Statistics Issues in Neutrino Physics…
…and Their Influence on HEP

Bob Cousins

Univ. of California, Los Angeles

Collaborator on NOMAD

Journée Jacques Bouchez

19 November 2010

Outline

1. **Introduction**

   “Classic” statistics issues have arisen in ν physics

2. **1980 Reines et al. “neutrino instability” claim**

   F. James re propagation of large uncertainties

3. **ν_μ mass (π decay)**

   Issue of limits on bounded physical parameter

4. **ν_e mass, (tritium β decay)**

   Issue becomes critical with measured $m_\nu^2 < 0$

5. **Karmen: 2.8 expected background, 0 events**

   The Likelihood Principle

6. **ν_τ mass ($\tau \rightarrow 5\pi\ \nu_\tau$) and “lucky” events**

   Statistical Principle of “Conditioning”

7. **Summary: The Issues haven’t gone away**

Physical Review Letters

VOLUME 45 | 20 OCTOBER 1980 | NUMBER 16

Evidence for Neutrino Instability

F. Reines, H. W. Sobel, and E. Pasierb
Department of Physics, University of California at Irvine, Irvine, California 92717
(Received 24 April 1980)

This Letter reports indications of neutrino instability obtained from data taken on the charged- and neutral-current branches of the reaction

$$\bar{\nu}_e + d \to n + n + e^+ \ \text{(ccd)}; \quad \bar{\nu}_e + d \to n + p + \bar{\nu}_e \ \text{(ncd)}$$

at 11.2 m from a 2000-MW reactor. These results at the (2–3)-standard-deviation level, based on the departure of the measured ratio (ccd/ncd) from the expected value, make clear the importance of further experimentation to measure the $\bar{\nu}_e$ spectrum versus distance.

$$\mathcal{R} = r_{\text{expt}}/r_{\text{theor}}$$

$$r_{\text{expt}} = \frac{\langle \eta' \rangle R_{2n}^{\text{ccd}}}{\langle \eta^2 \rangle (R_{1n} - R_{1n}^{\text{ccp}}) - 2(0.89)(\langle \eta \rangle - \langle \eta^2 \rangle) R_{2n}^{\text{ccd}}}.$$

Inserting numerical values for the two modes (B and C) and then combining the results, we find

$$r_{\text{expt}} = 0.167 \pm 0.093,$$

where

$$\sigma_{r_{\text{expt}}}^2 = \left(\frac{\partial r}{\partial R_{1n}}\right)^2 \sigma_{R_{1n}}^2 + \left(\frac{\partial r}{\partial R_{2n}}\right)^2 \sigma_{R_{2n}}^2 + \cdots.$$

$$\mathcal{R} = 0.38 \pm 0.21 \text{ or } 0.40 \pm 0.22,$$

which is a (3.0–2.7)-standard-deviation departure from unity, if it is assumed that the σ_r calculated above is representative of a normal distribution.

CERN-DATA HANDLING DIVISION
DD/81/02
F. James
13 February 1981

DETERMINING THE STATISTICAL SIGNIFICANCE OF EXPERIMENTAL RESULTS

Lectures presented at the
1980 CERN School of Computing,
Vraona, Attiki, Greece
14 September-27 September 1980

1.6.5 The propagation of large errors.

The question now arises of what to do in the general case for continuous variables when the linear approximation for error propagation is suspected of being poor. Straightforward calculation of the distribution of the new variable R involves complicated integrals over the component distributions which, even if they are independent Gaussians, quickly become intractable, and one must resort to numerical calculations even in relatively simple cases.

One such case came up recently in the analysis of an experiment by Reines, Sobel, and Pasierb which gives evidence for the instability of the neutrino. This result is of the greatest importance in high energy physics since it has generally been believed that all neutrinos were massless and could not decay. In view of the consequences of neutrino decay, it is necessary to determine the significance of these results accurately. The final result of the experiment is the measurement of the ratio of two cross sections, let us call this R. Expressed in terms of the elementary quantities measured in the experiment, it can be written as:

$$R = \frac{a}{\frac{d}{k^2 e}(b-c) - 2\left(1 - \frac{k^2 d}{ke}\right) a}$$

where
$a = 3.84 \pm 1.33$
$b = 74 \pm 4$
$c = 9.5 \pm 3$
$d = 0.112 \pm 0.009$
$e = 0.32 \pm 0.002$
$k = 0.89$

Straightforward application of the linear approximation gives:

$$R \simeq 0.191 \pm 0.073$$

But theoretical calculations show that the neutrino is unstable if R is less than about 0.42. Therefore, based on approximate error analysis, the result appears to be very significant: 3.2 standard deviations or about one chance in a thousand that the neutrino is stable.

However, two of the elementary quantities have large errors, and two quantities enter into the formula twice, producing correlations. In addition, there are several fractions, which we have seen cause non-Gaussian distributions, so let us try to calculate the exact confidence intervals for R. The easiest (and perhaps the only) way to do this is by Monte Carlo. Choose values of a,b,c,d,e randomly according to the appropriate Gaussian distributions (we will be optimistic and assume that at least the elementary measurements are Gaussian with known variances), and plot the resulting values of R. The FORTRAN program to do this is so simple that I include it here (Calls to subroutines beginning with H are for the HBOOK histogramming package; NORRAN is a Gaussian random number generator; all subroutines called here are from the CERN Program Library):

```
      PROGRAM REINES(INPUT,OUTPUT)
C         CALCULATION OF ERROR ON NEUTRAL TO CHARGED CURRENT
C         NEUTRINO INTERACTIONS, D'APRES REINES AND ROOS.
C
C     SET UP HISTOGRAM OF R
      CALL HBOOK1(1,10H N OVER D  , 50 ,0.,0.5,0.)
C
C  FILL HISTOGRAM BY LOOPING OVER RANDOM SAMPLES OF R
      DO 100 I= 1, 10000
      CALL NORRAN(XN)
      XN = XN*1.33 + 3.84
      CALL NORRAN(X112)
      X112 = X112 * .009 + 0.112
      CALL NORRAN(X74)
      X74 = X74 * 4. + 74.
      CALL NORRAN(X95)
      X95 = X95 * 3. + 9.5
      CALL NORRAN(X32)
      X32 = X32 * 0.02 + 0.32
      X89 = 0.89
      D1 = X112*(X74-X95)/(X89*X32)
      D2 = 2.0 * XN * (1.0 - (X89*X112/X32))
      XXX = XN/(D1-D2)
      CALL HFILL (1,XXX)
  100 CONTINUE
C
C  ASK FOR PRINTING OF HISTOGRAM, WITH INTEGRATED CONTENTS
      CALL HINTEG(1, 3HYES)
      CALL HISTDO
C
      STOP
      END
```

The histogram showing the distribution of the 10000 Monte Carlo values of R is shown in Figure 4. Those of you familiar with the reading of HBOOK output will quickly find the significant number, namely the number of entries falling above 0.42. This is almost 4%, so that the true significance of the result is only 4% instead of the apparent 0.1%. Notice also the skew, non-Gaussian distribution of R.


```
CHANNELS  10   0         1         2         3         4         5
           1   12345678901234567890123456789012345678901234567890

CONTENTS 100        1112233444444444433322222111111
          10   12358238693201265965493168531876401089754543322221
          1.   90992724101391557441788476794339467366374772477661

INTEGRAT1000            11122334445566677778888889999999999999999
         100      12346925826150594826924792357890123455666777888
          10   13731476214678173294982409580852779988515159257034
          1.   99879682334767274823086073093698285840304180418401

LOW-EDGE   1.            1111111111222222222233333333334444444444
*10**  1   0   01234567890123456789012345678901234567890123456789

 ENTRIES = 10000   UNDERFLOW = 23   OVERFLOW = 136
 MEAN VALUE = .2009E+00       R . M . S =.9219E-01
```


Figure 4. Distribution of R for experiment of Reines et al.

PHYSICAL REVIEW D VOLUME 20, NUMBER 11 1 DECEMBER 1979

Precision measurement of the muon momentum in pion decay at rest

M. Daum, G. H. Eaton, R. Frosch, H. Hirschmann, J. McCulloch,* R. C. Minehart,† and E. Steiner
Swiss Institute for Nuclear Research, SIN, 5234 Villigen, Switzerland

We have used the p_{μ^+} value of Eq. (7) to calculate the upper limit for the muon-neutrino mass. The muon and pion masses of Ref. 6, $m_{\mu^+} = 105.659\,46 \pm 0.000\,24$ MeV/c^2, $m_{\pi^-} = 139.5679 \pm 0.0015$ MeV/c^2, were used. Assuming that m_{π^+} is equal to m_{π^-} (CPT theorem) one obtains from Eq. (1) the squared neutrino mass,

$$m_{\nu_\mu}{}^2 = 0.13 \pm 0.14\ (\mathrm{MeV}/c^2)^2 . \quad (8)$$

The uncertainty of $m_{\nu_\mu}{}^2$ (one standard deviation) has been obtained by adding the three contributions given by Eqs. (2)–(4) in quadrature. The contributions of Δm_π and Δp_{μ^+} to $\Delta(m_{\nu_\mu}{}^2)$ are about equal, whereas the contribution of Δm_μ is much smaller.

Following the method recommended by the Particle Data Group,[33] illustrated in Fig. 22, we calculated the upper limit of the muon-neutrino mass. The result is

$$m_{\nu_\mu} \leq 0.57\ \mathrm{MeV}/c^2\ (90\%\ \text{confidence level}) . \quad (9)$$

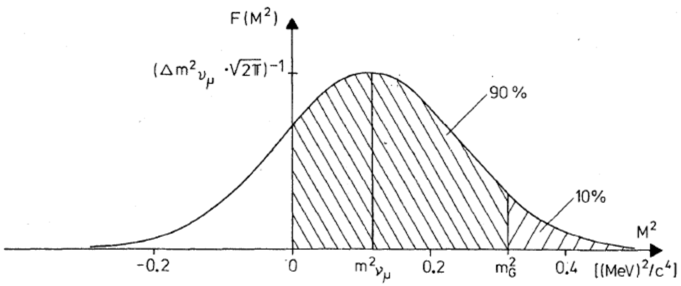


FIG. 22. According to the prescription of the Particle Data Group (Ref. 33) the upper limit m_G of the muon-neutrino mass is calculated from the squares mass $m_{\nu_\mu}{}^2$ and its uncertainty $\Delta(m_{\nu_\mu}{}^2)$ by setting the probability function $F(M^2)$ to zero for $M^2 < 0$, as indicated in the figure.

[33] T. G. Trippe, private communication, 1976.

Estimation of Upper Limits from Experimental Data

Virgil L. Highland

July 1986, Revised February 1987

Temple University

Philadelphia, PA 19122

Upper limit on mean of Gaussian based on one sample, x.

Physical values of mean are non-negative.

Numbers are in units of sigma (Gaussian rms).

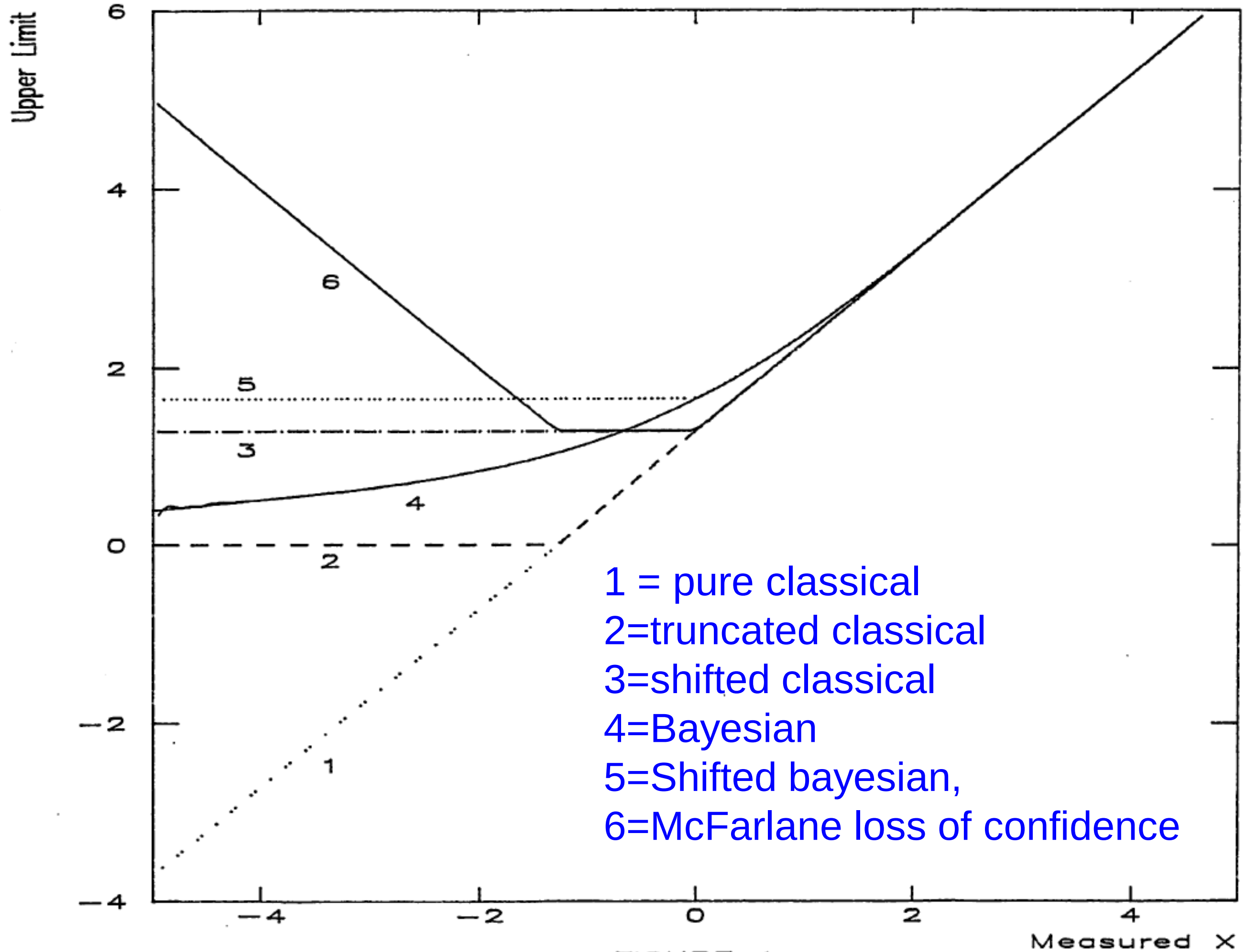


FIGURE 1

PDG RPP: 1986, same in 1988 (nothing in 1984)

B.3 Limits in Case of Bounded Physical Regions

If we assume μ is bounded from below by μ_{min} … we may estimate an upper limit for μ at the C.L. (e.g., 90%…) by the following procedure:

1) *Renormalize* the normal probability distribution …such that the integral of [Gaussian] from μ_{min} to infinity to 1.0.
2) Find the value μ_1 such that the integral… from μ_{min} to μ_1 is equal to the desired value of CL.
3) Set μ_1 to be the desired upper limit with confidence CL.

…this is *conservative*…

Ann. Rev. Nucl. Part. Sci. 1988. 38: 185–215

DIRECT MEASUREMENTS OF NEUTRINO MASS

R. G. H. Robertson

Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

D. A. Knapp

Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550

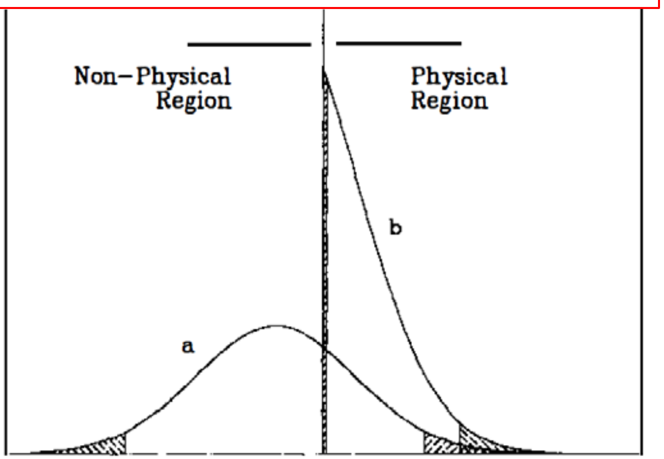


Figure A.2 The prescription recommended by the Particle Data Group for setting confidence levels on the true value of a parameter confined to a physical regime. Curve "*a*" shows a likelihood function (LF) centered on a measurement, which may fall in the non-physical regime. Curve "*b*" is the tail of the LF renormalized to unit area in the physical regime. The hatched areas then exclude small and large values of the parameter at a selected confidence level.

Limit on $\bar{\nu}_e$ Mass from Observation of the β Decay of Molecular Tritium

R. G. H. Robertson, T. J. Bowles, G. J. Stephenson, Jr., D. L. Wark,[a] and J. F. Wilkerson
Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

D. A. Knapp
Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550
(Received 6 May 1991)

We report the most sensitive direct upper limit set on the mass m_ν of the electron antineutrino. Our measurements of the shape of the β decay spectrum of free molecular tritium yield, under the assumption of no new physics other than that of mass, a central value for m_ν^2 of $-147 \pm 68 \pm 41$ eV^2, which corresponds to an upper limit of 9.3 eV (95% confidence level) on m_ν. The result is in clear disagreement with a reported value of 26(5) eV.

tainties. In order to set confidence limits on the true value of a quantity that is inherently non-negative, a Bayesian approach is needed [11]. Adding the uncertainties in quadrature, one finds an upper limit of 9.3 eV on the neutrino mass at the 95% confidence level. If the

$$m_\nu^2 < 96 \text{ eV}^2$$

[11] See, R. G. H. Robertson and D. A. Knapp, Annu. Rev. Nucl. Part. Sci. **38**, 185 (1988).

1995 PDG Review of Particles Properties

ν_e MASS

[1] PDG 94 formal upper limit, as obtained from the m^2 average in the next section, is 5.1 eV at the 95%CL. Caution is urged in interpreting this result, since the m^2 average is positive with only a 3.5% probability. If the weighted average m^2 were forced to zero, the limit would increase to 7.0 eV.

ν_e MASS SQUARED

VALUE (eV^2)	*DOCUMENT ID*	*TECN*	*COMMENT*
− 54± 30 OUR AVERAGE			
− 39± 34± 15	[14] WEINHEIMER 93	SPEC	^3H β decay
− 24± 48± 61	[15] HOLZSCHUH 92B	SPEC	^3H β decay
− 65± 85± 65	[16] KAWAKAMI 91	SPEC	$\overline{\nu}_e$, tritium
−147± 68± 41	[17] ROBERTSON 91	SPEC	$\overline{\nu}_e$, tritium

Neutrino-*less* physics

Nuclear Physics B (Proc. Suppl.) 13 (1990) 547–550
North-Holland

RECENT RESULTS FROM THE UCSB/LBL DOUBLE BETA DECAY EXPERIMENT[§]

D.O. CALDWELL,* R.M. EISBERG,* F.S. GOULDING,† B. MAGNUSSON,* A.R. SMITH,† and M.S. WITHERELL*

Presented by David O. CALDWELL

Physics Department, University of California, Santa Barbara, CA 93106, USA

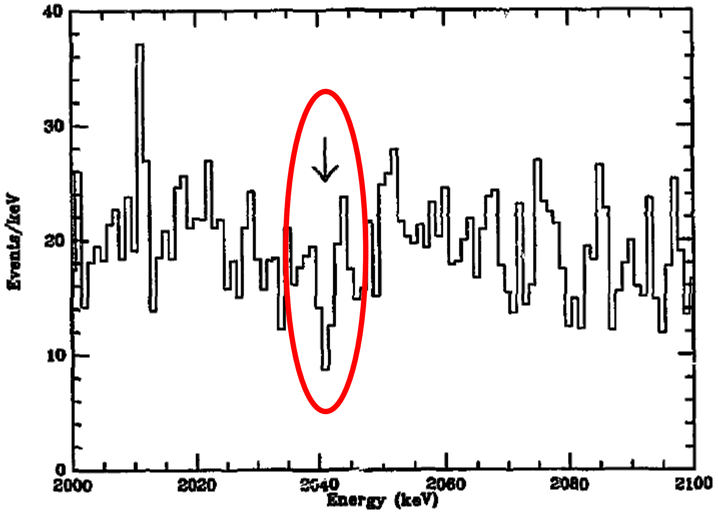


FIGURE 1

UCSB/LBL Ge multidetector data in the vicinity of the possible $0^+ \rightarrow 0^+$ $\beta\beta_{0\nu}$ peak (arrow) for 21 kg·y of sensitivity.

The data in this energy region are shown in Fig. 1, where a dip is observed at the energy where a peak is sought. An analysis using Bayesian statistics (an approach we believe to be most nearly correct) agrees with a maximum likelihood calculation to 0.1% in giving from these data a 90% confidence level lower limit on the halflife of ^{76}Ge ($0^+ \rightarrow 0^+$ transition) of 1.2×10^{24} years. In this field, 68% confidence levels

Phys. Rev. D57 3873 (1998)

Unified approach to the classical statistical analysis of small signals

Gary J. Feldman*
Department of Physics, Harvard University, Cambridge, Massachusetts 02138

Robert D. Cousins†
Department of Physics and Astronomy, University of California, Los Angeles, California 90095

“Test for $\theta=\theta_0$” ↔
“Is θ_0 in confidence interval for θ”

Using the likelihood ratio hypothesis test, this correspondence is the basis of intervals/regions F-C advocated.

Kendall and Stuart

CHAPTER 22

LIKELIHOOD RATIO TESTS AND TEST EFFICIENCY

The LR statistic

22.1 The ML method discussed in Chapter 18 is a constructive method of obtaining estimators which, under certain conditions, have desirable properties. A method of test construction closely allied to it is the likelihood ratio (LR) method, proposed by Neyman and Pearson (1928). It has played a role in the theory of tests analogous to that of the ML method in the theory of estimation.

As before, we have the LF

$$L(x|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ is a vector of $r+s=k$ parameters ($r \geq 1, s \geq 0$) and x may also be a vector. We wish to test the hypothesis

$$H_0 : \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r0}, \quad (22.1)$$

which is composite unless $s=0$, against

$$H_1 : \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r0}.$$

We know that there is generally no UMP test in this situation, but that there may be a UMPU test – cf. **21.31**.

The LR method first requires us to find the ML estimators of $(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$, giving the unconditional maximum of the LF

$$L(x|\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\theta}}_s), \quad (22.2)$$

and also to find the ML estimators of $\boldsymbol{\theta}_s$, when H_0 holds,[1] giving the conditional maximum of the LF

$$L(x|\boldsymbol{\theta}_{r0}, \hat{\hat{\boldsymbol{\theta}}}_s). \quad (22.3)$$

$\hat{\hat{\boldsymbol{\theta}}}_s$ in (22.3) has been given a double circumflex to emphasize that it does not in general coincide with $\hat{\boldsymbol{\theta}}_s$ in (22.2). Now consider the likelihood ratio[2]

$$l = \frac{L(x|\boldsymbol{\theta}_{r0}, \hat{\hat{\boldsymbol{\theta}}}_s)}{L(x|\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\theta}}_s)}. \quad (22.4)$$

Since (22.4) is the ratio of a conditional maximum of the LF to its unconditional maximum, we clearly have

$$0 \leq l \leq 1. \quad (22.5)$$

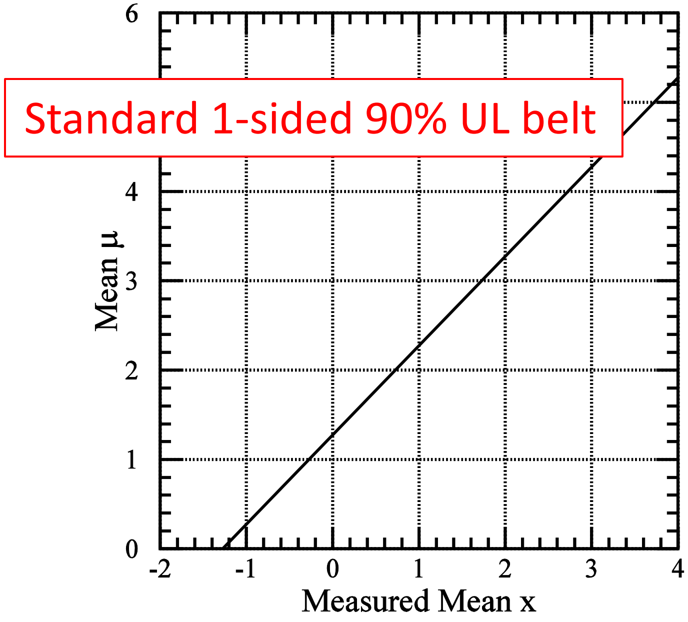
Intuitively, l is a reasonable test statistic for H_0: it is the maximum likelihood under H_0 as a fraction of its largest possible value, and large values of l signify that H_0 is reasonably acceptable. The critical region for the test statistic is therefore

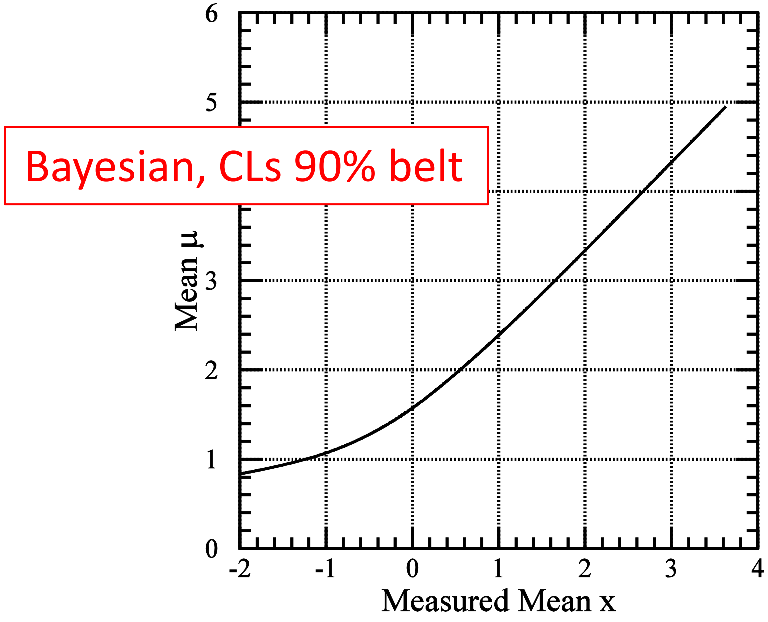
$$l \leq c_\alpha, \quad (22.6)$$

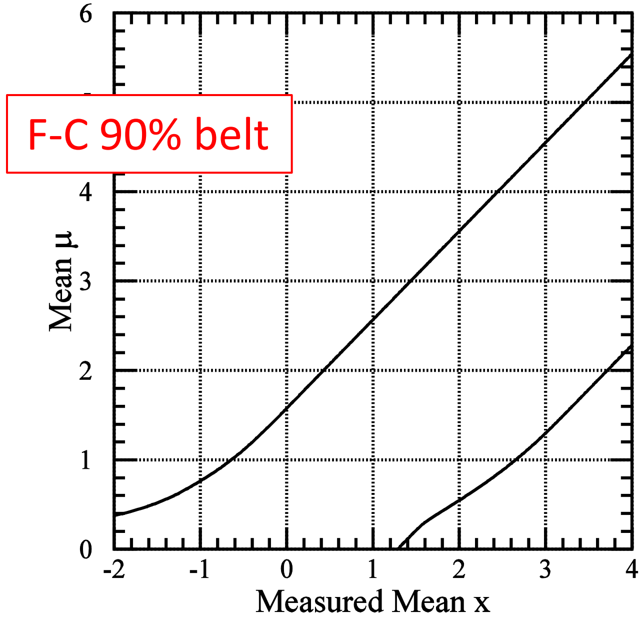
where c_α is determined from the distribution $g(l)$ of l to give a size-α test, that is,

$$\int_0^{c_\alpha} g(l)\,dl = \alpha. \quad (22.7)$$

Neither maximum value of the LF is affected by a change of parameter from $\boldsymbol{\theta}$ to $\tau(\boldsymbol{\theta})$, the ML estimator of $\tau(\boldsymbol{\theta})$ being $\tau(\hat{\boldsymbol{\theta}})$ – cf. **18.3**. Thus the LR statistic is invariant under reparametrization.


Standard 1-sided 90% UL belt
Mean μ
Measured Mean x
0
1
2
3
4
6
-2
-1
0
1
2
3
4


Bayesian, CLs 90% belt
Mean μ
Measured Mean x
0
1
2
3
5
6
-2
-1
0
1
2
3
4


F-C 90% belt
Mean μ
Measured Mean x
0
1
2
3
4
6
-2
-1
0
1
2
3
4

Confidence belts on bounded parameters

J.Bouchez
DAPNIA/SPP
CEA-Saclay
91191 Gif-sur-Yvette Cedex, France

January 13, 2000

Abstract

We show that the unified method recently proposed by Feldman and Cousins to put confidence intervals on bounded parameters cannot avoid the possibility of getting null results. A modified bayesian approach is also proposed (although not advocated) which ensures no null results and proper coverage.

http://arxiv.org/abs/hep-ex/0001036

Gary and I felt there was a misunderstanding, led to important clarification in PDG.

Statistica Sinica **19** (2009), 301-314

ON THE UNIFIED METHOD WITH NUISANCE PARAMETERS

Bodhisattva Sen, Matthew Walker and Michael Woodroofe

The University of Michigan

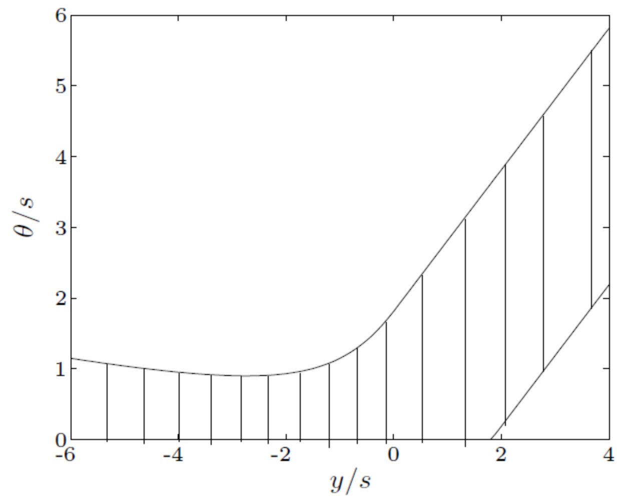


Figure 2.1. Confidence limits for θ/s as a function of y/s when $r = 10$ and $\alpha = 0.1$. Observe that the upper limit starts to increase as y decreases for $y < 0$.

[Recall McFarlane "Loss of Confidence"]

Contributed to 18th International Conference on Neutrino Physics and Astrophysics (NEUTRINO 98), Takayama, Japan, 4-9 Jun 1998.

Nuclear Physics B (Proc. Suppl.) 77 (1999) 212–219

The Search for Neutrino Oscillations $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ with KARMEN

K. Eitel[a] and B. Zeitnitz[a] for the KARMEN collaboration[1]

In the investigated data, no sequential structure fulfilled all the required properties for a (e^+,n) sequence. After all cuts, the remaining background amounts to only 2.88 ± 0.13 events caused by sequential cosmic background and ν induced sequences. These background sources are described in detail in the following section. The probability of measuring zero events with an expected number of 2.88 ± 0.13 background events is 5.6%. Applying a unified approach [8], we deduce an upper limit of $N < 1.07$ (90% CL) for a potential $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillation signal. With an expectation of $N = 811 \pm 89$ for $\sin^2(2\Theta) = 1$ and large Δm^2 this corresponds to a limit of

$$\sin^2(2\Theta) < 1.3 \cdot 10^{-3} \qquad (90\%\,\mathrm{CL}) \qquad (2)$$

for $\Delta m^2 \geq 100\,\mathrm{eV}^2/\mathrm{c}^4$. Fig 3 shows the KARMEN2 exclusion curve in comparison with other experiments.

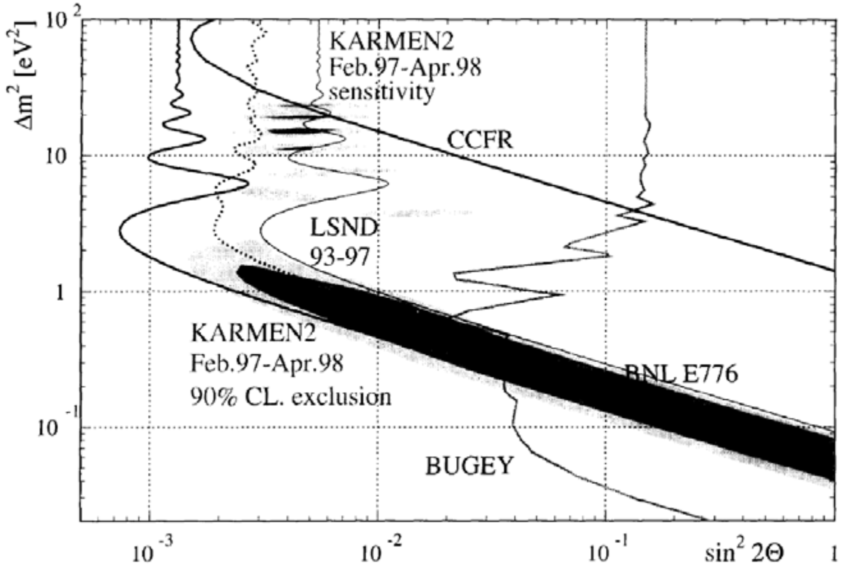


Figure 3. KARMEN2 90% CL exclusion limit and sensitivity compared to other experiments: BNL [9], CCFR [10], BUGEY [11] and the evidence for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ oscillations reported by LSND [12].

8. G.J. Feldman and R.D. Cousins, Phys. Rev. D **57**, 3873 (1998).

Quick reminder of intervals on Poisson mean

Adapted from R. Cousins, Am. J. Phys. 63 398 (1995)

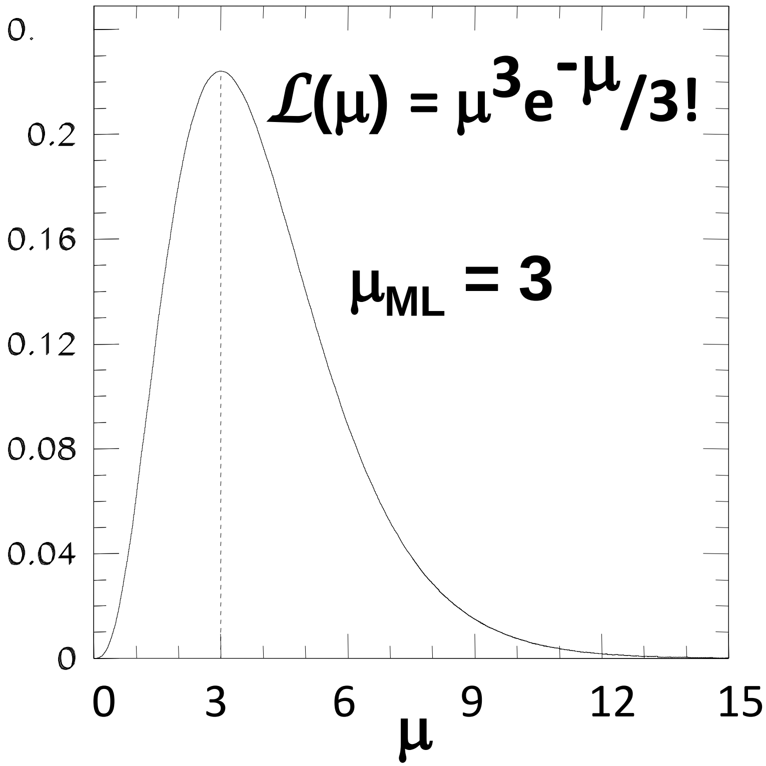
Poisson process $P(n|\mu) = \mu^n e^{-\mu}/n!$

Measurement of n yields n=3.

Substituting n=3 into $P(n|\mu)$ yields the *Likelihood function* $\mathcal{L}(\mu)$.

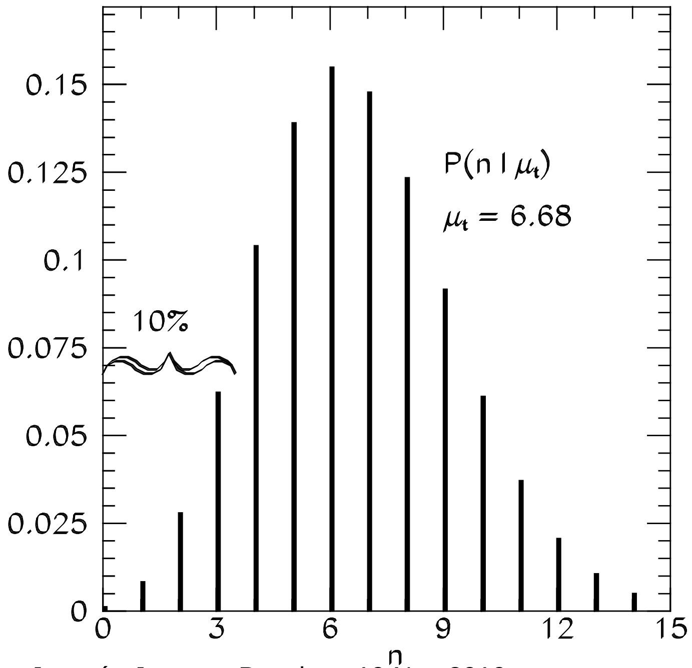
It is tempting to consider area under $\mathcal{L}$, but $\mathcal{L}(\mu)$ is *not* a probability density in μ:

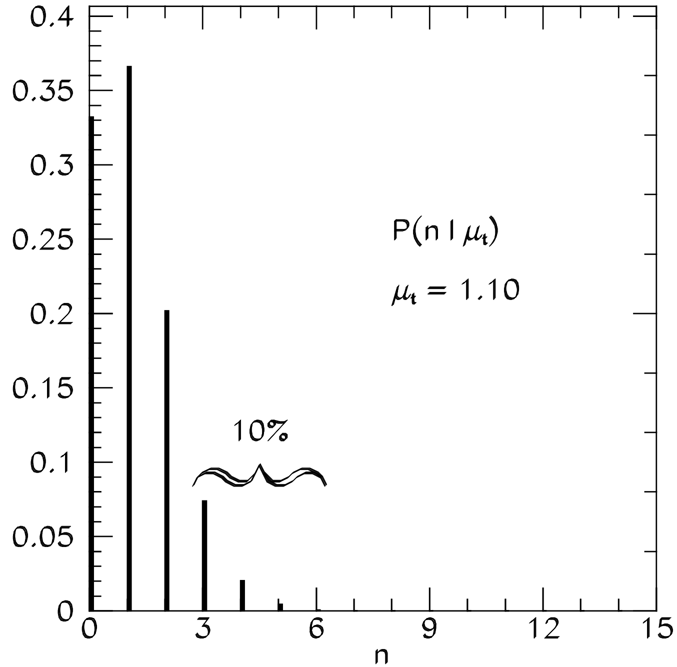
Area under $\mathcal{L}$ is meaningless.



How to get upper (or lower) limit on μ ?
Consider 90% *upper* and 90% *lower* limits on μ.
Together they form an 80% *central interval* for μ.

1) *Frequentist confidence limit* method:
Find μ_u s.t. Poisson $P(n \leq 3 \mid \mu_u) = 0.1$. $\mu_u = 6.68$
Find μ_ℓ s.t. Poisson $P(n \geq 3 \mid \mu_\ell) = 0.1$. $\mu_\ell = 1.10$

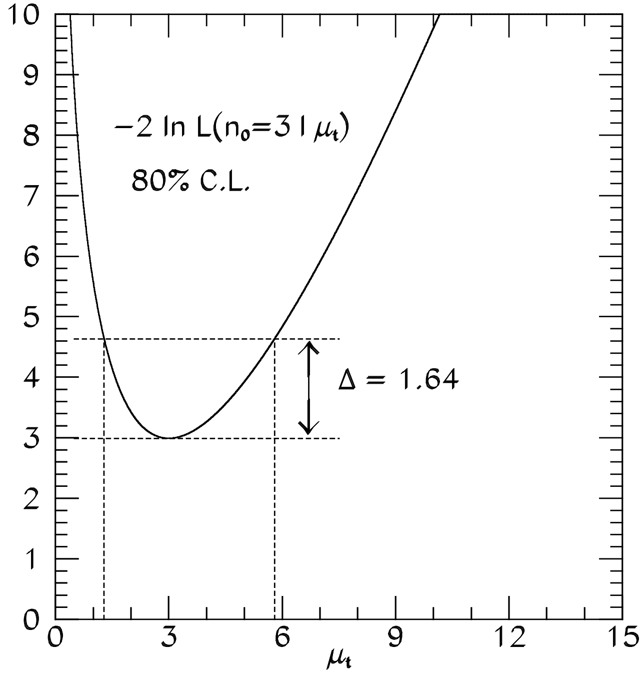




2) *Likelihood ratio* method.
Based on $\mathcal{L}(\mu) / \mathcal{L}(\mu_{ML})$, equivalently:

$-2\ln\mathcal{L}(\mu) - (-2\ln\mathcal{L}(\mu_{ML})) \leq Z^2$, for Z real.

Asymptotically (note regularity conditions) this interval approaches a frequentist central confidence interval with C.L. corresponding to ± Z Gaussian standard deviations.



For 80% central interval, Z=1.28.
90% upper and lower limits are:
$\mu_u = 5.80$
$\mu_\ell = 1.29$

3) *Bayesian* method.

Different definition of probability: *degree of belief*.

With that definition, one can have pdf's in μ (!)

$p(\mu|n=3) \propto \mathcal{L}(\mu)\, p(\mu)$,

- **$p(\mu|n=3)$ = *posterior* pdf for μ, given n=3**
- **$\mathcal{L}(\mu)$ = Likelihood function from above for n=3**
- **$p(\mu)$ = *prior* pdf for μ, before incorporating n=3.**

Vast literature on Bayesian methods and priors.
This literature has largely been ignored in HEP, where most papers use uniform prior for μ.
Bayesian statisticians call this "pseudo-Bayesian".

Deep Foundational Issue: Confidence Principle (Frequentist Coverage) vs Likelihood Principle

The Likelihood ratio interval and the Bayesian interval use $\mathcal{L}(\mu)$ given the *observed* n=3, but make *no* use of P(n|μ) for any n≠3. This is the essence of the *Likelihood Principle*.

The confidence interval relying on P(n≤3 | μ) and P(n≥3 | μ) used *probabilities of data not observed*. This violates the L.P.

This turns out to be *very important*:
In general, cannot have both coverage and L.P.
Whole approach of tail probabilities violates L.P. !

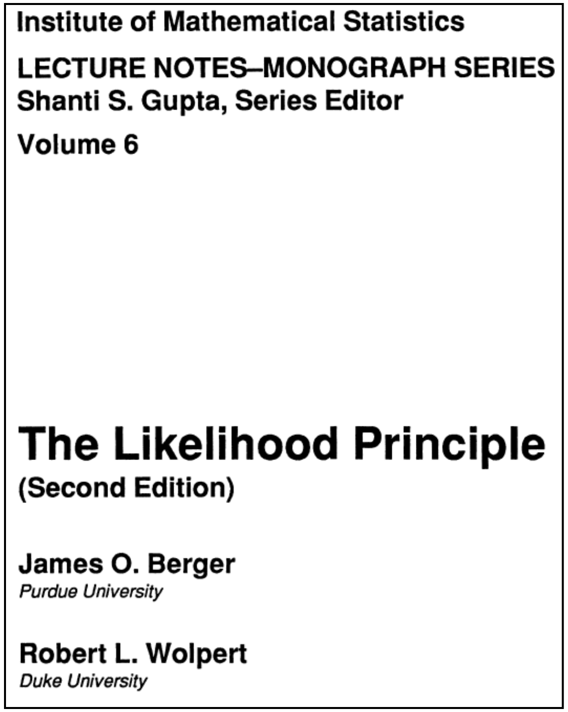
The Karmen Problem is a Classic L.P. Issue!

- The "Karmen Problem"
  - You expect background events sampled from a Poisson mean b=2.8, assumed known precisely.
  - For signal mean μ, the total number of events n is then sampled from Poisson mean μ+b.
  - So $P(n) = (\mu+b)^n \exp(-\mu-b)/n!$
  - Observe n=0.
  - $\mathcal{L}(\mu) = (\mu+b)^0 \exp(-\mu-b)/0! = \exp(-\mu) \exp(-b)$
- Changing b from 0 to 2.8 changes $\mathcal{L}(\mu)$ only by the constant factor exp(-b). This gets renormalized away in any Bayesian calculation, and is irrelevant for likelihood *ratios*.
- So for observed n=0, likelihood-based inference about signal mean μ is *independent of expected b*.
- For essentially all frequentist confidence interval constructions, the fact that n=0 is less likely for b=2.8 than for b=0 results in *narrower* confidence intervals for μ as b increases. Clear violation of the L.P.

Likelihood Principle Discussion

We will not resolve this issue, but should be aware of it.

- See book by Berger & Wolpert, but be prepared for the "Stopping Rule Principle" to set your head spinning.
- When frequentist intervals and limits badly violate the L.P., use great caution in interpreting them!
- And when Bayesian inferences badly violate the Confidence Principle (frequentist coverage), again use great caution!


Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES
Shanti S. Gupta, Series Editor
Volume 6

The Likelihood Principle
(Second Edition)

James O. Berger
Purdue University

Robert L. Wolpert
Duke University

“Lucky” Data: Events with high power

Physics Letters B 292 (1992) 221–228

A measurement of the tau mass

ARGUS Collaboration

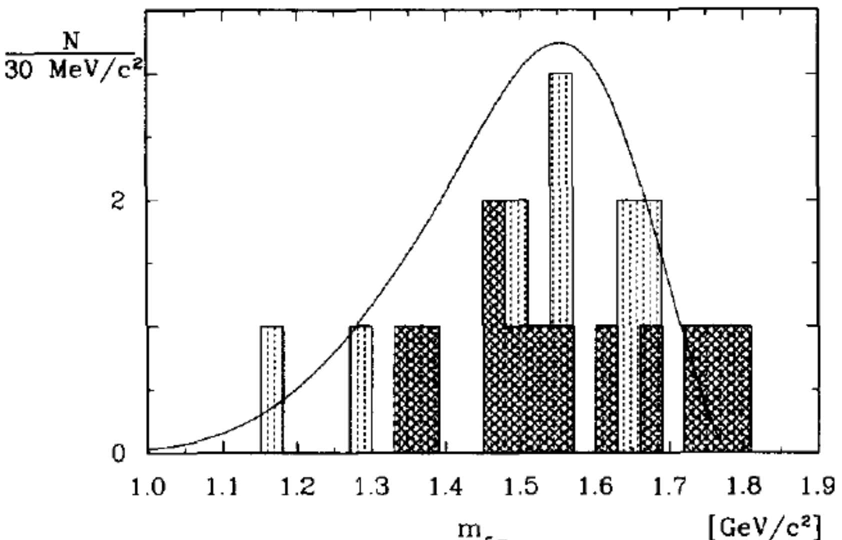


Fig. 6. Measured invariant 5π mass spectrum (histogram), where the hatched part displays the result of our previous analysis (see text). The curve corresponds to the expected shape of a phase-space decay weighted with the weak matrix element (assuming $m_{\nu_\tau} = 0$ MeV/c^2). Note that the curve has not been normalized to the data.

1987 limit with 12 events was so “lucky” that 1992 limit with 20 events was the same.

(Event through in both cases, highest 5π mass was removed in order to account for possible uncertainties in background.)

$m_{\nu_\tau}^2 < 35$ MeV ^{2}

(<31 MeV ^{2} with new τ mass)

Limit on the Tau Neutrino Mass

(CLEO Collaboration)

To compare these results to the ARGUS result it is informative to calculate the quantity P_{MC}, defined as the percentage of signal Monte Carlo experiments in which a 95% C.L. limit less than or equal to that of the data is obtained. Since each of these experiments is generated at

TABLE I. The size of the final event samples, the 95% C.L. limit on M_{ν_τ} after correction for systematic efforts, $M_{\nu_\tau}^{95}$, and the estimated Monte Carlo probability of having obtained this limit, P_{MC}, for each of the studied decay modes and for the published ARGUS result. The errors on P_{MC} are statistical only.

Decay mode	Sample size	$M_{\nu_\tau}^{95}$ (MeV)	P_{MC} (%)
$\tau^- \to 3h^- 2h^+ \nu_\tau$	60	47.5	34.9 ± 0.8
$\tau^- \to 2h^- h^+ 2\pi^0 \nu_\tau$	53	33.7	4.3 ± 0.3
Combined	113	32.6	13.9 ± 0.6
ARGUS	20	31	0.041 ± 0.012

Conditioning

- **"Ancillary statistic": a function of your data which carries information about the precision of your measurement, but no info about parameter's value.**
- **E.g.: branching ratio measurement in which the total number of events N can fluctuate if the experimental design is to run for a fixed length of time.**
  Then N is an ancillary statistic.
- **You perform an experiment and obtain N total events, and then do a toy M.C. of repetitions of the experiment. Do you let N fluctuate, or do you fix it to the value observed?**
- **It may seem that the toy M.C. should include your *complete* procedure, including fluctuations in N.**
- **But there are strong arguments, going back to Fisher, that inference should be based on probabilities *conditional on the value of the ancillary statistic actually obtained*!**

Conditioning (cont.)

- **1958 thought experiment of David R. Cox focused the issue:**
  - **Your procedure for weighing an object consists of flipping a coin to decide whether to use a weighing machine with a 10% error or one with a 1% error; and then measuring the weight. (Coin flip result is ancillary stat.)**
  - **Then "surely" the error you quote for your measurement should reflect which weighing machine you actually used, and not the average error of the "whole space" of all measurements!**
  - **But classical most powerful Neyman-Pearson hypothesis test uses the whole space!**
- **In more complicated situations, ancillary statistics do not exist, and it is not at all clear how to restrict the "whole space" to the relevant part for frequentist coverage.**

Conclusion

Some of the statistics controversies in neutrino physics are "classic" cases of foundational issues in the professional statistics literature…

Procedures pioneered by neutrino physicists have had large impact on the greater HEP community…

…and problematic data sets still arise:

arXiv:1010.4439

Search for Quark Compositeness with the Dijet Centrality Ratio in pp Collisions at $\sqrt{s} = 7$ TeV

The CMS Collaboration*

From Abstract: A statistical analysis of the data provides a lower limit on the energy scale of quark contact interactions. The sensitivity of the analysis is such that the expected limit is 2.9 TeV; because the observed value of the centrality ratio at high invariant mass is below the expectation, the observed limit is 4.0 TeV at the 95% confidence level.

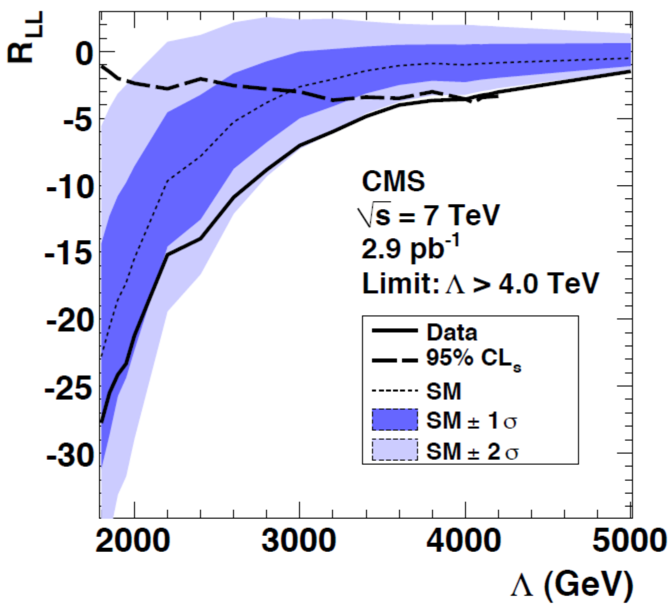
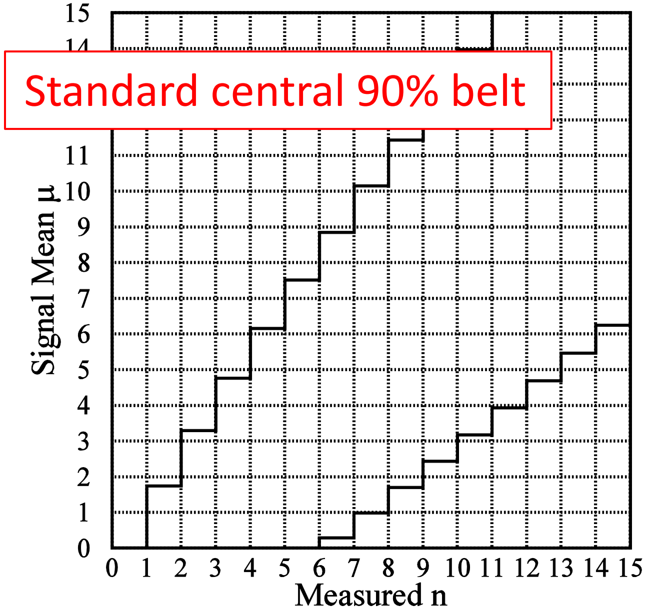
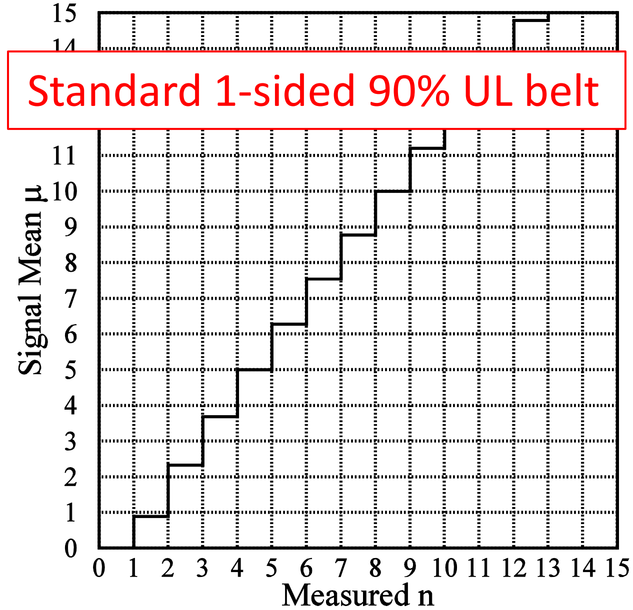
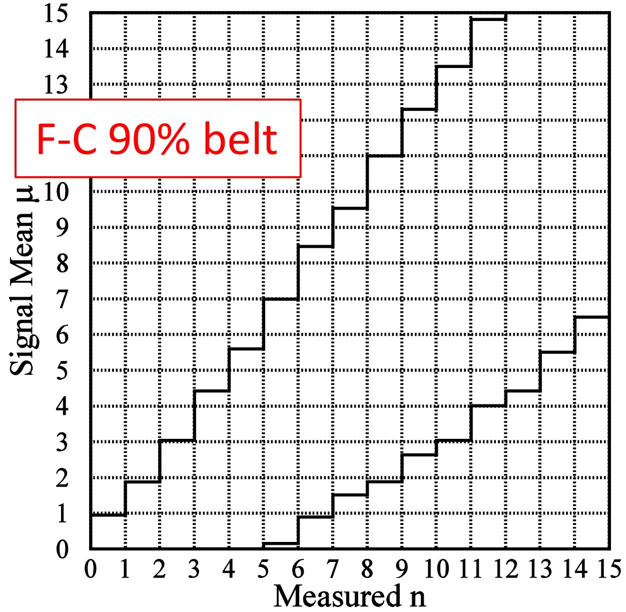


Figure 3: Summary of the limit for the contact interaction scale Λ. We show $\mathcal{R}_{LL}$ versus Λ for the data (solid line), the 95% CL_s (dashed line), and the SM expectation (dotted line) with 1σ (dark) and 2σ (light) bands.





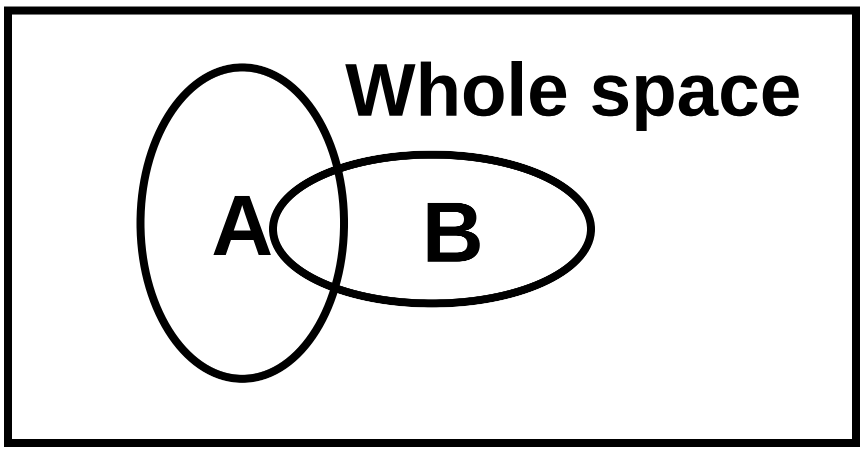


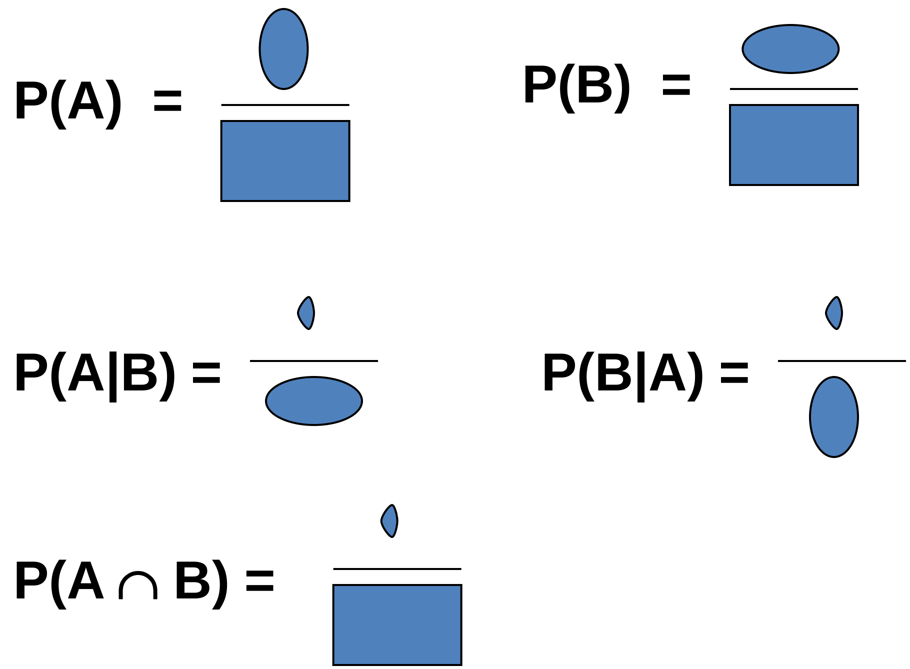
Discussion on next page.

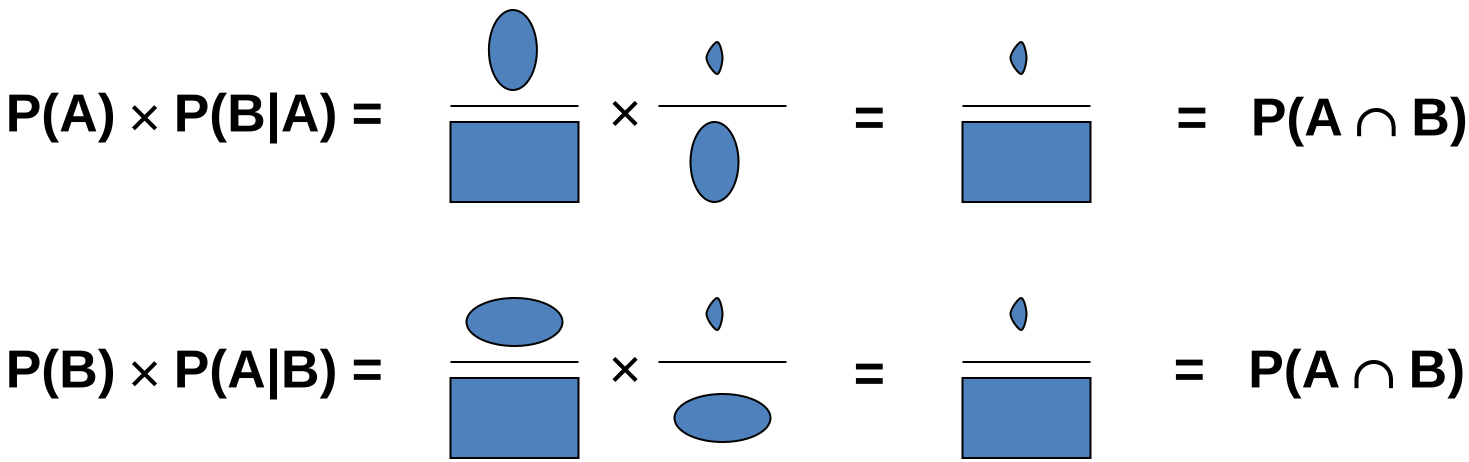
Summary of Three Ways to Make Intervals

	Bayesian Credible	Frequentist Confidence	Likelihood Ratio
Requires prior pdf?	Yes	No	No
Obeys likelihood principle?	Yes (exception re Jeffreys prior)	No	Yes
Random variable in "$P(\mu_t \in [\mu_1, \mu_2])$":	μ_t	μ_1, μ_2	μ_1, μ_2
Coverage guaranteed?	No	Yes (but over-coverage…)	No
Provides P(parameter|data)?	Yes	No	No

P, Conditional P, and Derivation of Bayes' Theorem in Pictures







$\Rightarrow$ **P(B|A) = P(A|B) × P(B) / P(A)**

2010 RPP

$\overline{\nu}$ MASS (electron based)

Those limits given below are for the square root of $m^{2(\mathrm{eff})}_{\nu_e} \equiv \sum_i |U_{ei}|^2 m^2_{\nu_i}$. Limits that come from the kinematics of ^{3}H $\beta^- \overline{\nu}$ decay are the square roots of the limits for $m^{2(\mathrm{eff})}_{\nu_e}$. Obtained from the measurements reported in the Listings for "$\overline{\nu}$ Mass Squared," below.

VALUE (eV)	*CL%*	*DOCUMENT ID*		*TECN*	*COMMENT*
< 2 OUR EVALUATION					
< 2.3	95	[1] KRAUS	05	SPEC	^{3}H β decay
< 2.5	95	[2] LOBASHEV	99	SPEC	^{3}H β decay

$\overline{\nu}$ MASS SQUARED (electron based)

Given troubling systematics which result in improbably negative estimators of $m^{2(\mathrm{eff})}_{\nu_e} \equiv \sum_i |U_{ei}|^2 m^2_{\nu_i}$, in many experiments, we use only KRAUS 05 and LOBASHEV 99 for our average.

VALUE (eV^2)	*CL%*	*DOCUMENT ID*		*TECN*	*COMMENT*
− 1.1± 2.4 OUR AVERAGE					
− 0.6± 2.2± 2.1		[15] KRAUS	05	SPEC	^{3}H β decay
− 1.9± 3.4± 2.2		[16] LOBASHEV	99	SPEC	^{3}H β decay

Tau Neutrino Mass

9 July 1998

ELSEVIER

PHYSICS LETTERS B

Physics Letters B 431 (1998) 209–218

A limit on the mass of the ν_τ

CLEO Collaboration

The resulting extended likelihood is shown in Fig. 4. We define [8] the 95% confidence level (CL) upper limit by integrating defined likelihood above zero mass to its ninety-fifth percentile. We find 95% CL upper limits of

[8] The method for extracting an upper limit from a likelihood distribution at a given confidence level is not unambiguously defined; the method used here differs from that used in the analysis of Ref. [8]. Therefore comparisons of upper limits among different experiments must be done with care.

20. STATISTICS

Revised June 1994 with the help of R. Cousins, F. James, G. Lynch, B. P. Roe, and M. Roos.

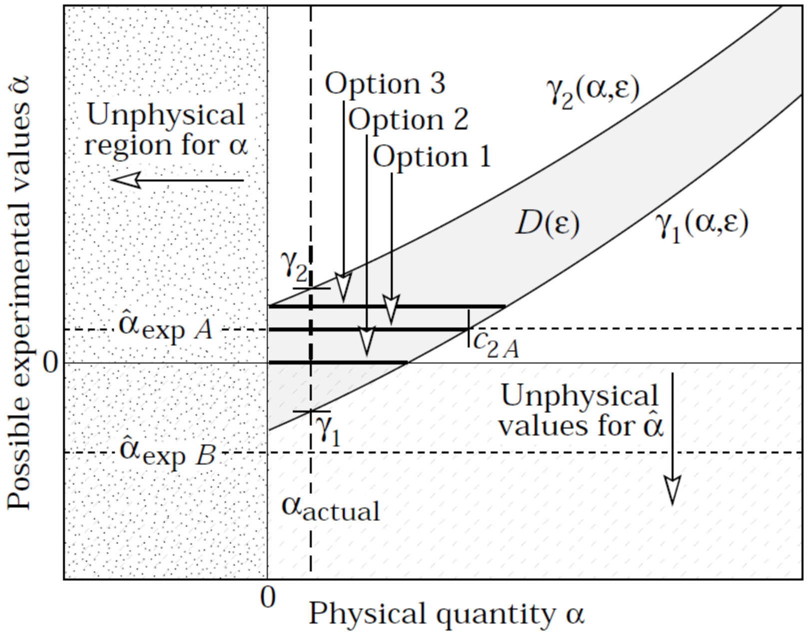


Figure 20.5: The situation near a physical boundary. In Fig. 20.1 the horizontal line for a given $\widehat{\alpha}_{\rm exp}$ crossed the domain $D(\varepsilon)$, bounded by $\gamma_1(\alpha,\epsilon)$ and $\gamma_2(\alpha,\epsilon)$ entirely in the physical region, entering at c_1 and leaving at c_2. The limits γ_1 and γ_2 cannot be defined in a region where α is not defined, so the functions cannot be continued into the unphysical region. As a result c_1 (for experiment A) or c_1 and c_2 (for experiment B) cannot be defined. Options 1, 2, and 3 label the ways one might define confidence intervals, as described in the text.

2. The Bayesian approach [3]. This is the approach favored in the older literature, and has (unfortunately and incorrectly) been referred to as the "PDG method" in certain papers. To begin with, it is argued

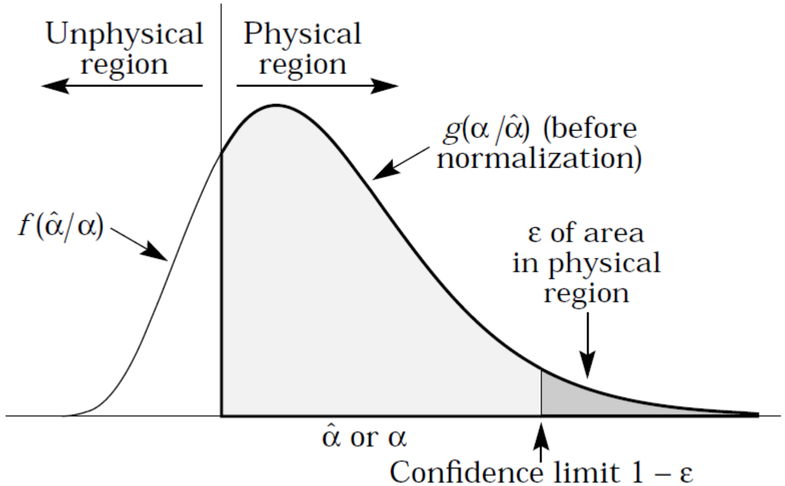


Figure 20.6: An example of a bounded physical region, in which a measurement $\widehat{\alpha}$ can fall in an unphysical region with significant probability. If we assume that α, the quantity we are trying to measure, cannot lie in the unphysical region (0 probability) but can lie anywhere in the physical region ("no prior knowledge"), then Bayes' theorem says that our new knowledge of the distribution of α, given our measurement $\widehat{\alpha}$, is given by the shaded function after appropriate renormalization.