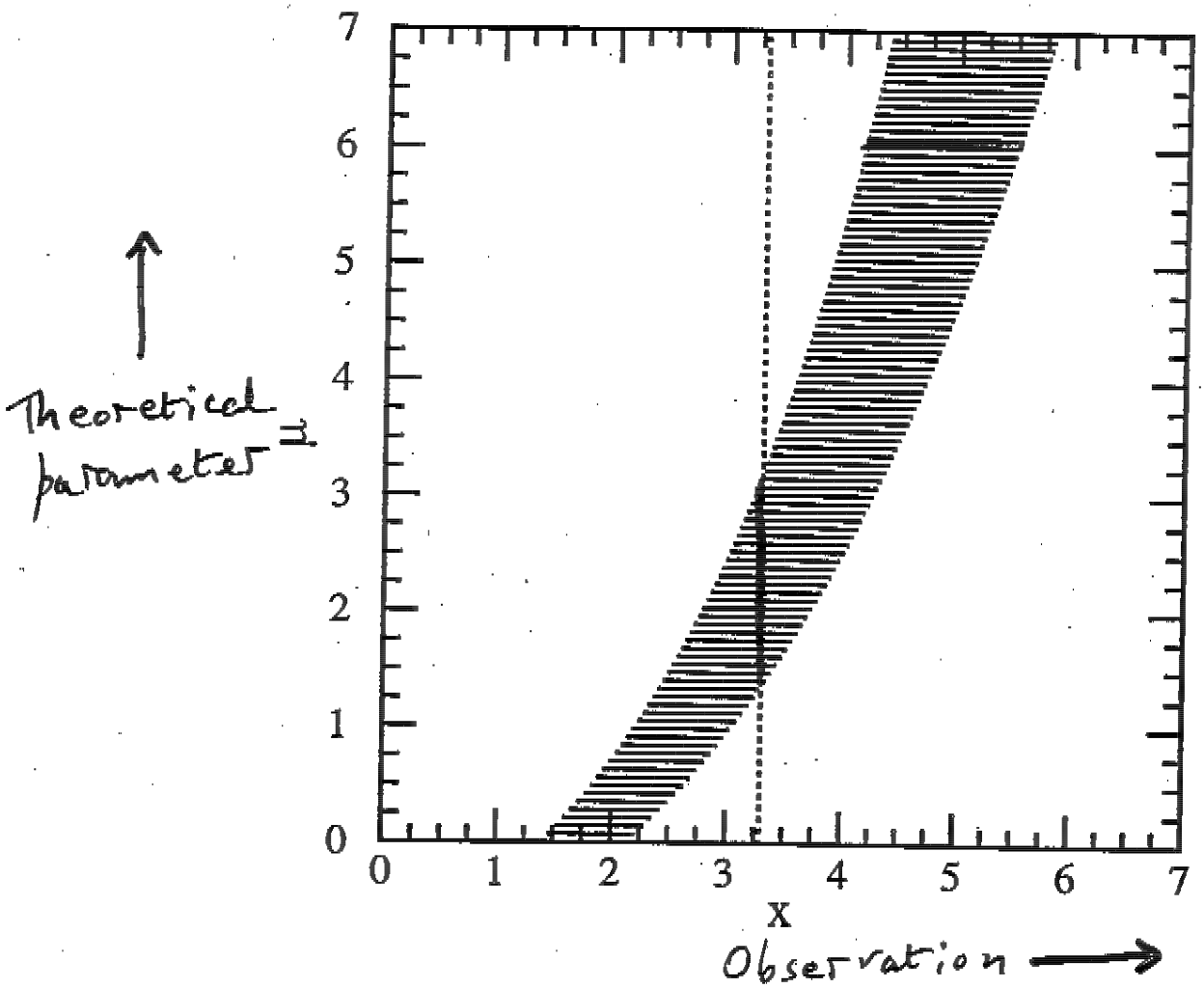



FIG. 1. A generic confidence belt construction and its use. For each value of $\mu$, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2]|\mu) = \alpha$. Upon performing an experiment to measure $x$ and obtaining the value $x_0$, one draws the dashed vertical line through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



BF8 8

# COVERAGE

If true for all $\mu$ : "correct coverage"

P< $\alpha$ for some $\mu$: "undercoverage"
(this is serious !)

P> $\alpha$ for some $\mu$: "overcoverage"

Conservative

Loss of rejection power



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 # 25/ 54
scan for file

$\mu_l \leq \mu \leq \mu_u$ 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 $\mu$



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 26/ 54

# 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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 27/ 54

90% classical interval for Gaussian

$\sigma = 1$

$\mu \geq 0$

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

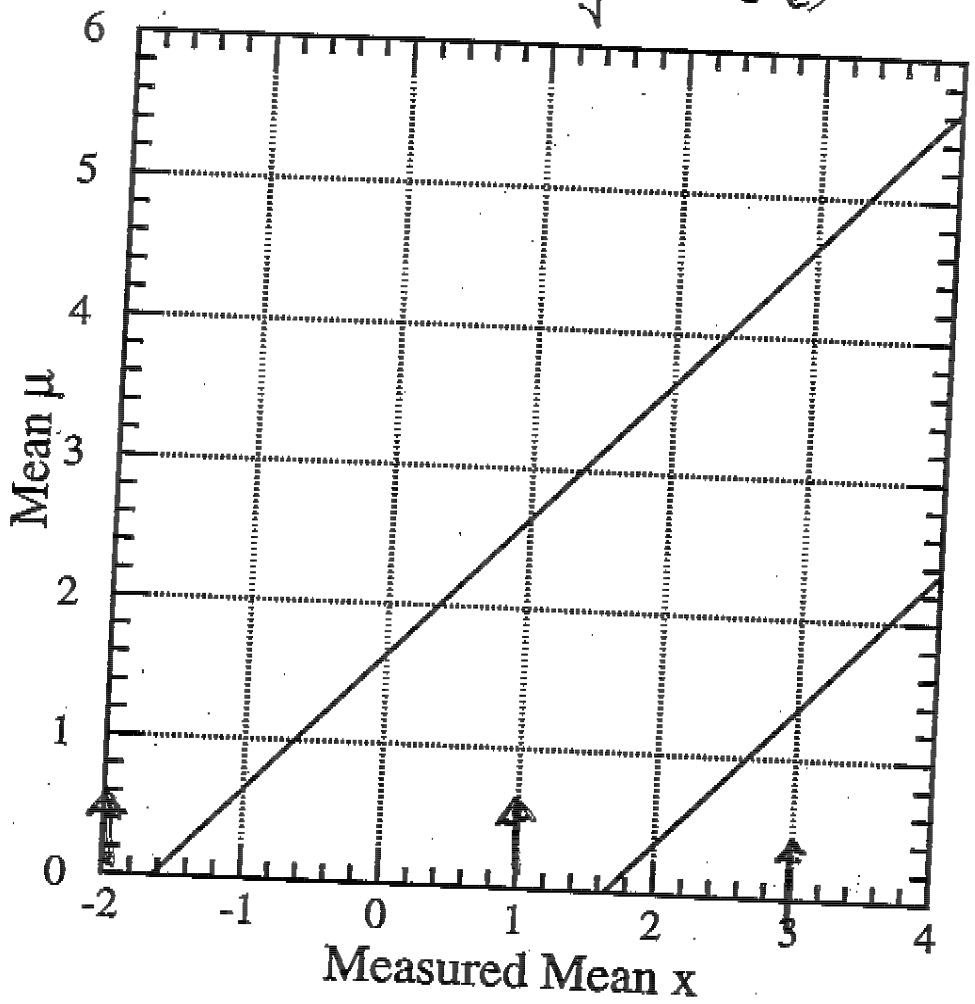



FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$x_{obs} = 3$ Two sided limit

$x_{obs} = 1$ Upper limit

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



RE.10

17

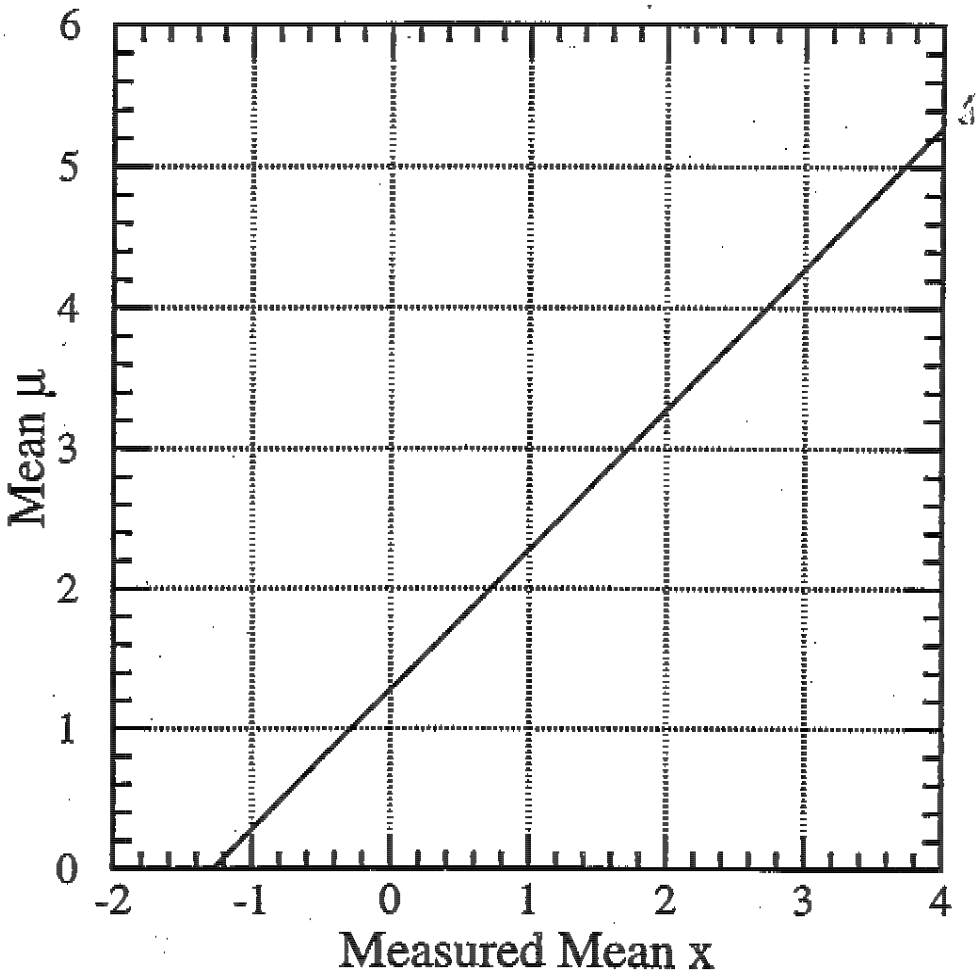



FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, in units of the rms deviation. The second line in the belt is at $x = +\infty$.

Feldman-Cousins 90% Conf interval

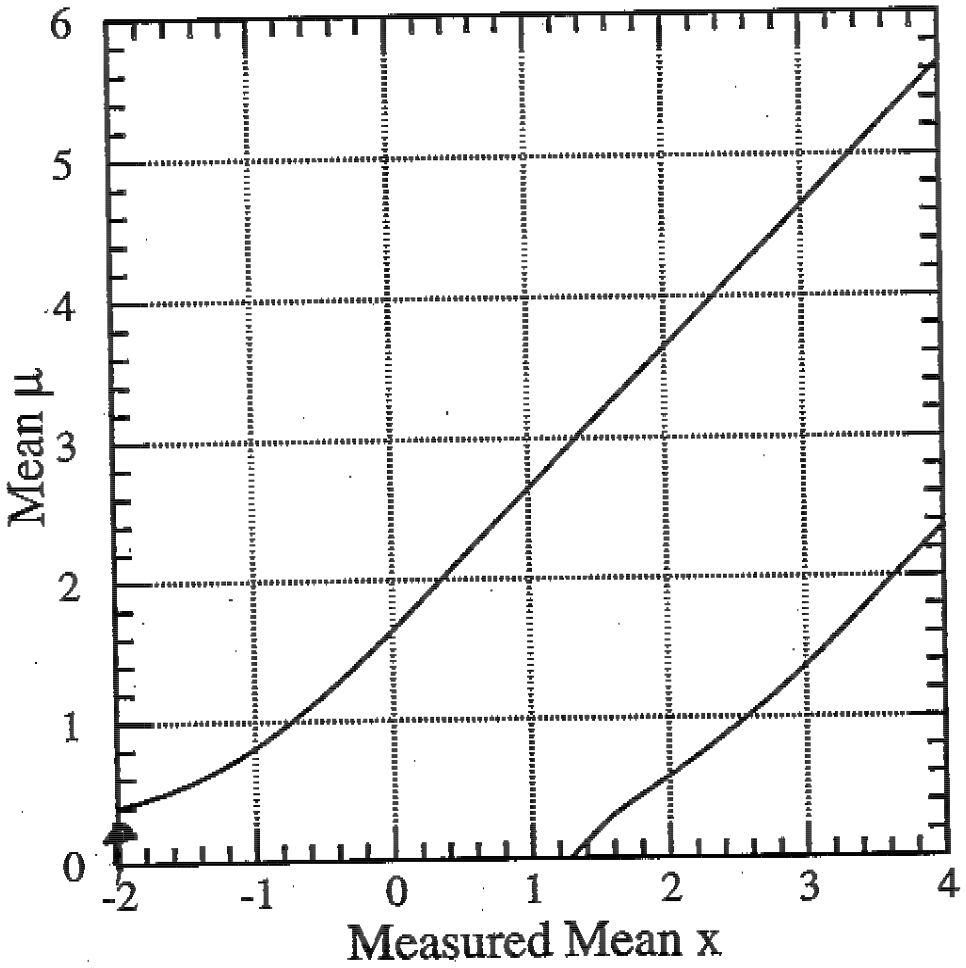



FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

$x_{obs} = -2$. Now give upper limit

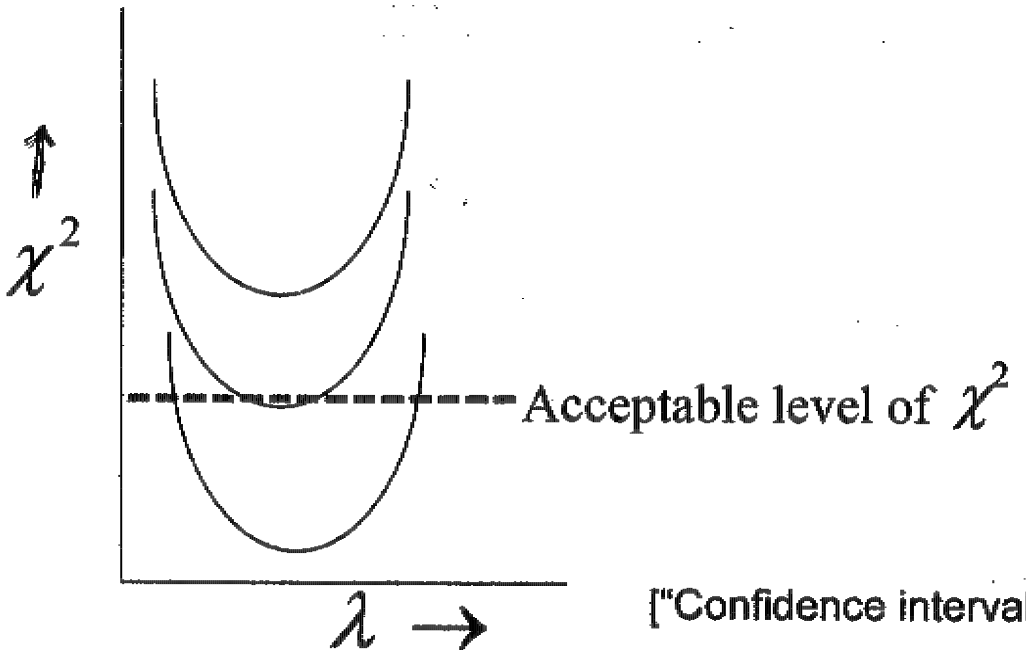
# Importance of Ordering Rule

Neyman construction in 1 parameter $\mu$

2 measurements $X_1$ $X_2$

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

An aside: Determination of single parameter $\lambda$ 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) $\chi^2(\lambda) < \chi^2_{min}(\lambda) + 1$

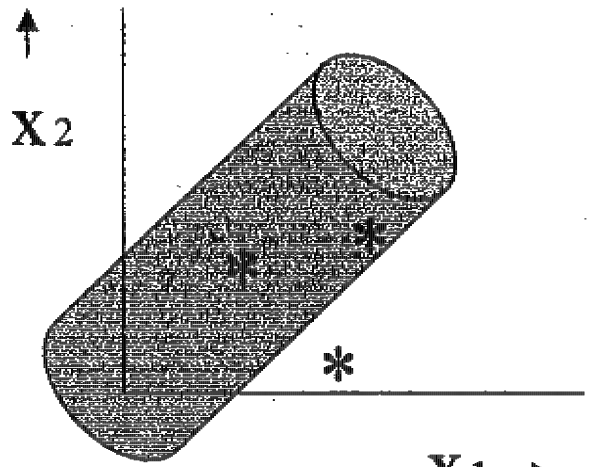
2) is good

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

["Confidence interval coupled to goodness of fit"]



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 31/ 54




# 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



03-10-15;17:58 ;BELLE GROUP
scan for file
;+81-298-64-5340
# 32/ 54

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:

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

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

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:

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



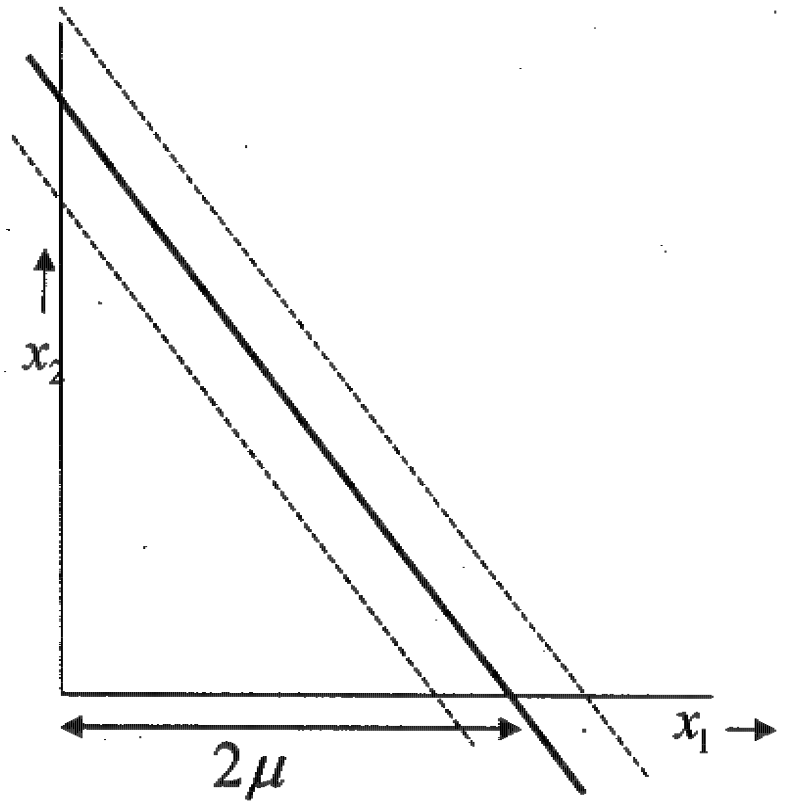
03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 33/ 54

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

Therefore, range of $\mu$ is

Constant Width

Independent of $x_1 - x_2$




Confidence Range and Goodness of Fit are completely decoupled



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 34/ 54

# Bayesian

Pros:

Easy to understand

Physical Interval

Cons:

Needs prior

Hard to combine

Coverage



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 # 35/ 54
scan for file

# Standard Frequentist

Pros:

Coverage

Cons:

Hard to understand

Small or Empty Intervals

Different Upper Limits



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 # 36/ 54
scan for file

# SYSTEMATICS

For example

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

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

Physics parameter

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

Uncertainties → error in $\sigma$

Some are arguably statistical errors

$$LA = LA_{\circ} \pm \sigma_{LA}$$

$$b = b_{\circ} \pm \sigma_{b}$$

Shift Central Value

Bayesian

Frequentist

Mixed



03-10-15;17:58 ;BELLE GROUP
scan for file ;+81-298-64-5340 # 37/ 54

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

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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 38/ 54

# Bayesian

Without systematics

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

prior

With systematics

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

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

Then integrate over LA and b

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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 39/ 54

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

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

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

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



03-10-15;17:58 ;BELLE GROUP ;+81-298-64-5340 # 40/ 54
scan for file

# Frequentist

Full Method

Imagine just 2 parameters $\sigma$ and LA

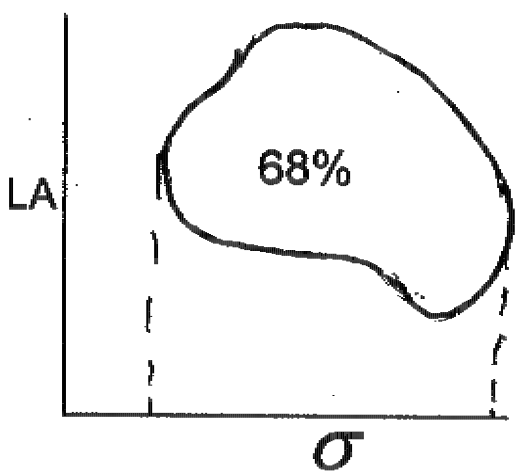
and 2 measurements N and M

↑ Physics ↑ Nuisance

Do Neyman construction in 4-D

Use observed N and M, to give

Confidence Region






03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 41/ 54

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

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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 42/ 54

# Full frequentist method hard to apply in several dimensions

Used in $\leq 3$ parameters

For example: Neutrino oscillations (CHOOZ)

$$\sin^2 2\theta \,, \Delta m^2$$

Normalisation of data

Use approximate frequentist methods that reduce dimensions to just physics parameters

e.g. Profile pdf

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

Contrast Bayes marginalisation

Distinguish "profile ordering"



Properties being studied by Giovanni Punzi

03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 43/ 54

Talks at FNAL CONFIDENCE LIMITS WORKSHOP

(March 2000) by:

Gary Feldman

Wolfgang Rolke hep-ph/0005187 version 2

Acceptance uncertainty worse than Background uncertainty

$$\text{Limit of C.L. as } \sigma \rightarrow 0$$

$$\neq \text{C.L. for } \sigma = 0$$

Need to check Coverage



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 44/ 54

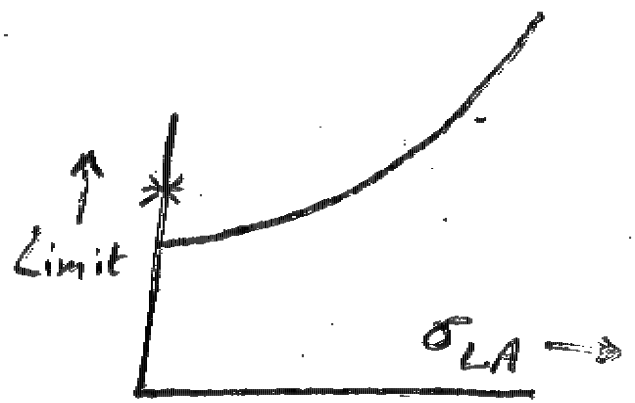
# Method: Mixed Frequentist - Bayesian

Bayesian for nuisance parameters and

Frequentist to extract range

Philosophical/aesthetic problems?

Highland and Cousins




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



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 45/ 54

# 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 |
| Problem 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 |



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 46/ 54

# 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 |



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 47/ 54

# 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"



03-10-15;17:58 ;BELLE GROUP scan for file ;+81-298-64-5340 # 48/ 54

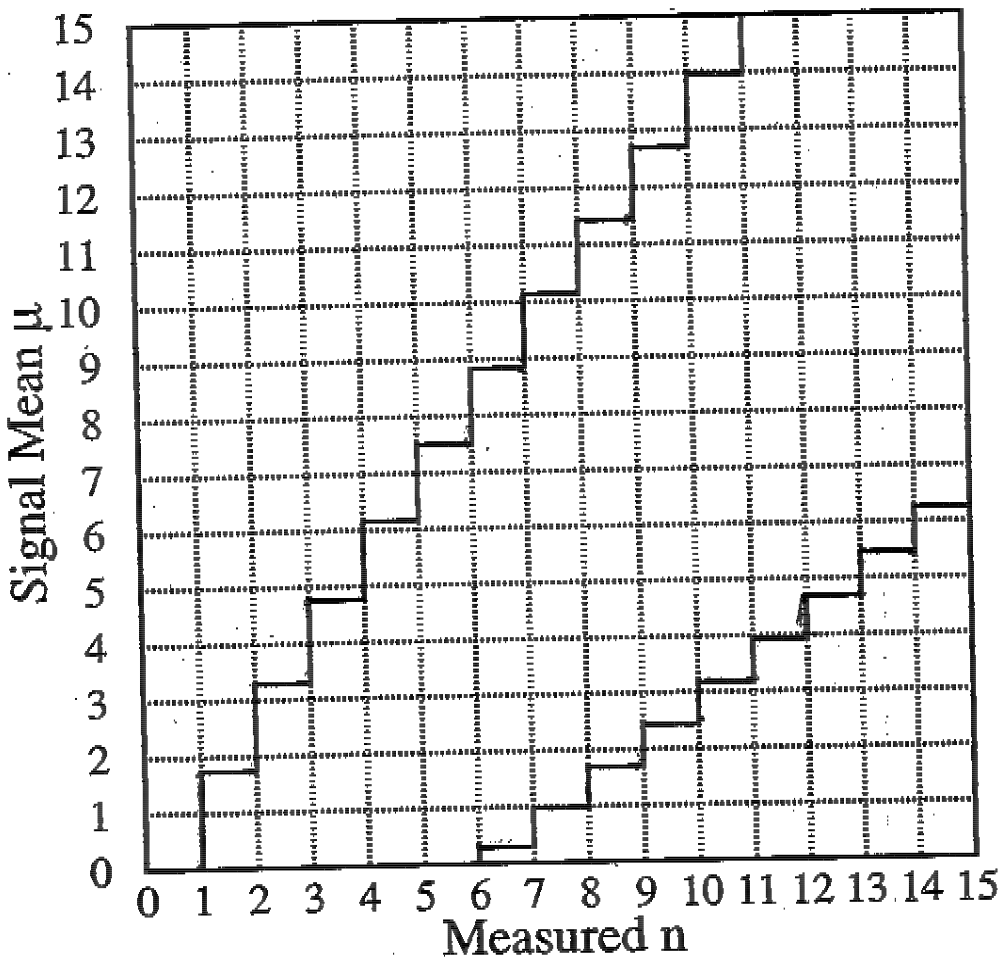
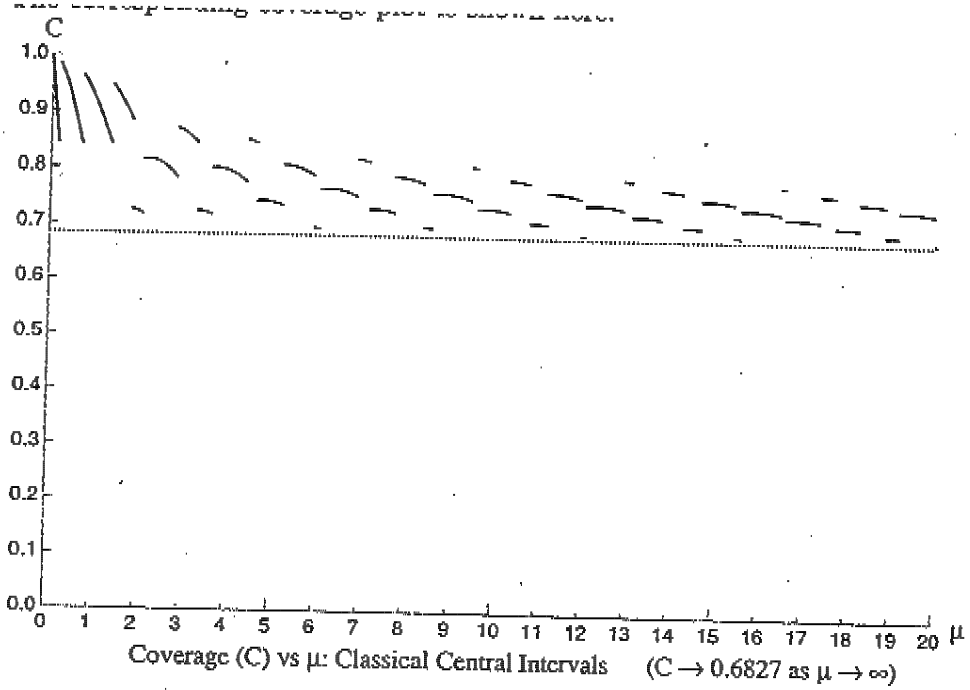



FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean $\mu$ in the presence of Poisson background with known mean $b = 3.0$.

Standard Frequentist for Poisson mean, ll.

# COVERAGE OF ERROR BARS FOR POISSON DATA

Joel Heinrich
CDF 6438

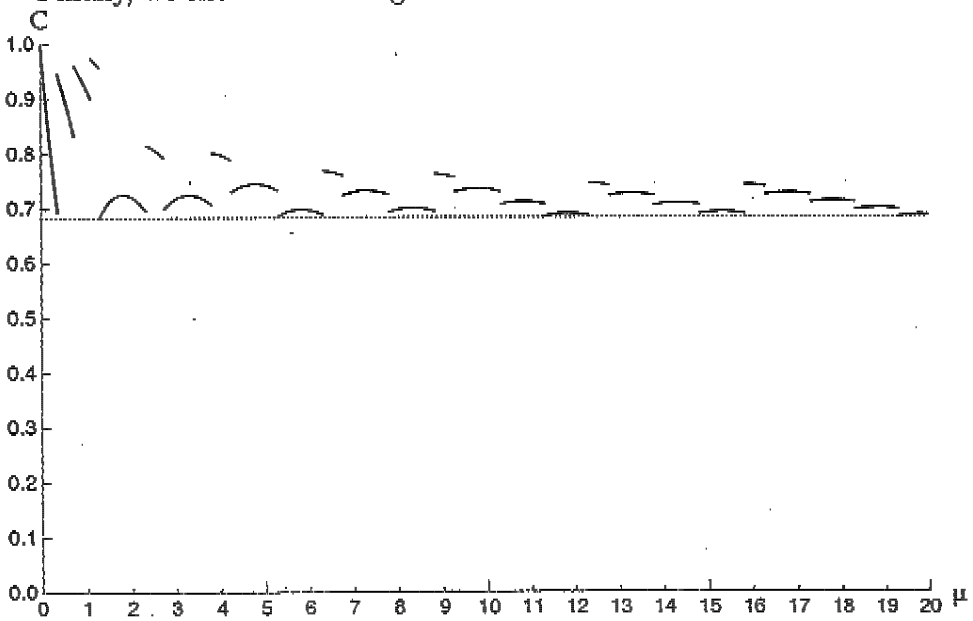



Coverage (C) vs μ: Classical Central Intervals ($C \to 0.6827$ as $\mu \to \infty$)

$$P(n,\mu) = e^{-\mu}\mu^n/n!$$

Classical central intervals at 68.3% coverage

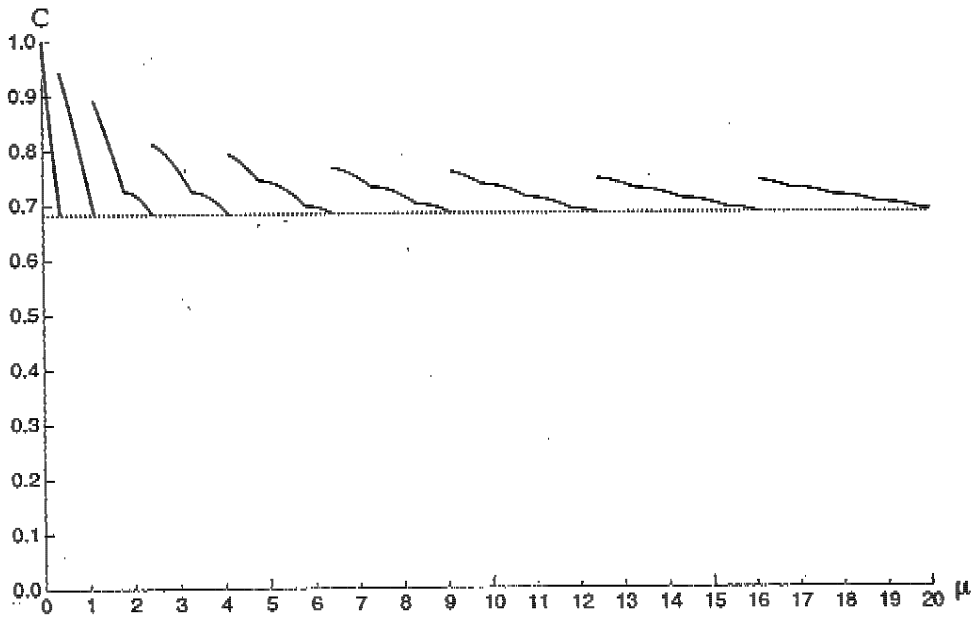
Finally, we show the coverage of the $1\sigma$ unified intervals:




Coverage (C) vs μ: Unified Intervals ($C \to 0.6827$ as $\mu \to \infty$)

$P(n, \mu) = e^{-\mu} \mu^n / n!$

Unified intervals → 68.3% coverage



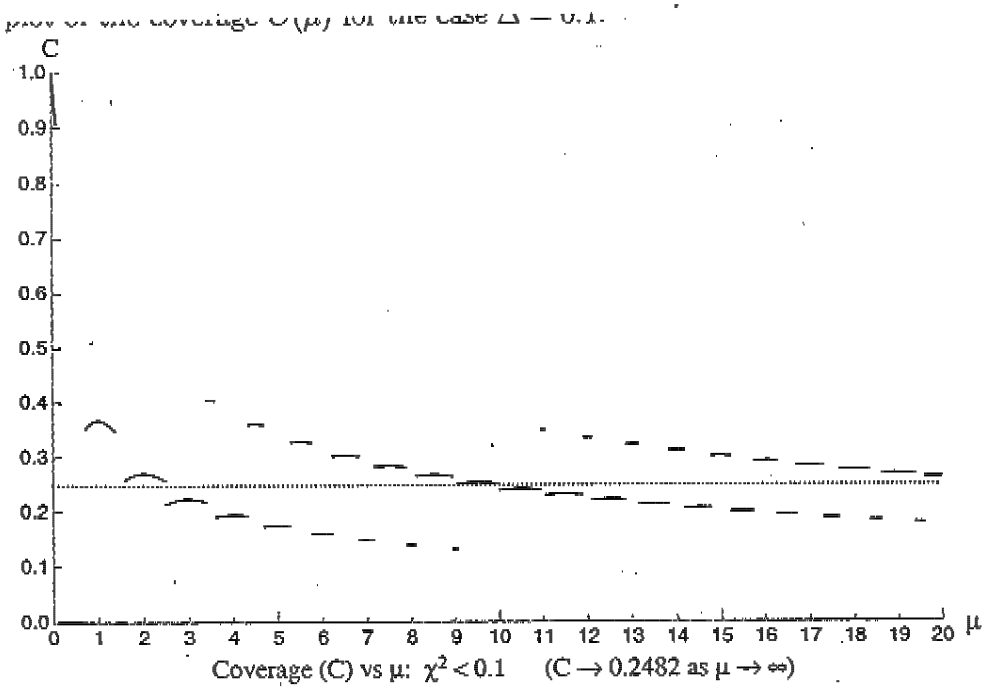

Coverage (C) vs μ: Probability Ordering Intervals ($C \to 0.6827$ as $\mu \to \infty$)

$$P(n,\mu) = e^{-\mu}\mu^n/n!$$

Probability ordering intervals

at 68.3% coverage

COVERAGE OF ERROR BARS FOR POISSON DATA — JOEL HEINRICH CDF 6438



Coverage (C) vs μ: $\chi^2 < 0.1$ ($C \to 0.2482$ as $\mu \to \infty$)

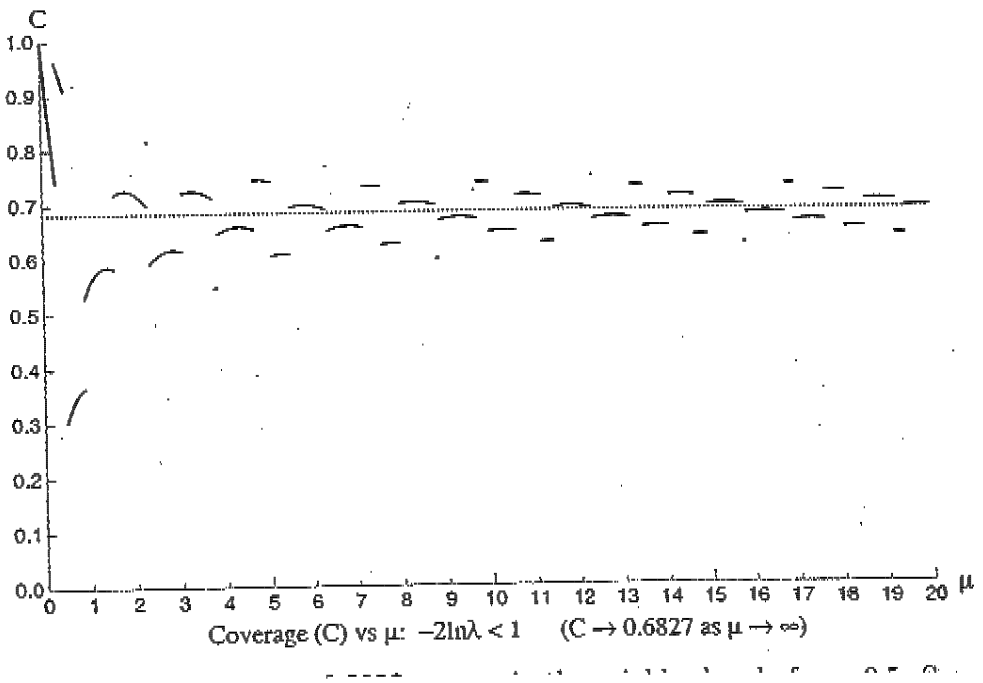
3

$$P(n,\mu) = e^{-\mu}\mu^n/n!$$

$$\chi^2 = \left(\frac{n-\mu}{\sqrt{\mu}}\right)^2$$

$$\Delta\chi^2 = 0.1 \Longrightarrow 24.8\% \text{ coverage}$$

NOT FREQUENTIST



Coverage (C) vs μ: $-2\ln\lambda < 1$ ($C \to 0.6827$ as $\mu \to \infty$)

$$P(n,\mu) = e^{-\mu}\mu^n / n!$$

$$-2\ln\lambda < 1$$

$$\left[\lambda = P(n,\mu) / P(n,\mu_{best}\right]$$