Multiple Measurements and Parameters in the Unified Approach

Gary Feldman
Workshop on Confidence Limits
Fermilab
March 28, 2000

Origins

- The Unified Approach was designed to be completely general. The exact same approach is used for simple problems and complex. Therefore, the method of combining signals is uniquely specified.

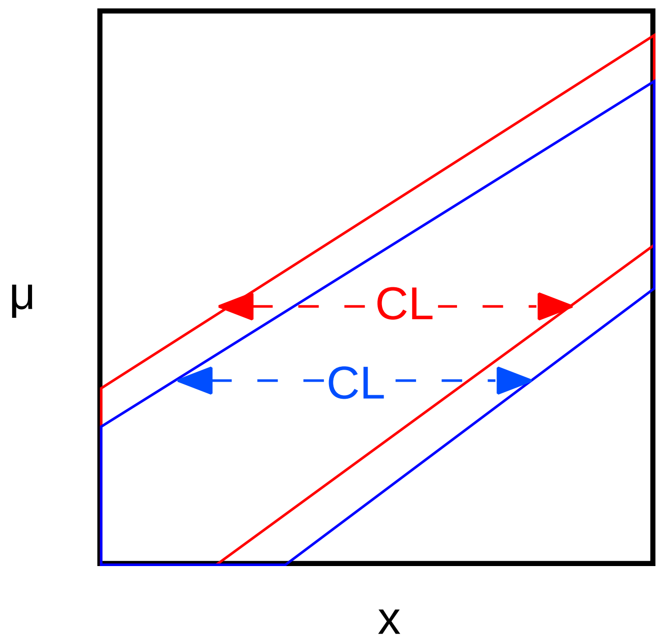
- Meeting with Harvard statisticians:
  - The Unified Approach is the "standard method," but no known prior examples.
  - Confidence intervals are equivalent to hypothesis tests.
  - The likelihood ratio provides the most powerful hypothesis test (Neyman-Pearson theorem).
  - Therefore, it is reasonable to use the likelihood ratio for constructing confidence intervals. However, no uniformly most powerful test. (see figure)
  - Discovery of prior publication by Kendall and Stuart in 1961, including treatment of nuisance parameters. (see figure)

Lack of Uniformly Most Powerful Test

- Error of the first kind:

  Rejecting a true hypothesis ≡ coverage.
- Error of the second kind:

  Accepting a false hypothesis ≡ power.



- Deciding which is more powerful is not possible because frequentists do not admit a prior distribution for μ.

Kendall and Stuart

From M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Volume 2: Inference and Relationship (1961):

CHAPTER 24

LIKELIHOOD RATIO TESTS AND THE GENERAL LINEAR HYPOTHESIS

Kendall and Stuart define:

x	vector of measurements
θ_r	vector of unknown parameters with θ_{r_0} representing the parameters of the null hypothesis H_0 (read unknown true parameters)
θ_s	vector of nuisance parameters
$\hat{\theta}_r, \hat{\theta}_s$	unconditionally maximize $L(x \mid \hat{\theta}_r, \hat{\theta}_s)$
$\hat{\hat{\theta}}_s$	conditionally maximizes $L(x \mid \theta_{r_0}, \hat{\hat{\theta}}_s)$

then

Kendall and Stuart (continued)

Now consider the likelihood ratio

$$l = \frac{L(x\,|\,\theta_{r0}, \hat{\hat{\theta}}_s)}{L(x\,|\,\hat{\theta}_r, \hat{\theta}_s)}. \qquad (24.4)$$

. . .

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

$$l \leqslant c_\alpha, \qquad (24.6)$$

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

$$\int_0^{c_\alpha} g(l)\,dl = \alpha. \qquad (24.7)$$

Or in readable form:

"Now consider the likelihood ratio

$$l = \frac{L(x\,|\,\theta_{r_0}, \hat{\hat{\theta}}_s)}{L(x\,|\,\hat{\theta}_r, \hat{\theta}_s)}. \qquad (24.4)$$

…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

$$l \le c_\alpha, \qquad (24.6)$$

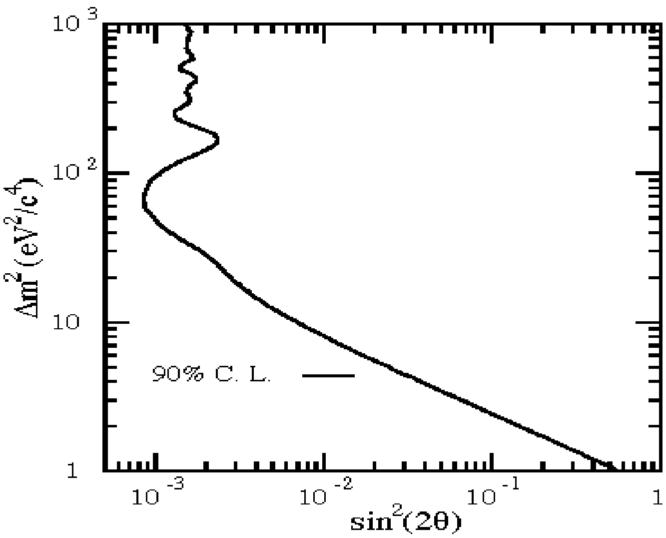
where c_α is determined from the distribution $g(l)$ of l to give a size-α test, i.e.

$$\int_0^{c_\alpha} g(l)dl = \alpha."$$

[Warning the c.l. is $1 - \alpha$.]

Examples from Neutrino Oscillations

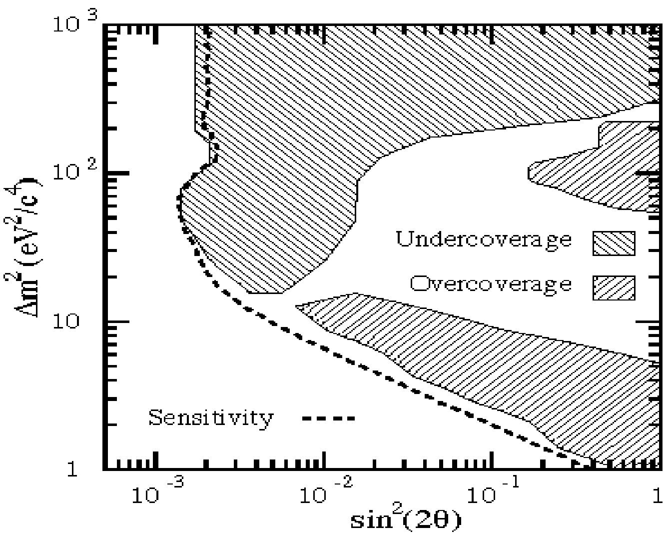
- The Unified Approach is often more intuitive for complicated problems than for simple ones, although it is identical for both.
- For example, in neutrino oscillations, most physicists' intuitive approach is to find a minimum of χ^2 and to "go up" by a fixed amount (4.61 for 90% c.l.) to set a confidence limit.



Neutrino Oscillations (continued)

- Since $\chi^2 = -2\ln L$, this is the same thing that one does in the Unified Approach, except that instead of a Gaussian approximation, one evaluates the integral $\int_0^{c_\alpha} g(l)dl = \alpha$ at each point to calculate the equivalent of the 4.61.

- In a toy model, the use of the Gaussian approximation leads to significant under- and overcoverage (76% and 94%):



Neutrino Oscillations (continued)

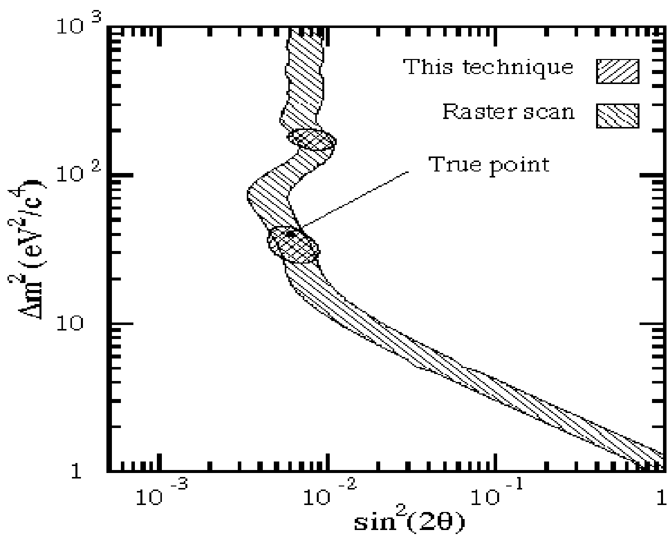
- In a simple case, the evaluation of the integral is just a sum over discrete probabilities, or a integral in one variable. In a more complicated case, such as neutrino oscillations, or combining results of several experiments, the integral is best done by Monte Carlo techniques.

- One computational simplification is that one only has to evaluate the integral in the region of the limit. The evaluation of the integral can be halted as soon as it is clear whether it is less than or greater than α. I.e., you know what c_α is for your experimental data. Thus you can simultaneously start evaluating two integrals and halt whenever one of the following conditions is met:

$$\int_0^{c_\alpha} g(l)dl > \alpha \text{ or } \int_{c_\alpha}^{\infty} g(l)dl > 1-\alpha .$$

Aside on Power

- One way that has been used to set confidence intervals in neutrino oscillation experiments is to do a “raster scan.” For each value of Δm^2, one finds the minimum of the likelihood and goes up 2.71 in χ^2. This gives exact coverage, but poor power compared to the Unified Approach*:



- Assuming, of course, that you do not have a highly peaked prior.

Nuisance Parameters

- A nuisance parameter is an unknown parameter whose value is not of interest, but for which coverage must be provided for all possible values.
- In this talk I will be mainly concerned with the true rate of background production as a nuisance parameter.
- Obtaining exact coverage for nuisance parameters is a cumbersome procedure at best, and computationally impossible in complicated cases. Therefore, statisticians often use the approximate procedure suggested by Kendall and Stuart of eliminating the nuisance parameters by maximizing the likelihood with respect to them.

$$l(x,\theta_{r_0}) = \frac{L(x\,|\,\theta_{r_0},\hat{\hat{\theta}}_s)}{L(x\,|\,\hat{\theta}_r,\hat{\theta}_s)}.$$

  The idea is that if one covers for $\hat{\hat{\theta}}_s$, the values most favorable to θ_{r_0}, then one is likely to cover for all θ_s. Our preliminary studies show that this is true to a high degree.
- The maximizations can be done analytically in simple cases, and numerically in more complex cases.

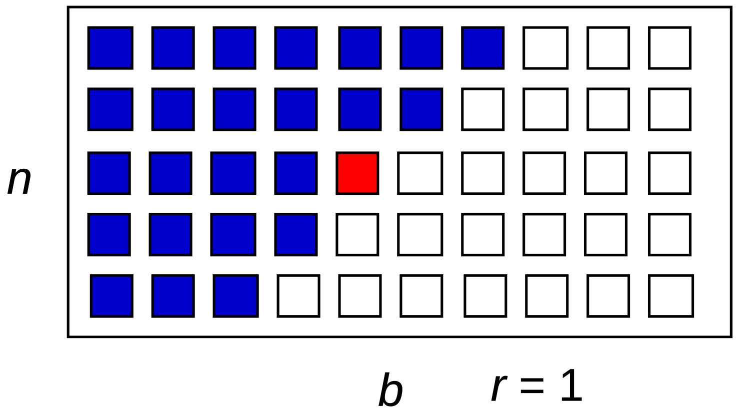
A Subtlety

- Consider the case of counting experiment in which *n* events are observed, and the background is estimated by an ancillary experiment (side-band, empty target, etc.) in which *b* events are observed, such that the expected background is *rb* events. We wish to find a confidence interval for μ, the unknown true rate of signal production, and β, the unknown true rate of background production is the nuisance parameter.

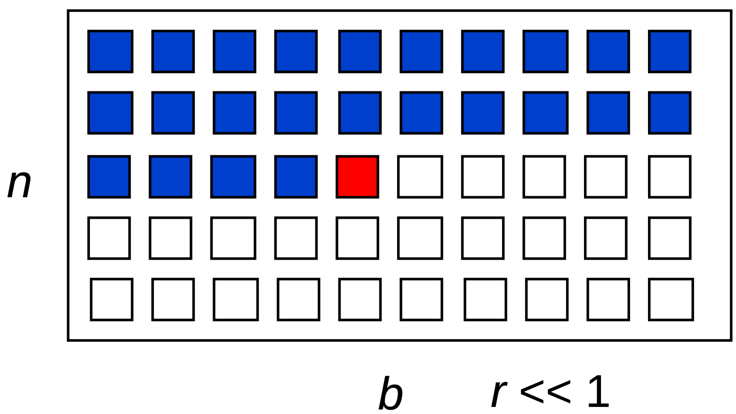
- As $r \to 0$, *b* becomes equal to β to high precision, and we expect the confidence interval to approach the value it would have if β were known exactly. This does not happen if we follow the outlined procedure. The reason is that we normally overcover due to discreteness. The introduction of a nuisance parameter reduces the effect of discreteness, and thus reduces the overcoverage.

- The problem and our tentative solution are illustrated on the next two transparencies.

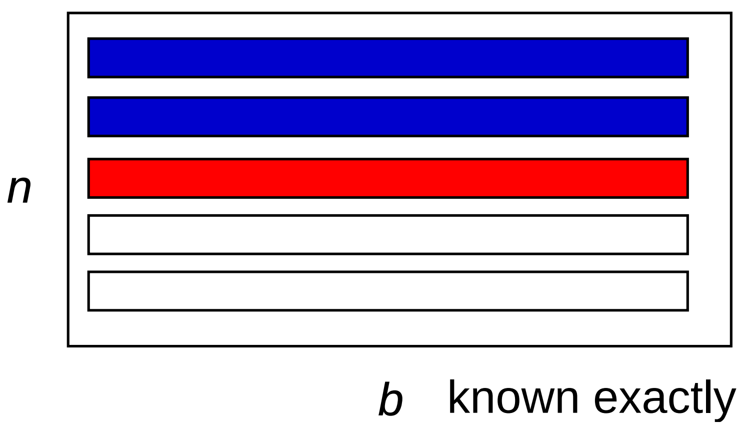
A Subtlety, Illustrated



$r = 1$

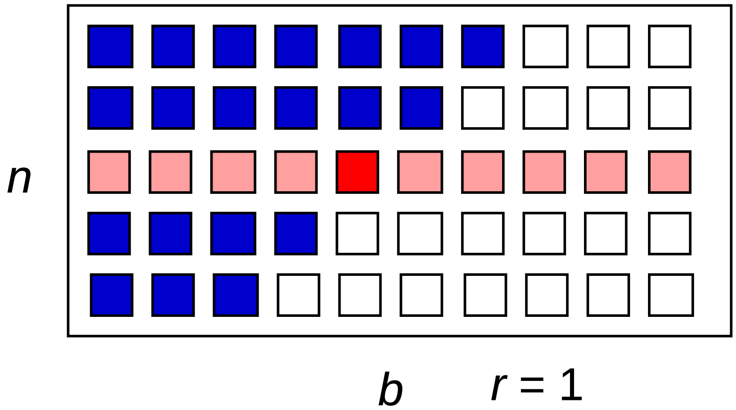


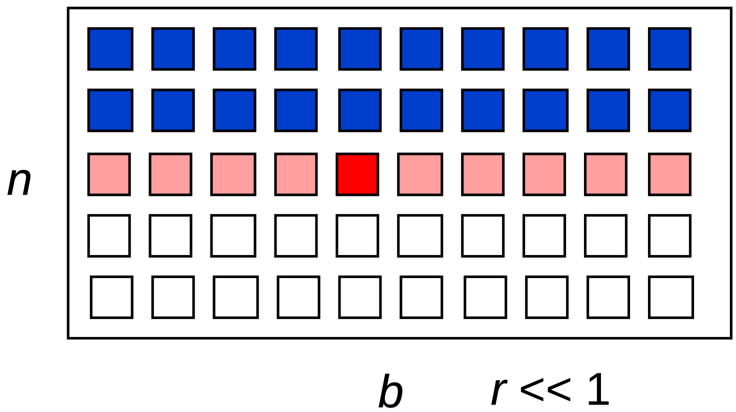
$r << 1$

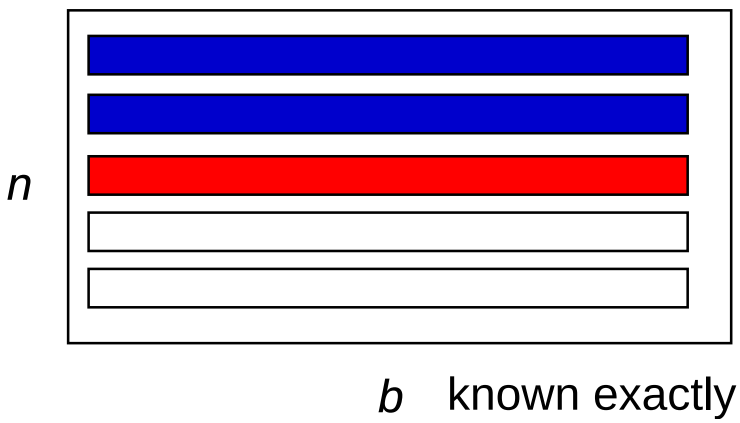


known exactly

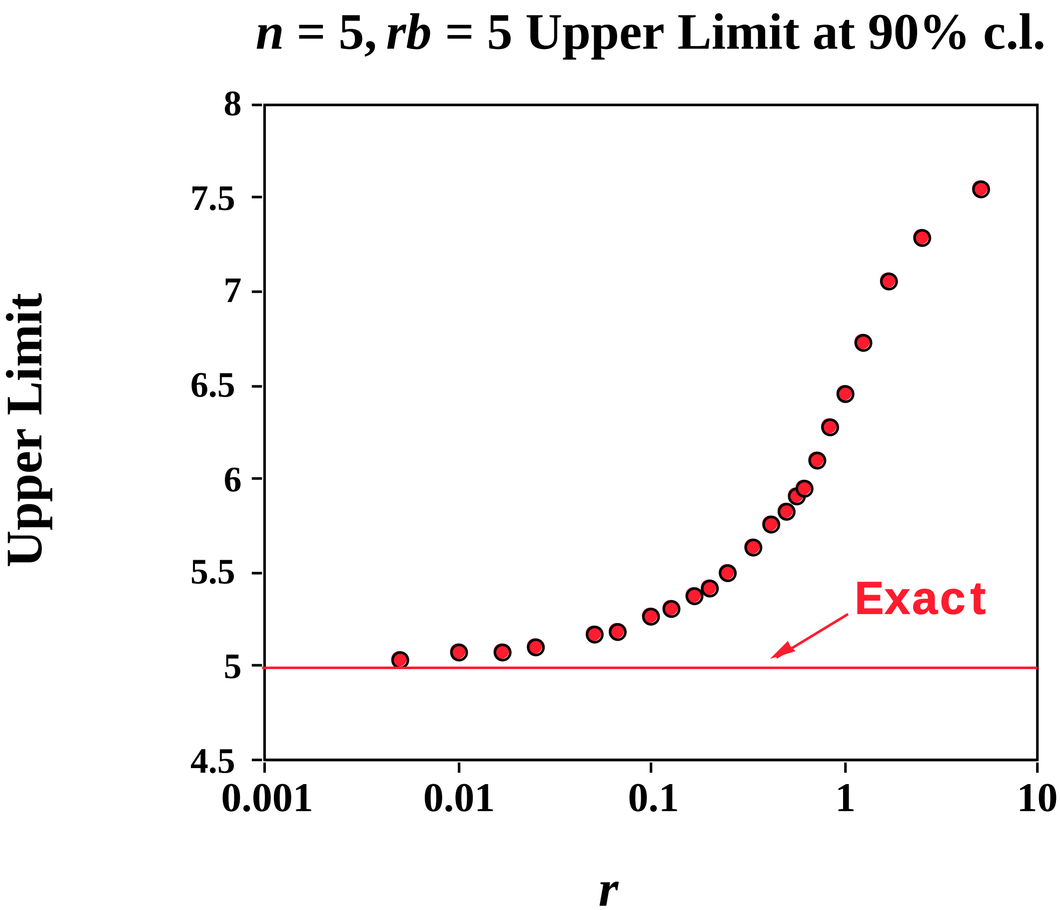
A Subtlety, Our Tentative Solution







A Simple Example



The NOMAD Experiment

- The NOMAD experiment at CERN searched for $\nu_\mu \rightarrow \nu_\tau$ oscillations in the mass region of cosmological significance, a few eV/c^2. The τ's from ν_τ interactions are identified purely by kinematical criteria.

- Searches for τ decays are made in several decay modes, and within each mode, the data may be binned by the kinematic criteria or by relative sensitivity (i.e., one gains sensitivity by treating regions of low background separately from regions of high background). Thus each bin is like a separate experiment.

- The Monte Carlo does not adequately describe the data, so backgrounds must be based in part on a data simulator: The muons in ν_μ charged current events are removed and replaced by electrons to simulate ν_e charged current events or by nothing to simulate neutral current events. The number of charged current events thus limits the accuracy with which backgrounds can be known. Thus, each mode has the true rate of background production as a nuisance parameter.

- The analyses are blind. The binning is determined prior to opening the box.

Results of the NOMAD Experiment

- Backgrounds are modeled as an equivalent Poisson measurement plus, optionally, a component that is known with high precision. Below is the approximate modeling for the 22 NOMAD bins.

Mode	Bin	N_τ	*rb*	*r*	*n*
$\tau \to e\bar{\nu}\nu$ DIS HE	1	134	0.9	0.08	2
	2	128	0.5	0.12	1
	3	639	0.2	0.20	0
	4	535	1.9	0.03	2
	5	389	0.8	0.03	0
	6	1388	0.2	0.05	0
$\tau \to e\bar{\nu}\nu$ DIS LE	1	247	0.8	0.09	0
	2	650	0.3	0.08	0
$\tau \to e\bar{\nu}\nu$ LM	1	282	3.1	0.15	3
	2	285	1.5	0.12	2
	3	292	0.8	0.20	1
$\tau \to \pi\nu$ DIS	1	817	4.4	0.88	3
	2	1205	2.4	0.27	2
$\tau \to \pi\nu$ LM	1	357	6.7	0.84	5

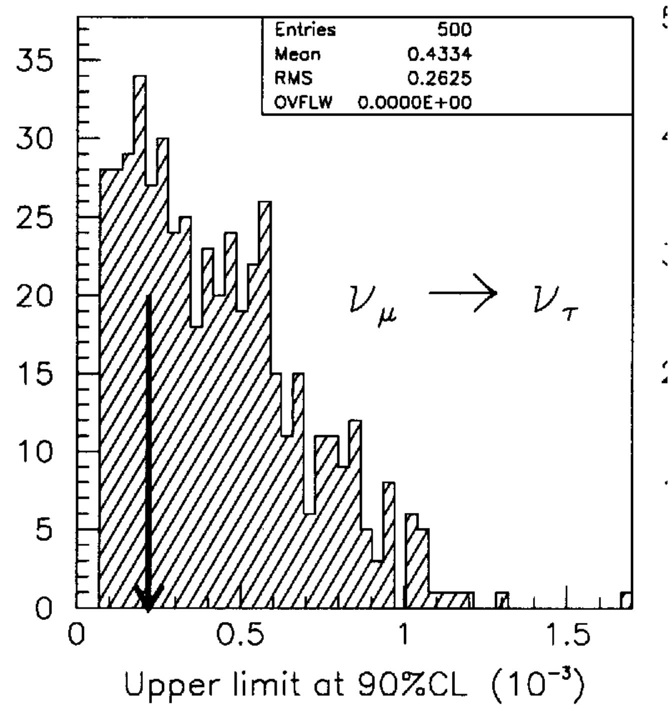
Continued…

Results of the NOMAD Experiment (cont.)

Mode	Bin	N_τ	*rb*	*r*	*n*
$\tau \to \rho\nu$ DIS 1γ	1	883	6.1	0.61	5
	2	1736	0.3	0.30	0
$\tau \to \rho\nu$ DIS 2γ	1	466	3.0	0.75	2
	2	222	0.0	0.88	0
$\tau \to \rho\nu / \pi\nu$ DIS	1	210	0.0	0.74	1
$\tau \to \rho\nu$ LM	1	458	5.2	0.65	7
$\tau \to 3\pi\nu$ DIS	1	1820	9.6	0.60	9
$\tau \to 3\pi\nu$ LM	1	288	3.5	0.44	5
Totals	22	13431	52.0		50

Results of the NOMAD Experiment (cont.)

- The upper limit at 90% c.l. on the oscillation probability is 2.2×10^{-4}. The experimental sensitivity is 4.3×10^{-4}. This is an indication that in the most sensitive bins the expected number of events was slightly lower than the expected background.



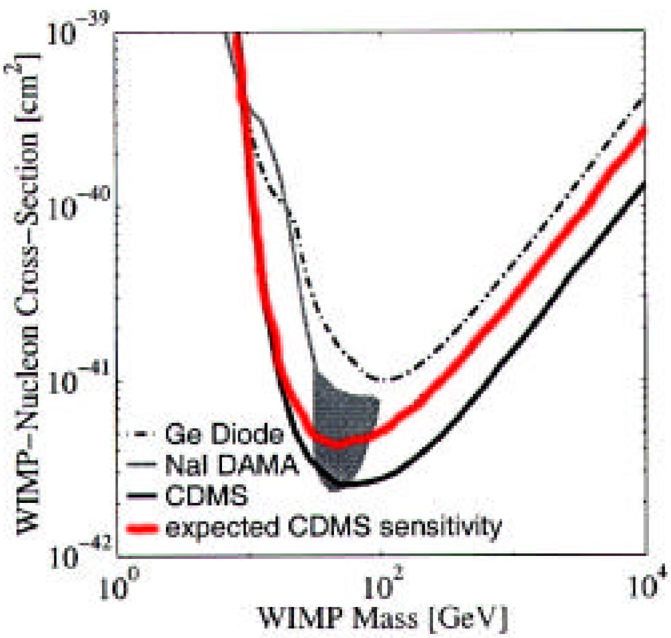
- However, this sensitivity does correspond to the mode of the distribution:

- If all of the bins had just been added together, the upper limit would have been 8.3×10^{-4} and the sensitivity would have been 9.7×10^{-4}.

Comments on the CDMS Limit

- Richard Schnee presented the CDMS analysis in the Unified Approach. It is an interesting example of a limit with a significant nuisance parameter.

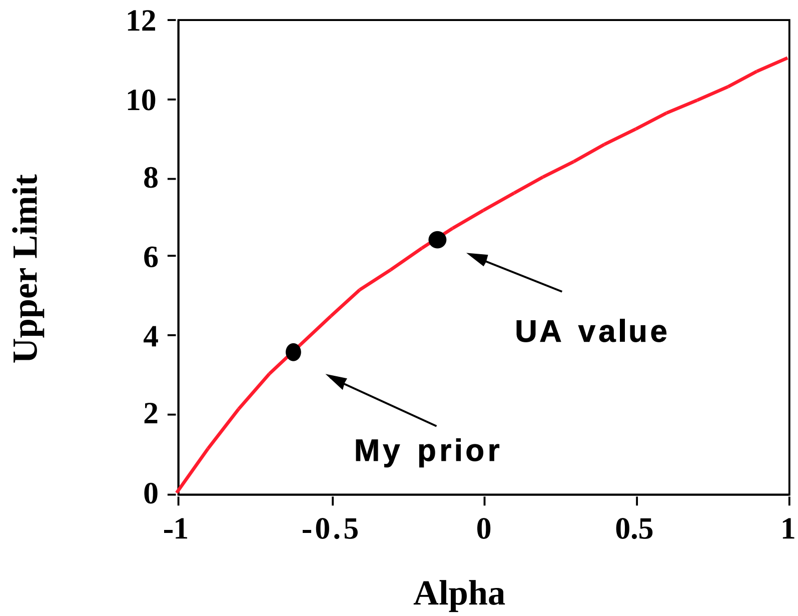
- The problem can be simplified to the observation of 13 signal plus background events with 8 background events having been measured in a control region of $1/r = 0.32$. (The two different background measurements combine in the likelihood function.) Thus, in this experiment the background is less well known than the measurement of signal plus background.



A Bayesian Analysis of CDMS

- In my simplified analysis, the Unified Approach gives a 90% c.l. upper limit of 6.3 signal events.

- I thought it would be interesting to see what a Bayesian analysis of this experiment would give. One has to choose a prior for both the background and the signal. The background prior does not matter much so I set it equal to the signal prior. For signal priors, I tried μ^{α}, where μ is the unknown parameter that is linear in the number of events. Statisticians prefer α = -0.5 or α = -1 for this type of a problem. The results:



A Bayesian Analysis of CDMS (cont)

- I also decided to do it right and use my subjective prior. I took 50% of the probability to be a δ-function at $\mu = 0$ and the rest flat in μ to 2 events and $1/\mu$ after that. The result was an upper limit of 3.6 events.

- I think this pretty accurately represents my degree of belief at the 90% c.l. It is lower, perhaps, than your degree of belief, but that is because I do not know much about WIMPs and am somewhat skeptical of them.

- This is the right way to use Bayesian statistics, but of course, it is not publishable.

Conclusion

- The Unified Approach can easily handle complicated problems involving the combination of results and nuisance parameters, yielding powerful frequentist results.