FNAL
27-28.3.00

CL2K

Are we $\gtrsim 95\%$ certain that we're on the right track?

I. Incomplete summary

II. My 2 ¢

Glen Cowan
Royal Holloway College, Univ. London

28.3.00

Glen Cowan
18.1.00

My Paper

by Joe Author

I. Introduction

blah blah

II. Method

blah blah blah blah

III. Results

classical statistics here

IV. Discussion & conclusions

Bayesian statistics here

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

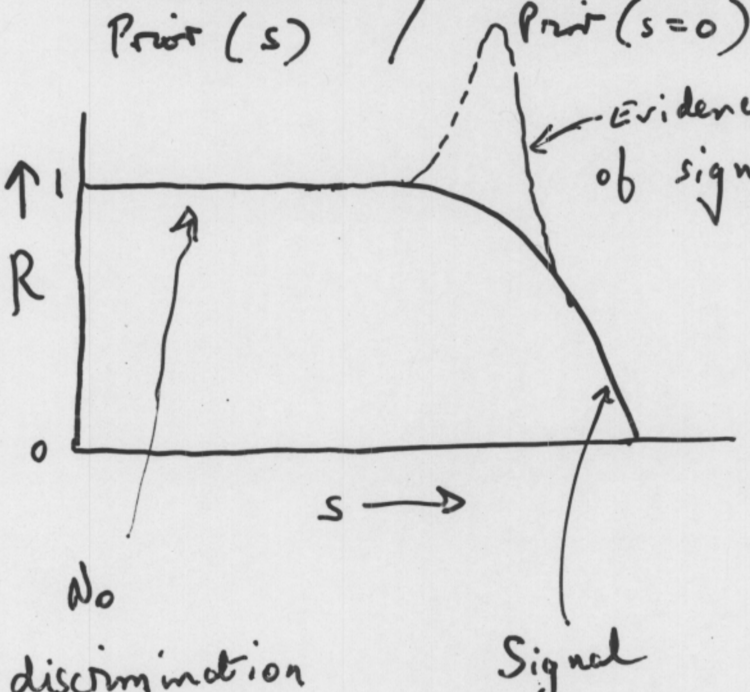
Uses in Higgs limit [d'A + Degrassi]

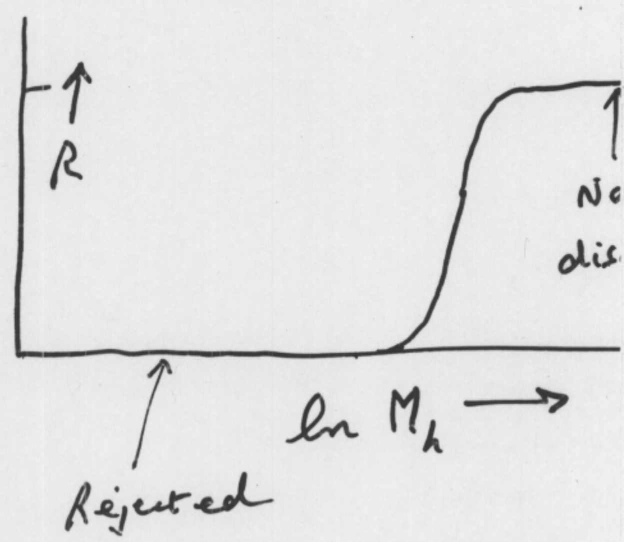
$$\frac{P(s_1 | n_{obs}, b)}{P(s_2 | n_{obs}, b)} = \frac{P(n_{obs} | s_1, b)}{P(n_{obs} | s_2, b)} \times \frac{\text{Prior}(s_1)}{\text{Prior}(s_2)}$$

$\mathcal{L}$ ratio

Rewrite as

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





R(s) is indep. of prior

From CERN workshop (L. Lyons)

→ report likelihood function $\mathcal{L}(\vec{x}|\theta)$

data ↗ ↑ param.

Or equivalent, e.g. $R = \dfrac{\mathcal{L}(\vec{x}|\theta)}{\mathcal{L}(\vec{x}|\theta=0)}$ (D'Agostini → plot

<u>Why?</u> So that a Bayesian consumer can compute:

consumer's (subjective) prior ↙

$$p(\theta|\vec{x}) \propto \mathcal{L}(\vec{x}|\theta)\pi(\theta)$$

Combining experiments: $\mathcal{L}(\theta) = \prod_i \mathcal{L}_i(\theta)$

↳ $\hat{\theta}_{ML}$

- exact choice of model? (Bill Murray)
- systematics?
- $\mathcal{L}(\theta)$ not enough to compute Neyman conf. int. in small sample case.

The Four Commandments (Fred James

I Physicists should learn statistics vocab.

p-value, marginalize, ...

II State all assumptions, methods etc. in publication.

(and give an unbiased estimator $\hat{\theta}$, even if in unphysical region)

III F-C in all searches

comments to follow . . .

IV Bayesian decision theory in policy decisions

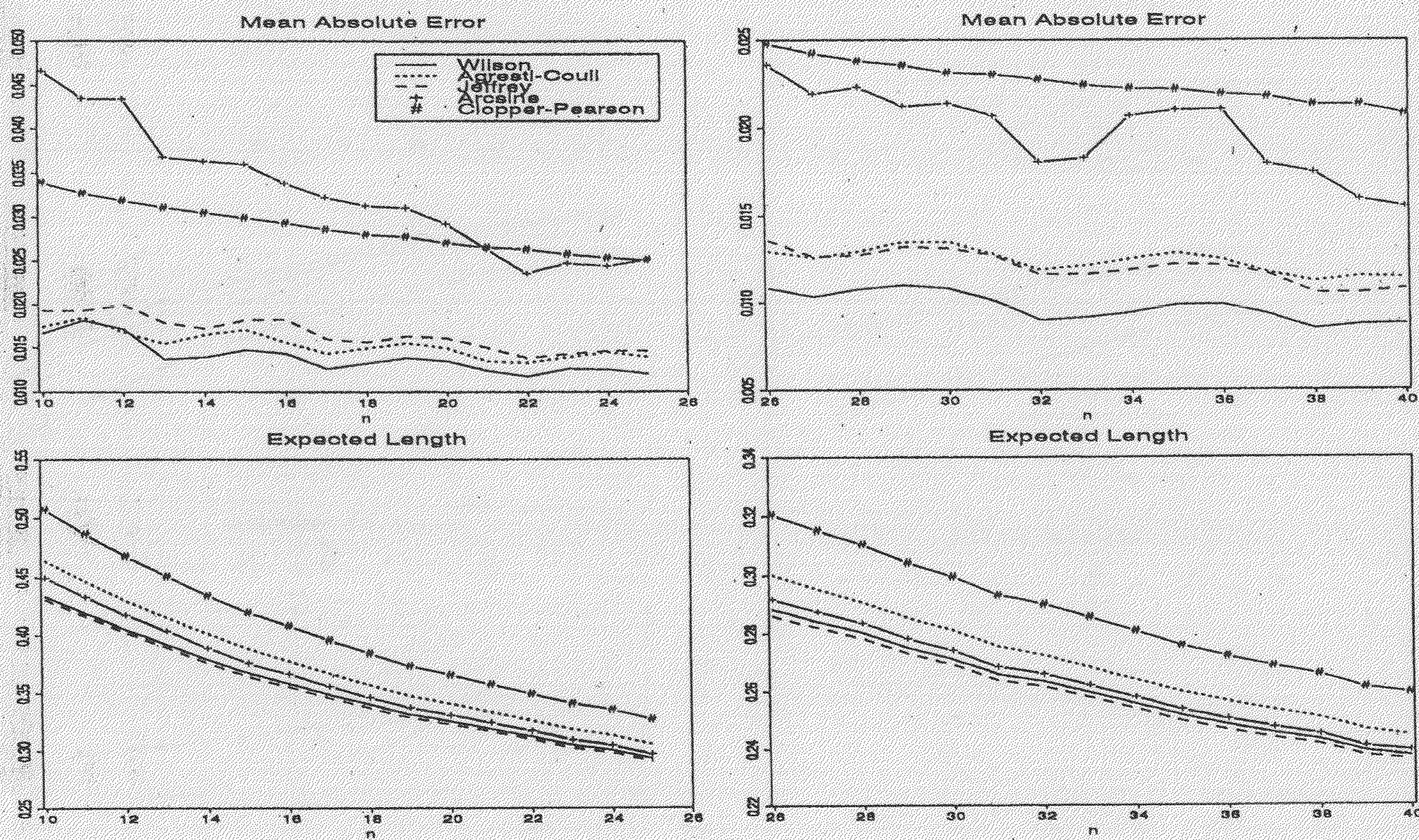
and in discussion of results

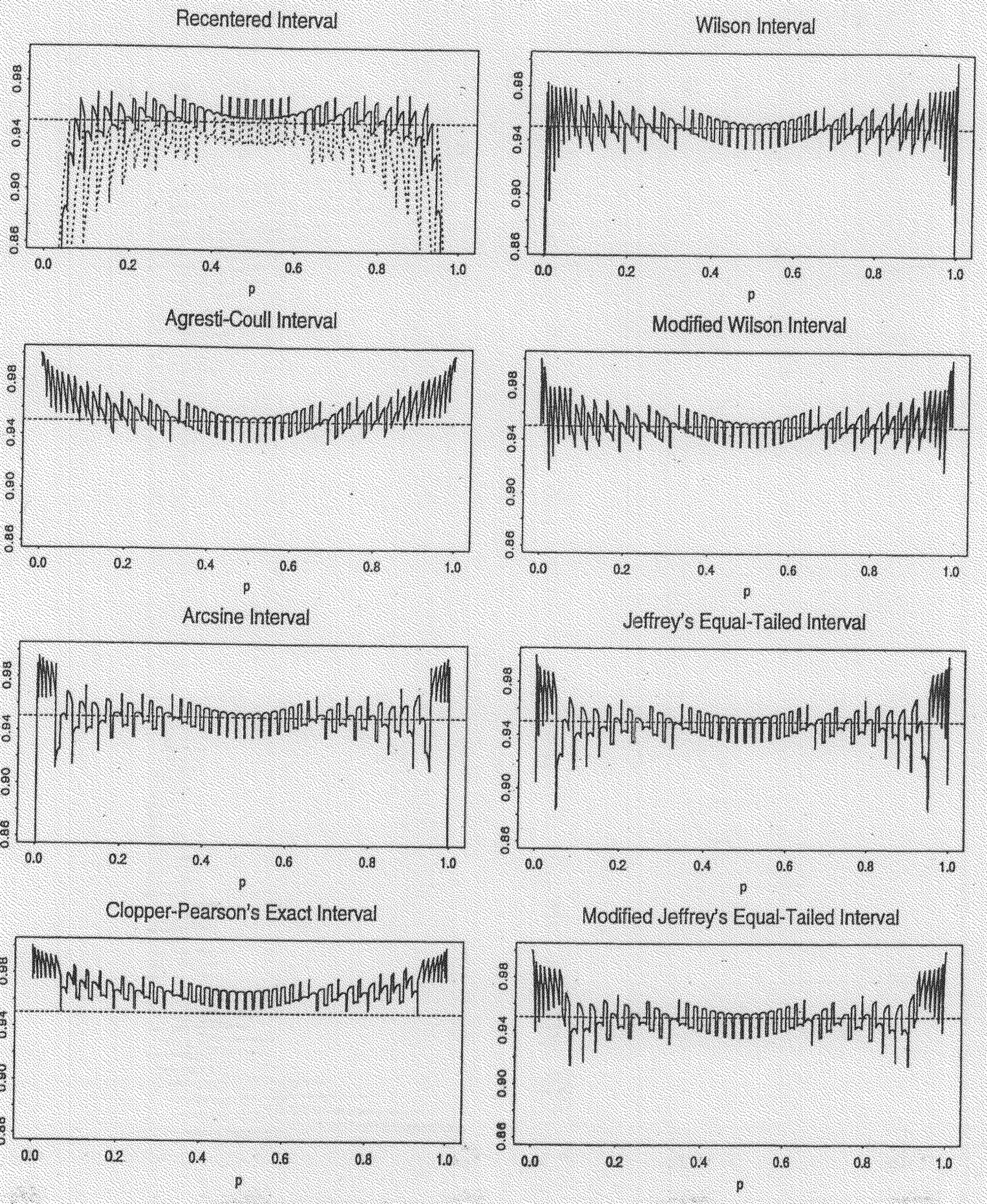
Subjective Bayes

$$p(\theta | \vec{x}) = \text{degree of belief}$$

$$\propto \mathcal{L}(\vec{x} | \theta)\, \pi(\theta)$$

- By itself, a bad way to report result of an observation (Reindeer problem)
- Good for discussion / conclusions section of paper.
- It's how the consumer will digest the result, using his / her own prior. (Explicitly or otherwise.)


Mean Absolute Error
Wilson
Agresti-Coull
Jeffrey
Arcsine
Clopper-Pearson
n
Mean Absolute Error
n
Expected Length
n
Expected Length
n



Recentered Interval
Wilson Interval
Agresti-Coull Interval
Modified Wilson Interval
Arcsine Interval
Jeffrey's Equal-Tailed Interval
Clopper-Pearson's Exact Interval
Modified Jeffrey's Equal-Tailed Interval
0.86
0.90
0.94
0.98
0.0
0.2
0.4
0.6
0.8
1.0
p

Frequentist Coverage of Textbook Confidence Interval

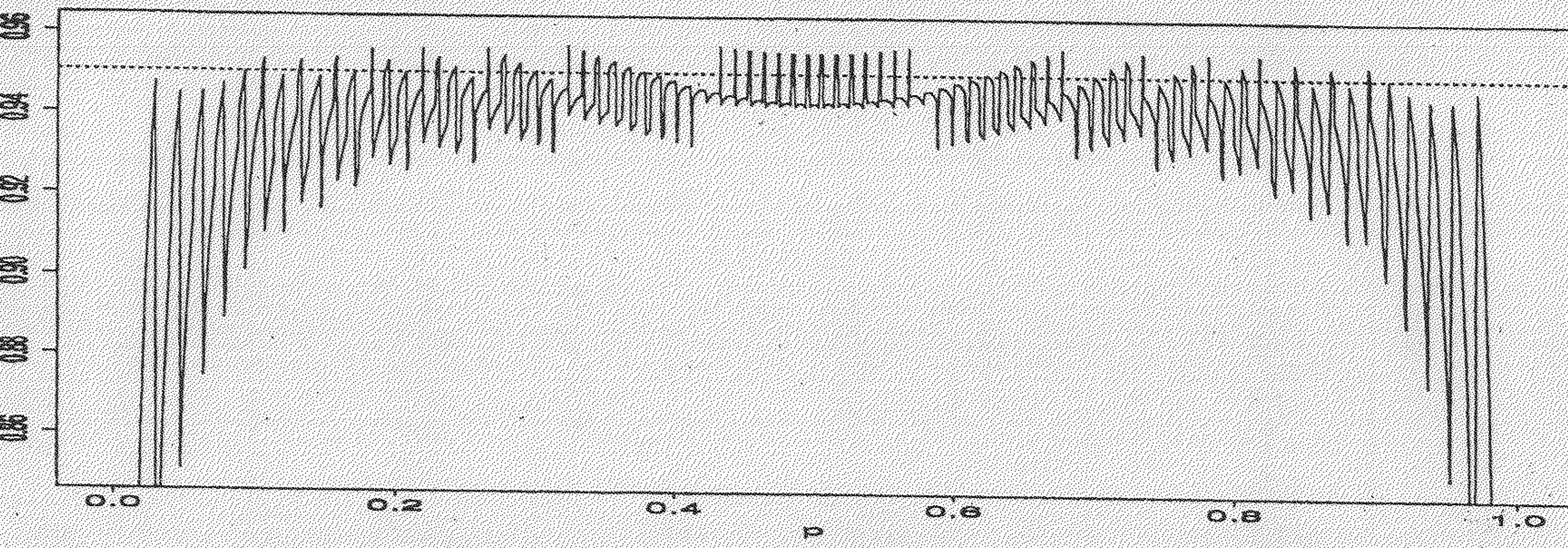


Figure 3: Oscillation phenomenon for fixed $n = 100$ and variable θ; Nominal 95% confidence interval

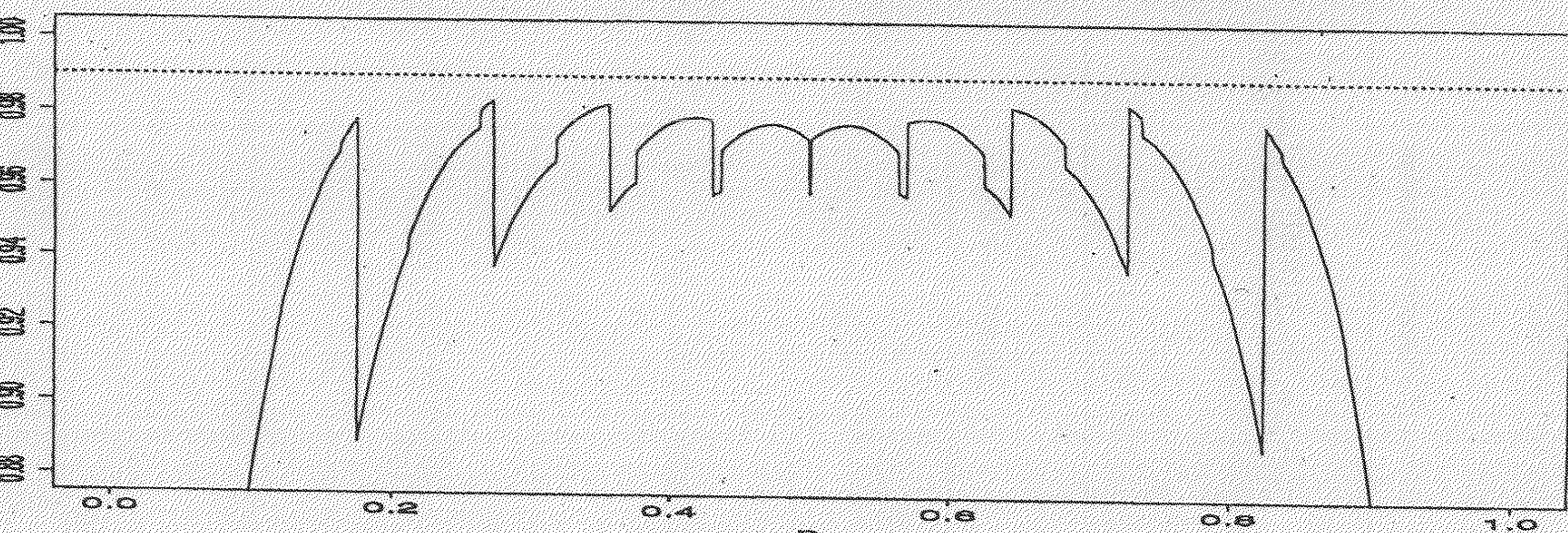


Figure: $n=20$ and variable θ; nominal 99% interval.

"Objective Bayes"

Berger
Prosper (?)
Maeshima
Kruse
Linnemann
Narsky
Roe

- A prescription with Bayesian origins for constructing functions of the data $\theta_{up}(\vec{x})$ having frequentist properties

  ~ coverage

  mean length (median?)

  $\sigma[\theta_{up}(\vec{x})]$?

- Exploits lucky data "automatically" (Berger, Roe ... ?)

- Easy to combine results

$$p(\theta \mid \vec{x}_1, \vec{x}_2) \propto \mathcal{L}_1(\vec{x}_1 \mid \theta)\, \mathcal{L}_2(\vec{x}_2 \mid \theta)\, \pi(\theta)$$

Feldman- Cousins Issues

Feldman
Cousins
Messier
Kafka
Schnee

- Avoids unphysical & null intervals.
- Coverage (~ independent of true θ)
- No flip-flopping

  1-sided interval in absence of convincing discovery } G. F., Maeshima, Conway, ⋮
- If $n=0$, upper limit on Poisson mean decreases for increasing expected background.

  Condition on ancillary var. (Roe - Woodroofe)

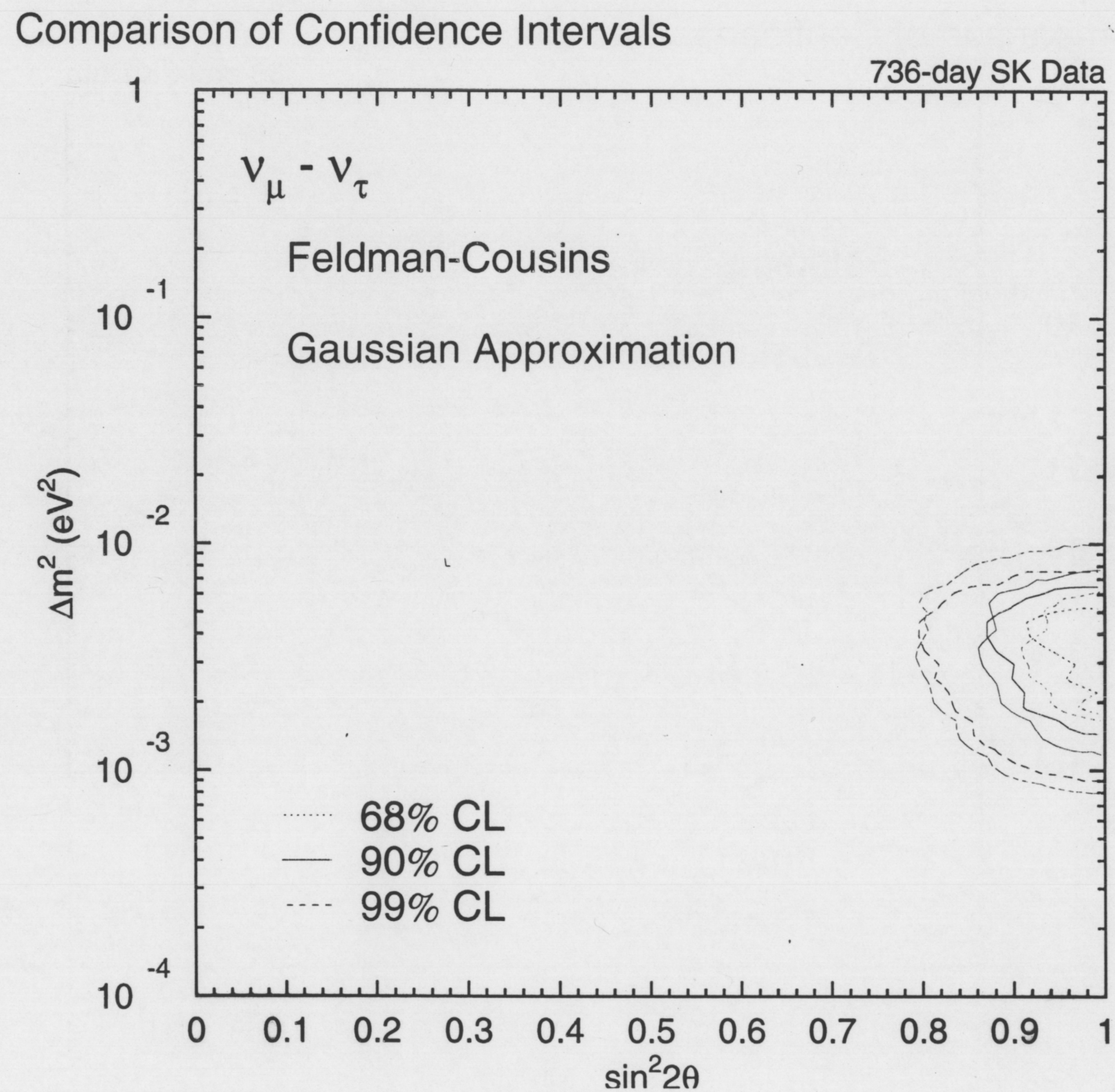
  Not always easy to find; (J. Berger)

  doesn't always give "desired" result (Roe)

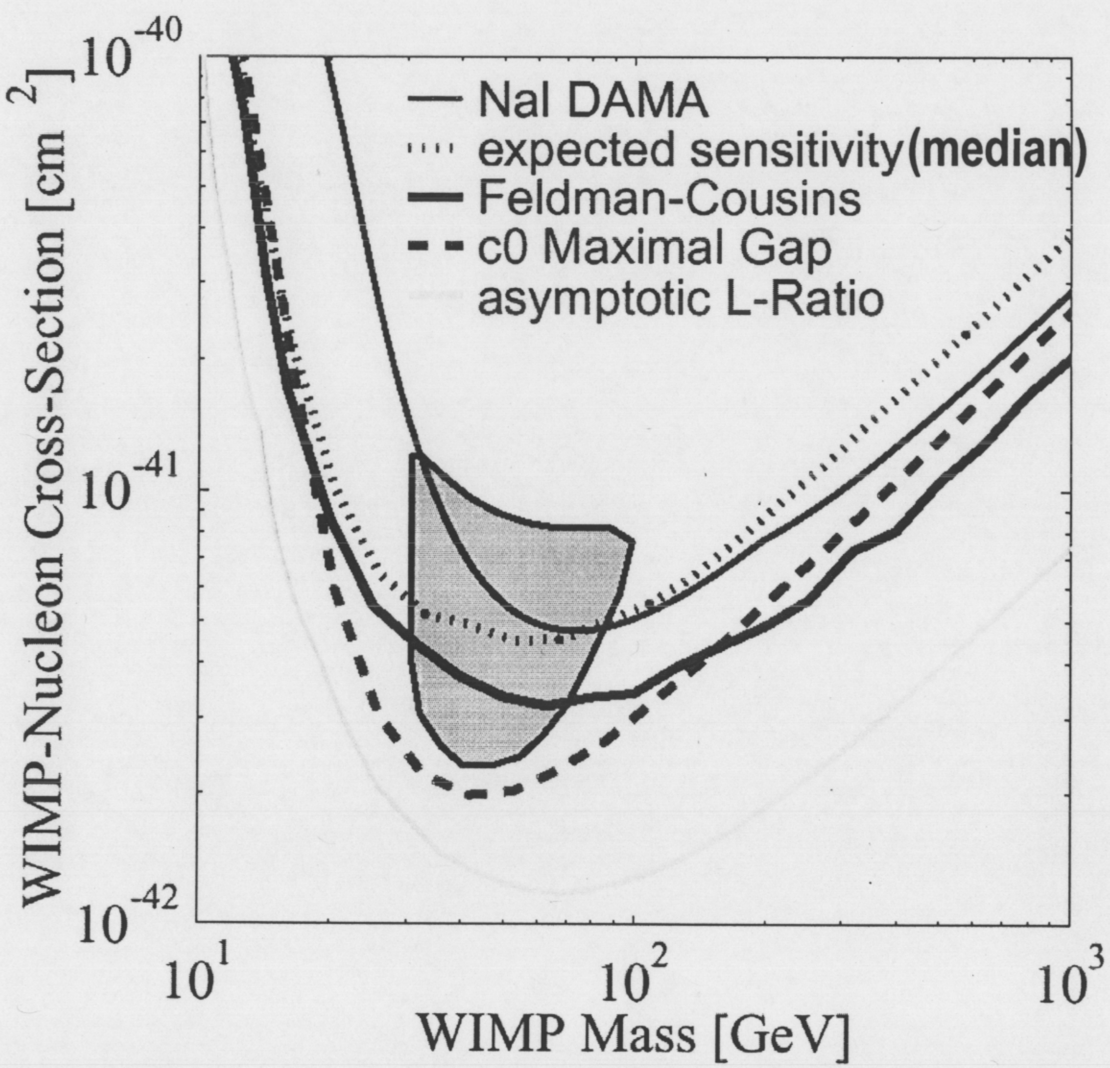
Can be computer intensive (Messier, Kafka)

What are the frequentist properties of these

[illegible]

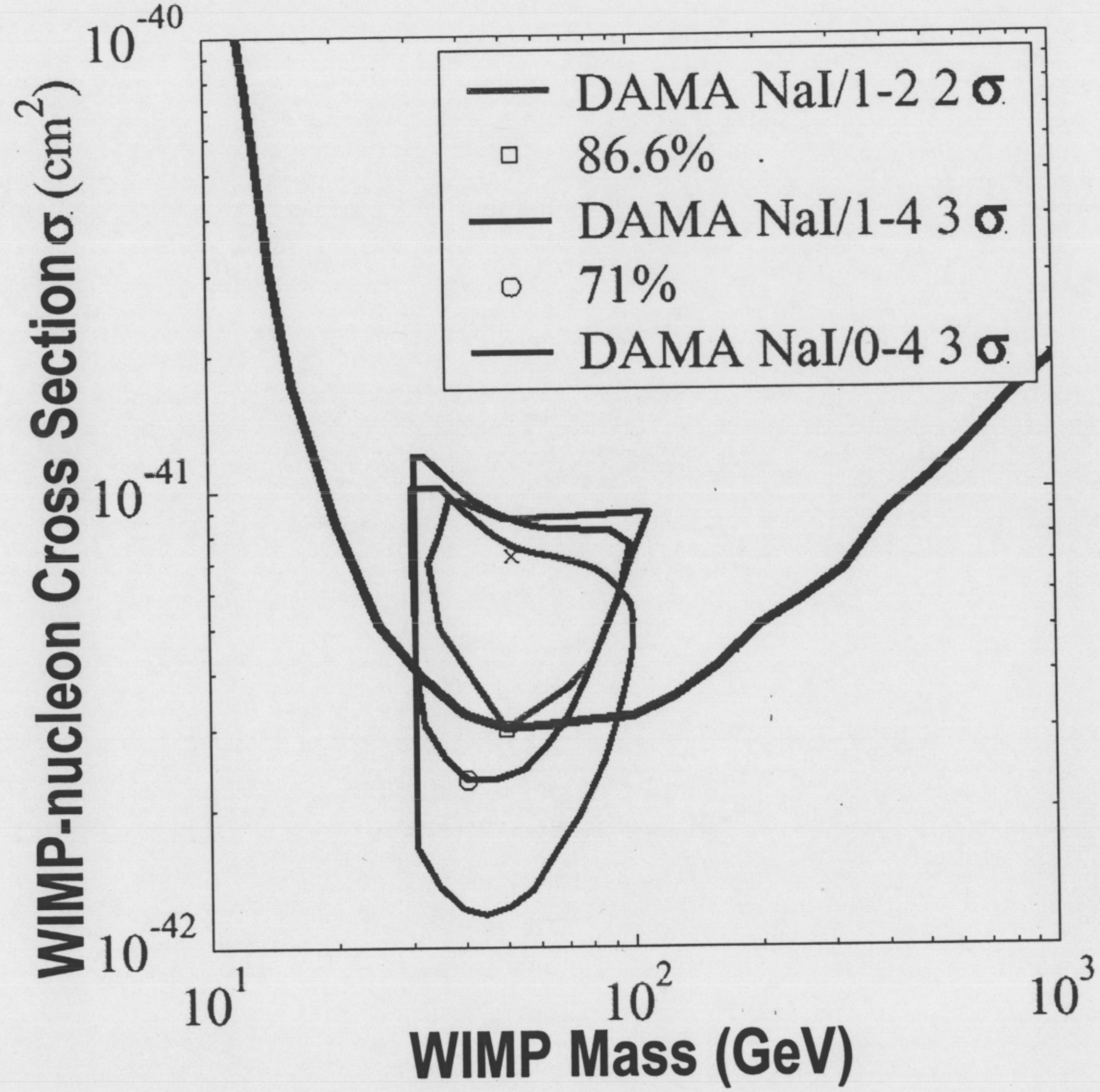

Comparison of Confidence Intervals
736-day SK Data
νμ - ντ
Feldman-Cousins
Gaussian Approximation
68% CL
90% CL
99% CL
Δm2 (eV2)
1
10-1
10-2
10-3
10-4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
sin22θ

CDMS Limits (90% CL)



- Because we see more multiple-scatter events than expected, limits are 50% better than expected sensitivity
- So far Bayesian method done only without energy info; results are similar to F-C.
- See http://dmtools.berkeley.edu/limitplots/ for interactive dark matter limit plotting

Compatibility with DAMA Regions



- Bottom of DAMA NaI/1-2 2σ (~87%) region excluded at 86.6% CL
- Bottom of DAMA NaI/1-4 3σ (~99%) region excluded at 71% CL
- It does not make sense to compare to DAMA NaI/0-4 region

CL_s

Bill Murray

Prescription: $$CL_s \equiv \frac{\text{P-value of } s+b}{\text{P-value of } b}$$

If $CL_s < \alpha$ exclude s

$\Rightarrow$ coverage $> 1-\alpha$ "frequentist-safe"
(?)

Limit for $n=0$ independent of b ✓

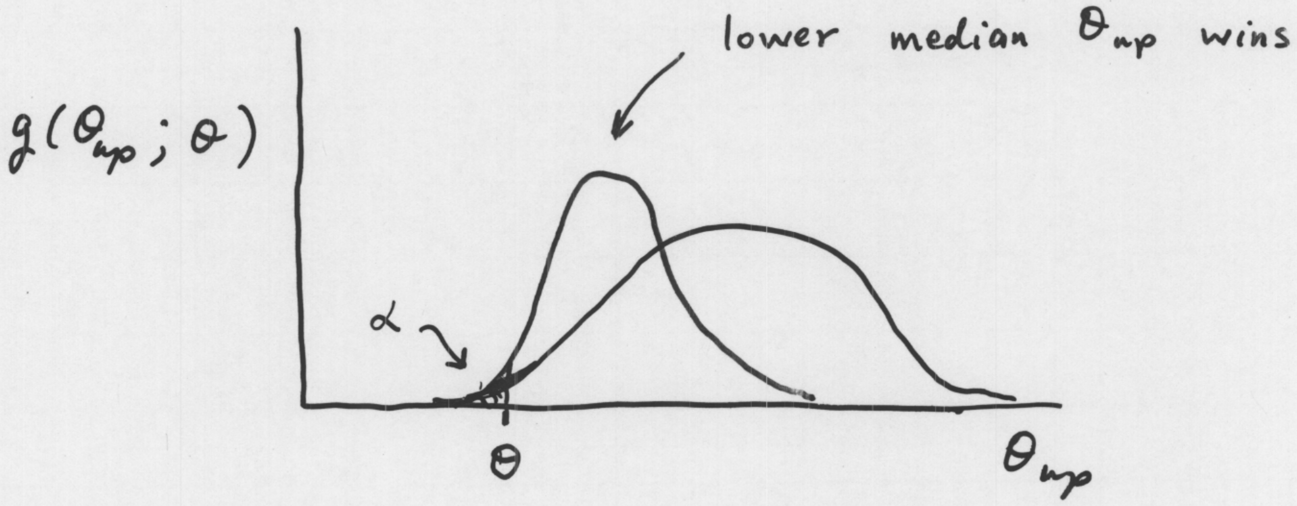
Claim: "Outperforms objective Bayes"

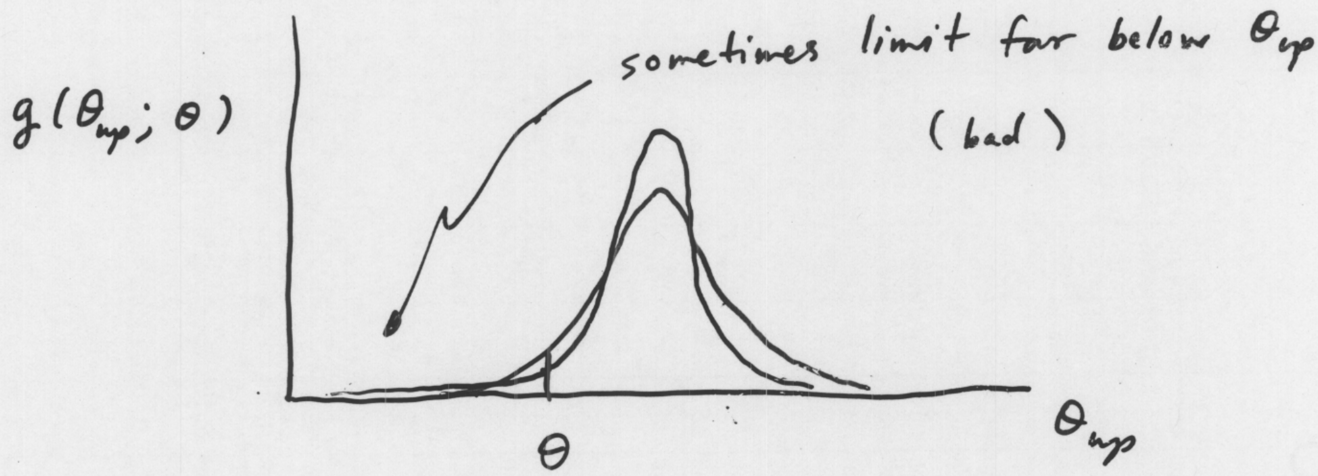
(when, where etc. is this true?)

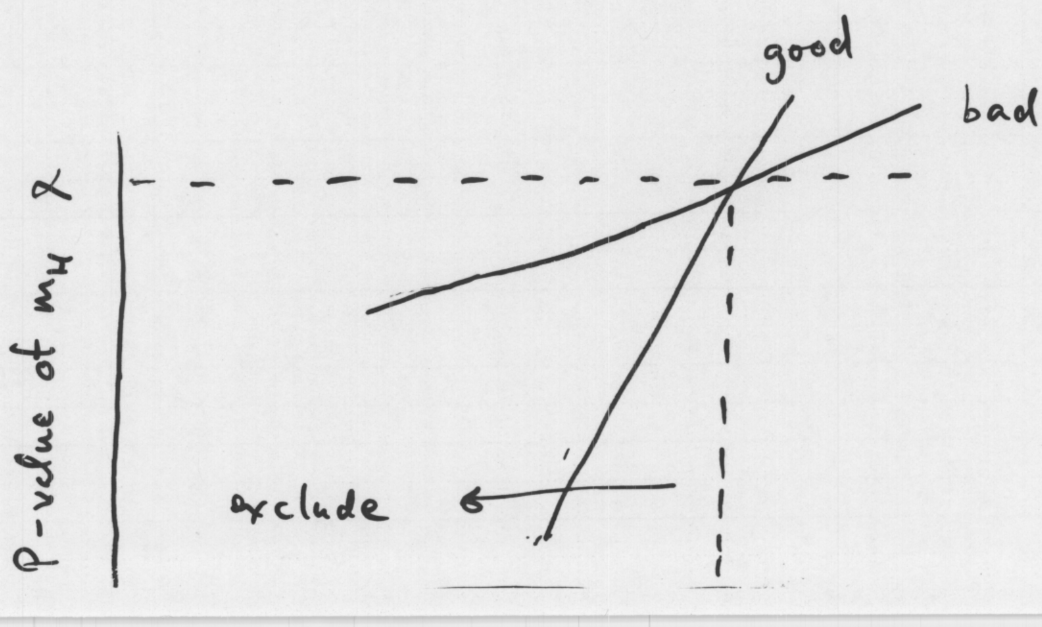
Others: "Signal Estimator method" McNamara
Hu
Nielson

⋮

Study distribution of limit (sampling pdf)







Systematics

Linnemann
Yellin

Bayes:

nuisance parameter (→ b)

$$p(\theta, b \mid \vec{x}) \propto \mathcal{L}(\vec{x} \mid \theta, b)\, \pi(\theta, b)$$

e.g. $$\pi(\theta, b) = \pi_\theta(\theta) \frac{1}{\sqrt{2\pi}\,\sigma_b} e^{-(b-b_0)^2/2\sigma_b^2}$$

Marginalize:

$$p(\theta \mid \vec{x}) = \int p(\theta, b \mid \vec{x})\, db$$

F-C + Cousins-Highland → problem with coverage?

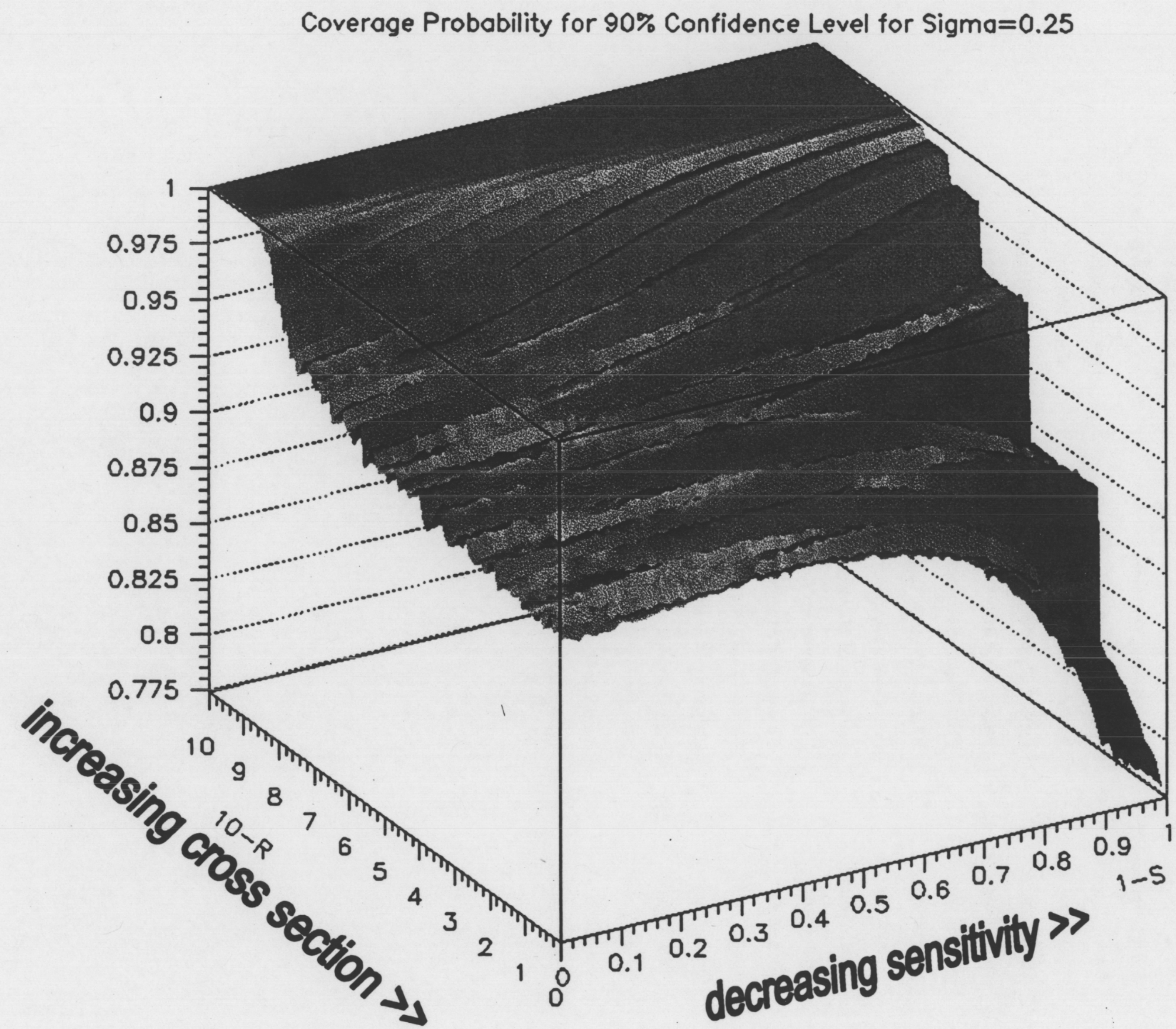
Objective Bayes: marginalize (becomes partly subj.)

Does it make sense to report a

"marginalized $\mathcal{L}$" ?

$$\mathcal{L}(\theta) = \int \mathcal{L}(\theta, b)\, \pi_b(b)\, db$$

?


Coverage Probability for 90% Confidence Level for Sigma=0.25
1
0.975
0.95
0.925
0.9
0.875
0.85
0.825
0.8
0.775
10
9
8
7
6
5
4
3
2
1
0
10-R
increasing cross section >>
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1-S
decreasing sensitivity >>

What is the killer argument for / against

F-C + modifications

Objective Bayes

CL_s ?

Which limit best allows the consumer to calculate

$$p(\theta \mid \text{limit}) \propto p(\text{limit} \mid \theta)\, \pi(\theta)$$

(consumer's prior → $\pi(\theta)$)

so that it gives the most (clearest?) information about θ where it counts?

N.B. limit alone not enough; need

$p(\text{limit} / \theta) \rightarrow E[\text{limit}]$

$\sigma[\text{limit}]$ (Raja.)

(Recall serious Bayesian consumer will want

$$p(\theta \mid \hat{x}) \propto \mathcal{L}(\hat{x} / \theta)\, \pi(\theta)$$

⇒ limit for "casual Bayesians".)

Expected Confidence Levels

What does expected mean?

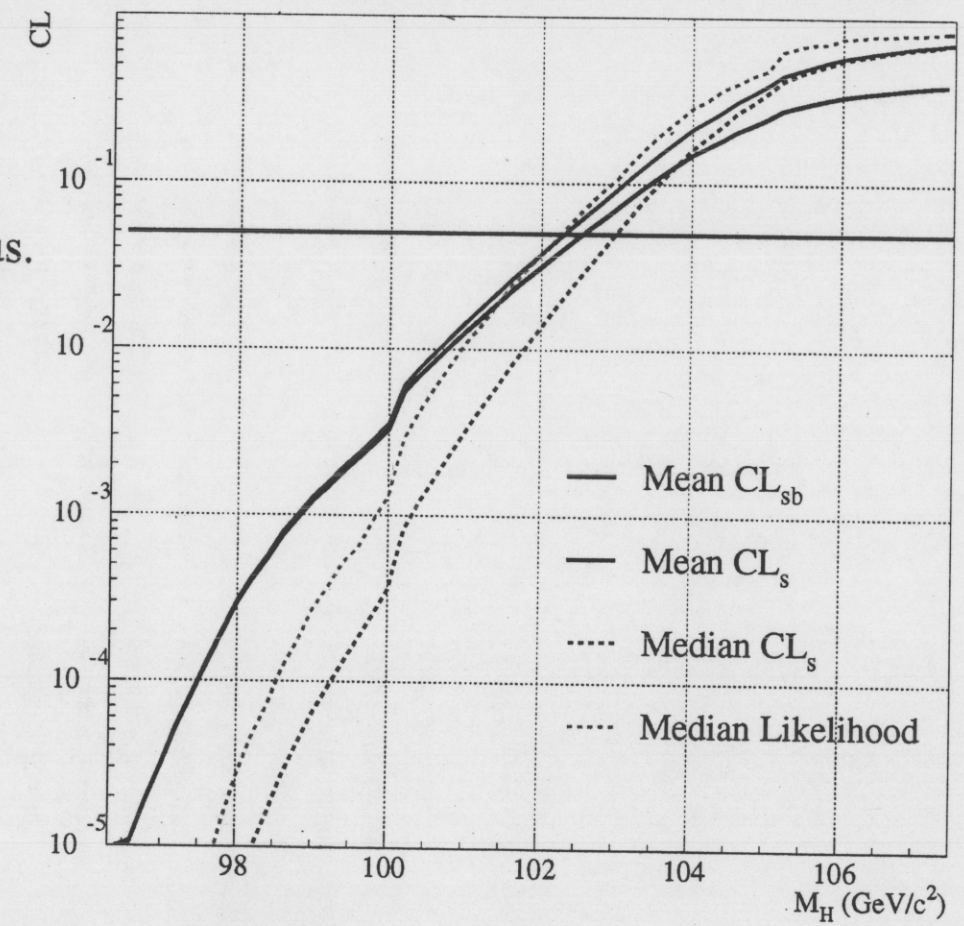
- **Mean**

  Has normally been used by us.

- **Median**

  No dependence on metric

Careful: Both are used here!



Expected limits:

CL_s .3GeV below CL_{sb}

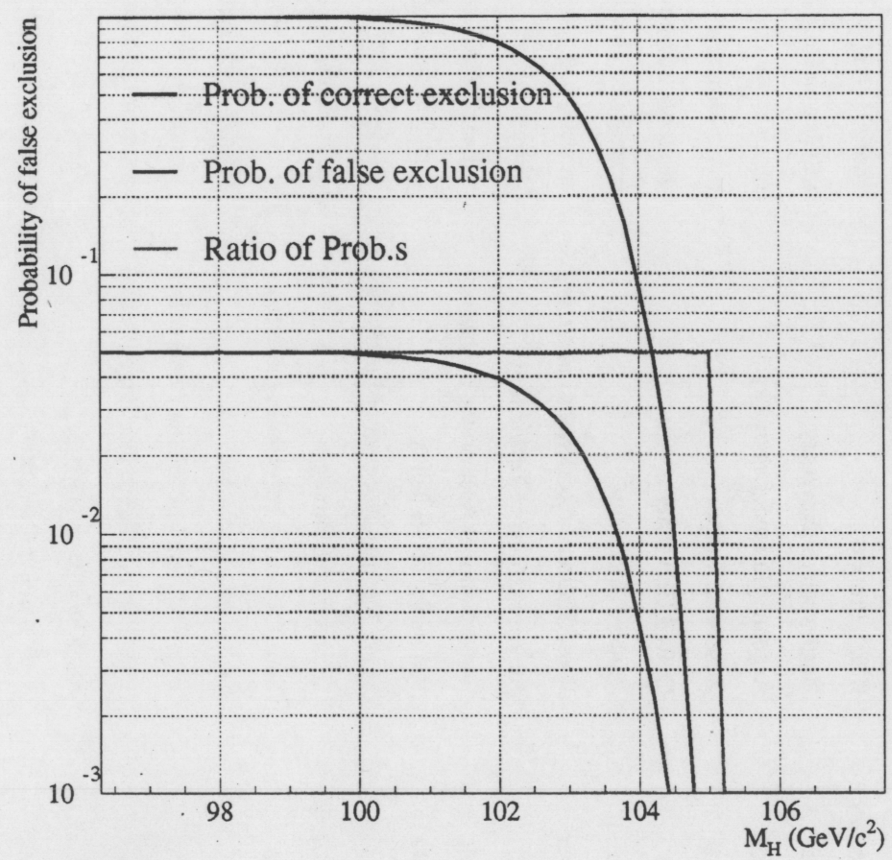
LR: 1GeV below CL_{sb}

Probability of Exclusion

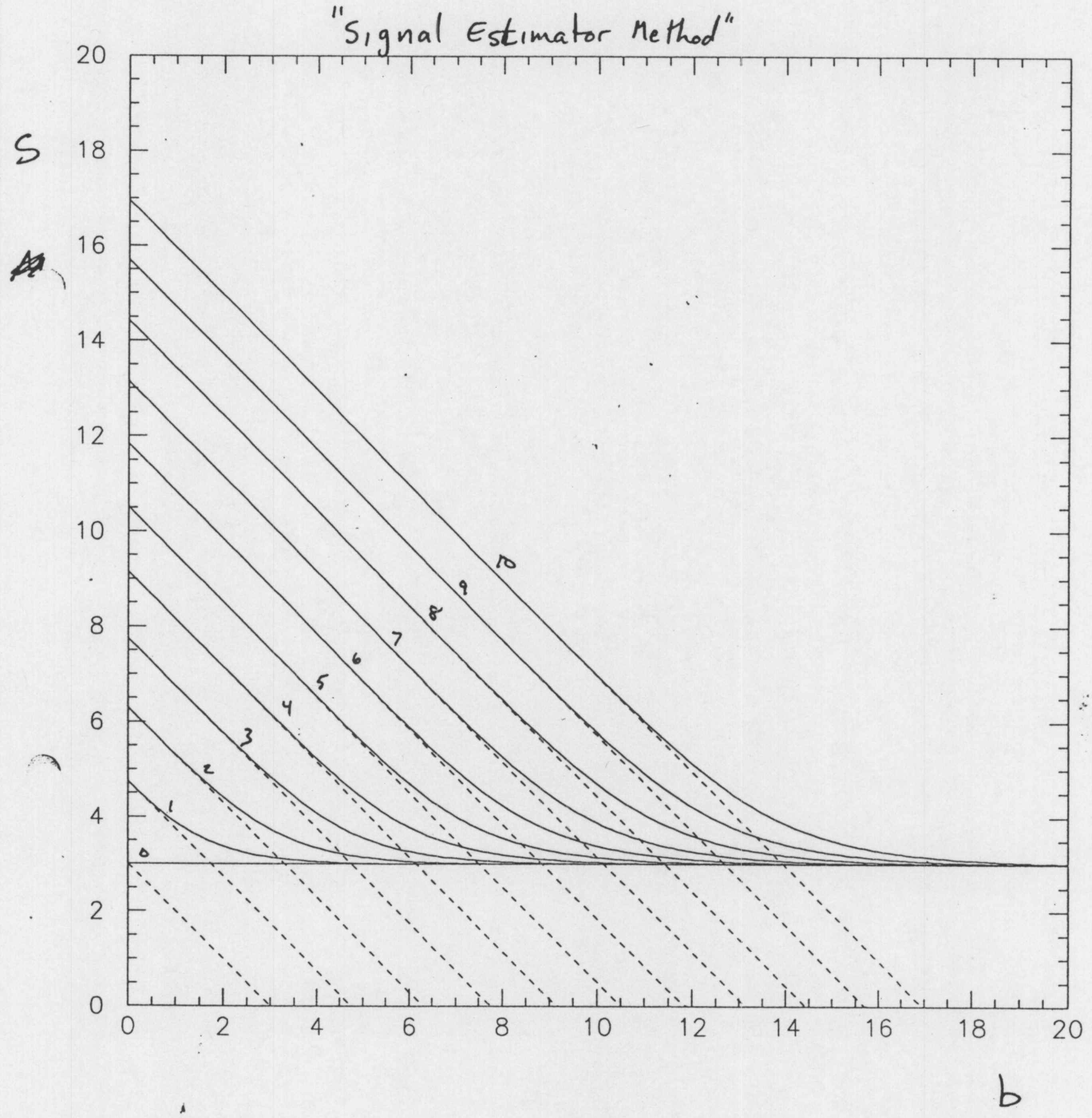
Define exclusion as CLs<0.05

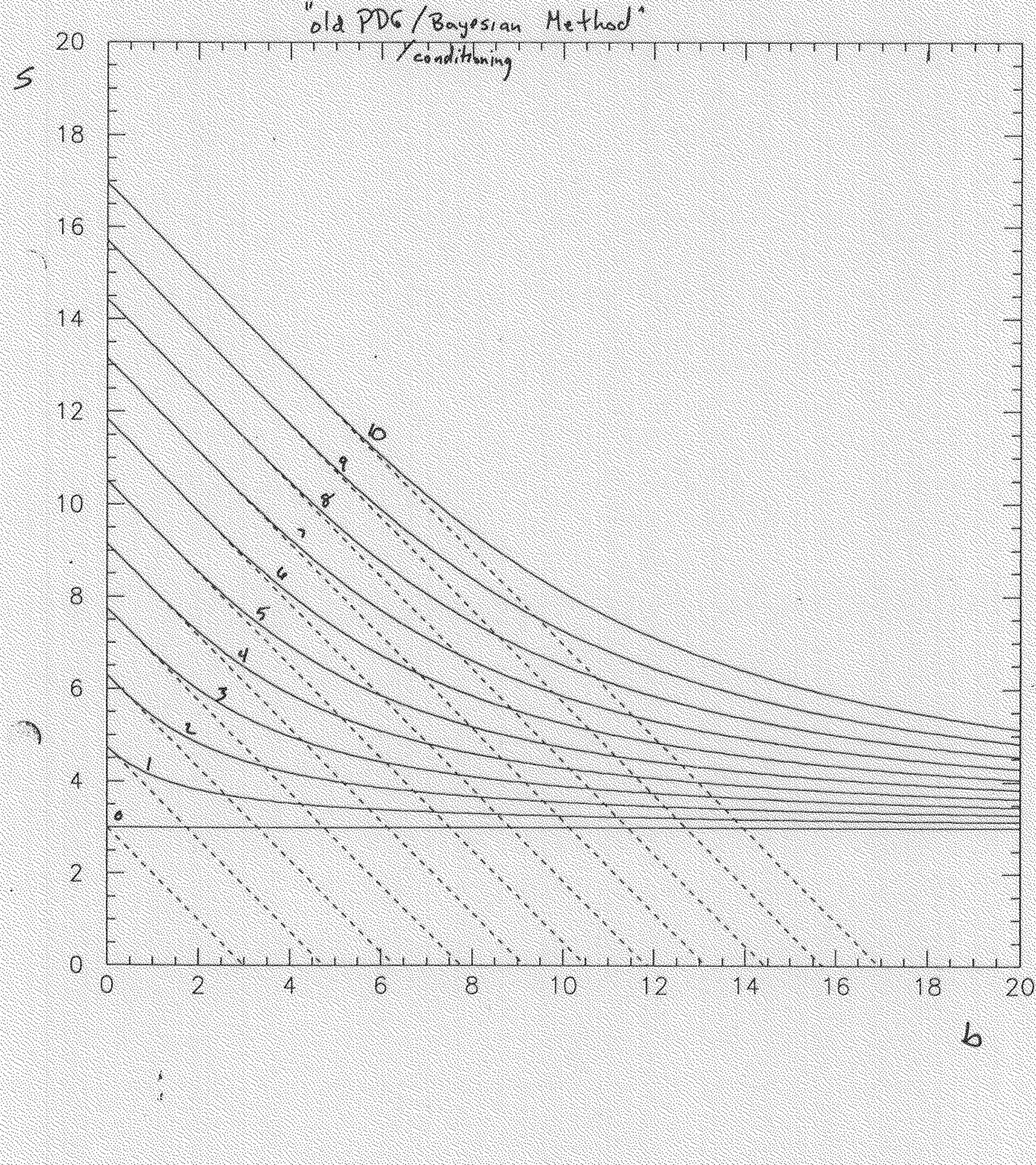
Probability of false exclusion should be 5% - but is less

Significant overcoverage



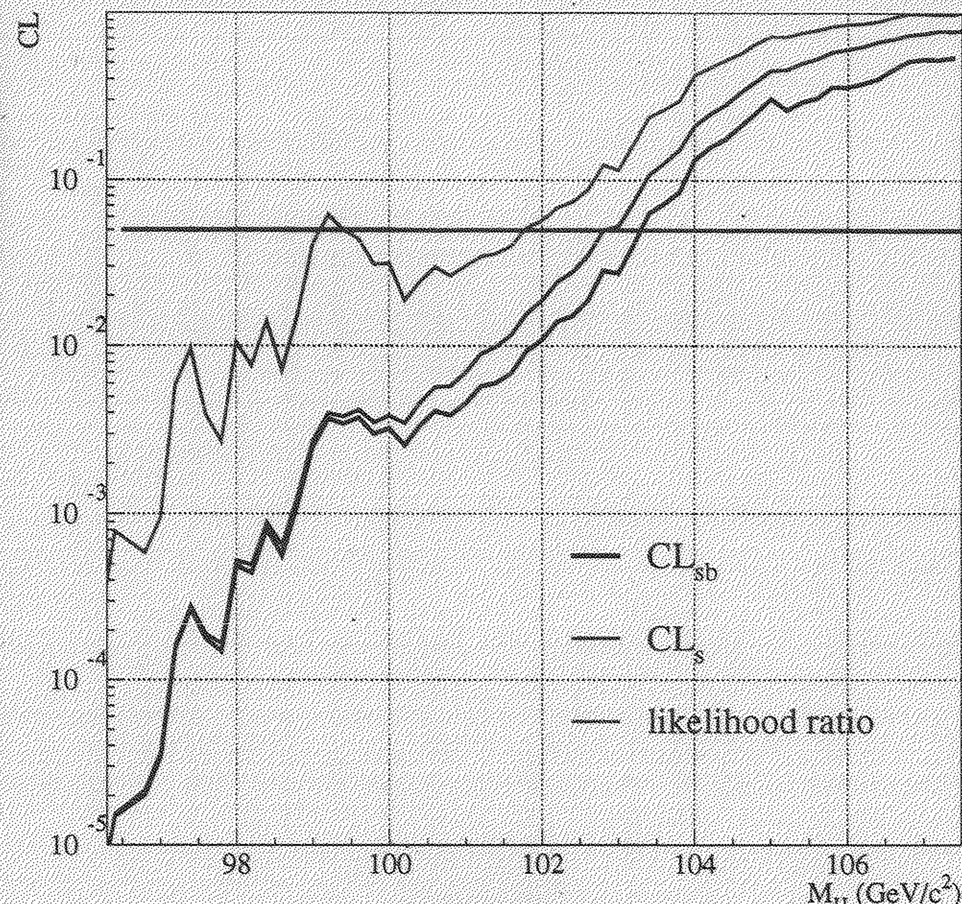
False exclusion rate is always 5% *of the true exclusion rate*


"Signal Estimator Method"
S
b
0
1
2
3
4
5
6
7
8
9
10
0 2 4 6 8 10 12 14 16 18 20


"old PDG / Bayesian Method"
/conditioning
s
b

Observed Confidence Levels

LR is same as R value proposed by d'Agostini



CLsb and CLs converge for good separation

Why because CLs near 1

Open Questions, etc.

- Is the stopping rule problem a non-problem?
- Does a Bayesian analysis prevent one from addressing goodness-of-fit?
- Should 95% coverage mean "minimum" or "typical"? (Why 95%?)
- Subjective $\neq$ arbitrary

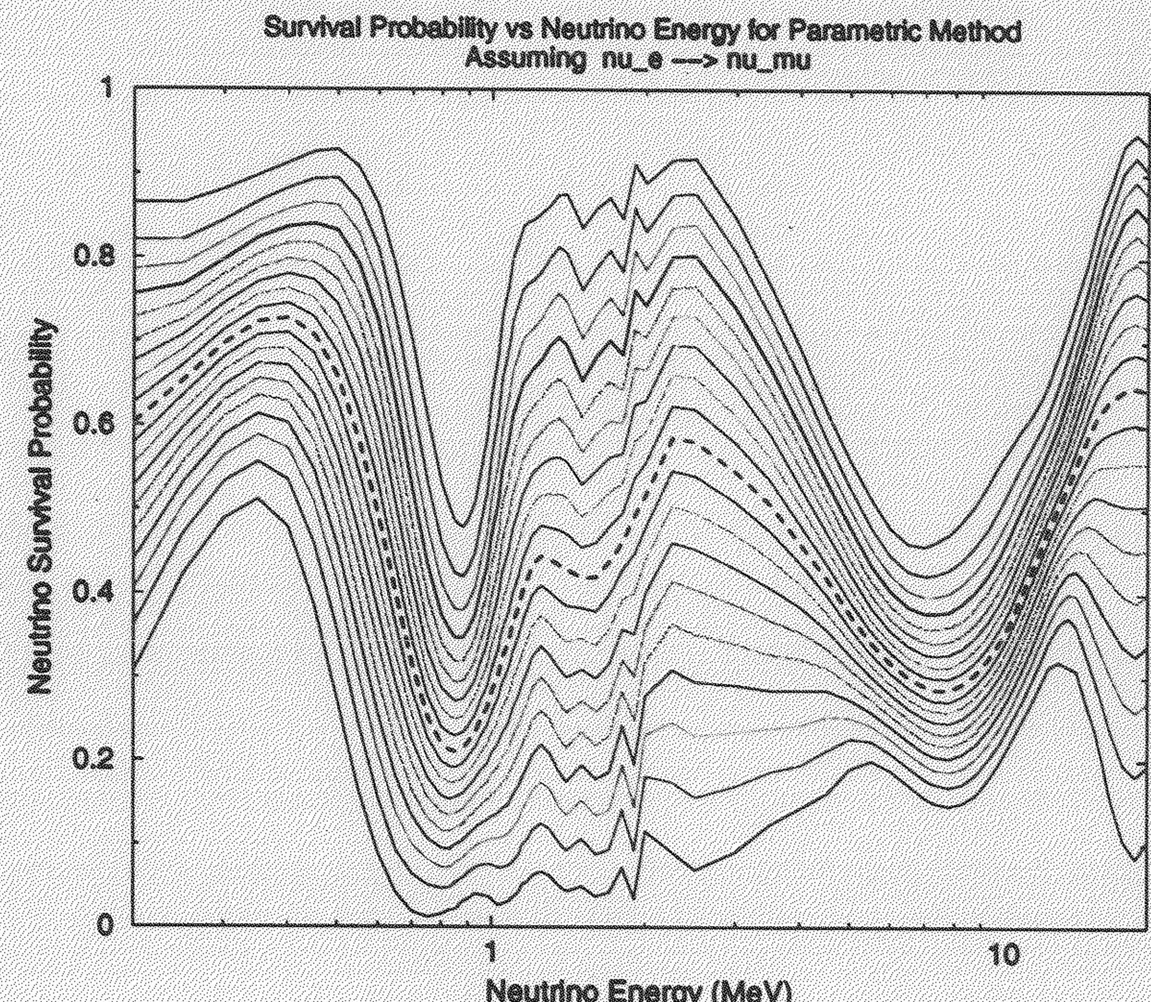
  Elements of arbitrariness:

  Bayes – choice of prior;
  "metric" $(\theta \rightarrow \ln \theta)$.

  Freq. – choice of method used to compactify result (estimator);
  choice of ensemble.

  → choose it, state it, use it, move on.

Pr(p|D): Active Neutrinos



For the "Results" section:

- Avoid subjectivity with respect to preference of theory (parameter)

- Summarize results:

  compact

  maximize useful information

  → functions of the data with well defined properties given a theory.

  likelihood function $\mathcal{L} = P(\text{data} \mid \text{theory})$

  limits

  P-values (CL) vs. parameter

  (goodness-of-fit statistics)

- Provide input for (Bayesian) consumer who will want

  consumer's prior ↘

$$p(\text{theory} \mid \text{result}) \propto p(\text{result} \mid \text{theory})\, \pi(\text{the[ory]}$$

e.g. $p(m_H \mid \text{your limit}) \propto p(\text{your limit} \mid m_H)\, \pi(m_H$

- coverage : like it or hate it ?

- ability to combine results

  $(\rightarrow \mathcal{L} \ ?)$

- digestable by non-specialists, e.g. your ed. board

- do we need a consensus ?

- Future workshop: <u>systematics</u>