Confidence Interval Estimation

JC Talk

Bayes and Frequentism: Old Controversy (AS THEY SAY)

Prob(parameter given data)

Neyman

Prob(data, given parameter)

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

	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

$$P(X) = N(X)/N \quad \text{for} \quad N \to \infty$$

The frequentist definition of probability

Examples: coins, dice, cards

- For continuous x extend to Probability Density

$$P(x \,\text{to}\, x + \mathrm{d}x) = p(x)\mathrm{d}x$$

$p(x)$ is the *probability density function (pdf)*

- Examples:
  - Measuring continuous quantities ($p(x)$ often Gaussian, Poisson, ...)
  - Counting rates
  - Physical Quantities: Parton momentum fractions (proton pdfs) ...
- Alternative: Define Probability $P(X)$ as "degree of belief that X is true"

The Bayesian definition of probability

Likelihood

- Probability distribution of random variable x often depends on some parameter a
- Joint function $p(x, a)$:
  - Considered as $p(x)|a$ this is the pdf.
  - Normalised: $\int p(x)\mathrm{d}x = 1$
  - Considered as $p(a)|x$ this is the Likelihood $L(a)$ (or $\mathcal{L}(a)$)
  - Not "likelihood of a" but "likelihood that a would give x"
  - Not normalised. Indeed, must never be integrated.
- This is going to be one of the central concepts/quantities for the rest of the talk
- If we want to know a parameter a, we are looking for the point where the likelihood that a would predict the data x is maximized

Point estimation

- If we want to test a Hypothesis H_0 against another one (H_1), we want to compare their likelihoods

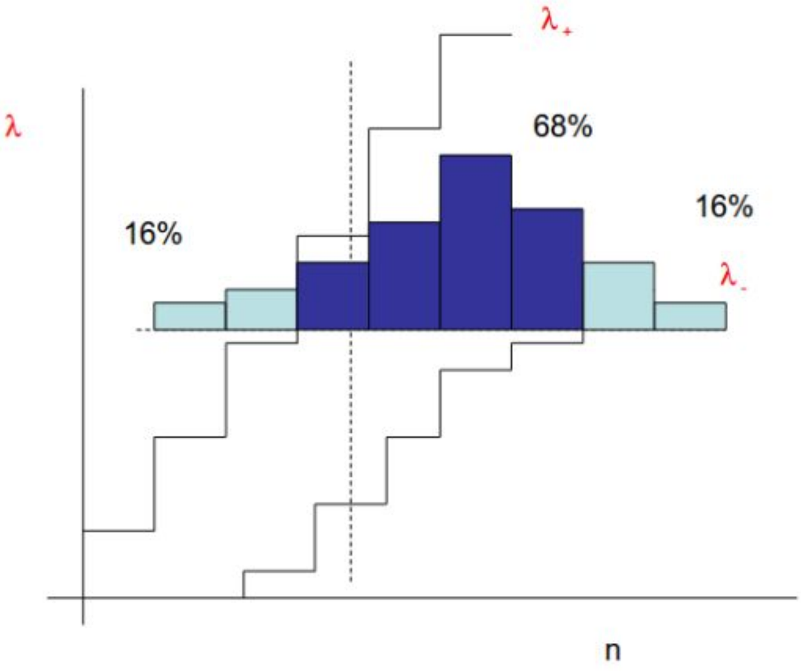
Hypothesis testing

- If we want to know what a cannot be, we want to know where $\mathcal{L}(a)|x$ is small

Interval estimation

Classical (Neyman's) Confidence Intervals

- Let's neglect systematics for the time being . . .
- Use Poisson-Distribution $p(n;\lambda) = e^{-\lambda}\lambda^n/n!$
- For any true λ the probability that $(n|\lambda)$ is within the belt is 68% (or more) by construction
- For any n, $[\lambda_-, \lambda_+]$ covers the true λ at 68% confidence
- Only integrated over n, not over λ!



Technique technically works for every CL, and single or double sided

Profile Likelihood method

Based on Neyman-Pearson Lemma

- Frequentist results are shown with the best-fit parameters and their errors using the profile likelihood technique. The profile likelihood function is defined as

$$\tilde{\mathcal{L}}^{\texttt{profile}}(\vec{\theta}) = \max_{\vec{\eta}} \mathcal{L}(\vec{\theta}, \vec{\eta}) \cdot \Pi(\vec{\theta}, \vec{\eta})$$

- Test statistic is defined as

$$TS = -log \left(\frac{\tilde{\mathcal{L}}^{profile}(\vec{\theta})}{\tilde{\mathcal{L}}^{profile}(\hat{\vec{\theta}})} \right)$$

- Errors are presented using Wilks' Theorem*

*S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," Ann. Math. Statist. 9, 60–62 (1938)

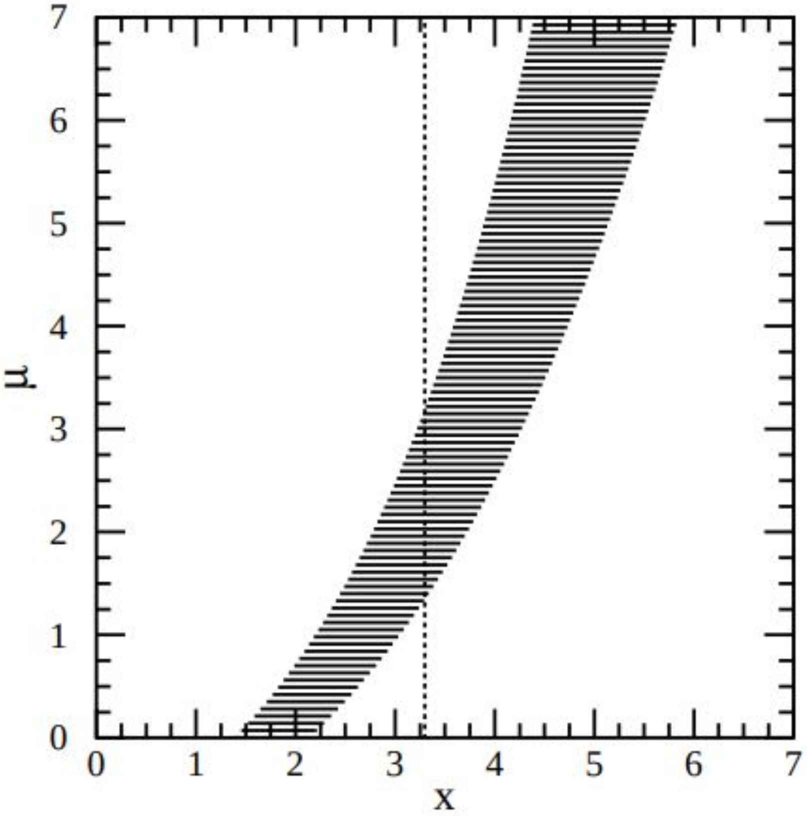
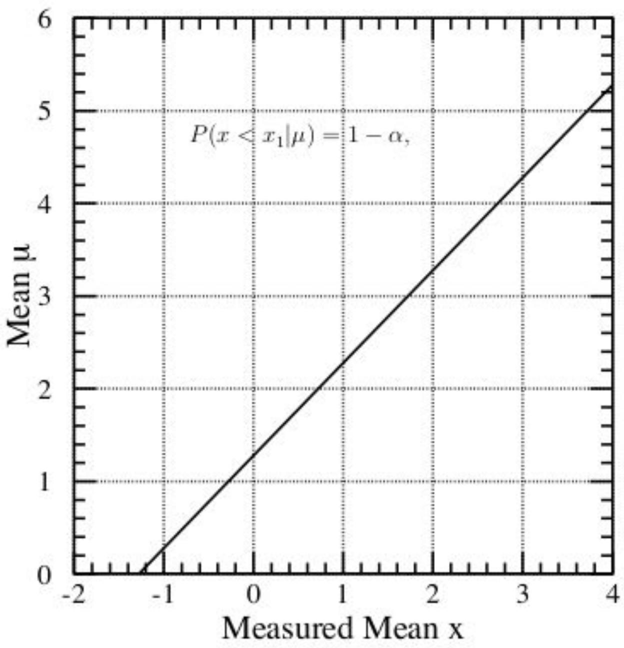


FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] \,|\mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through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

Gaussian with Boundary at origin

$$P(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp(-(x-\mu)^2/2).$$



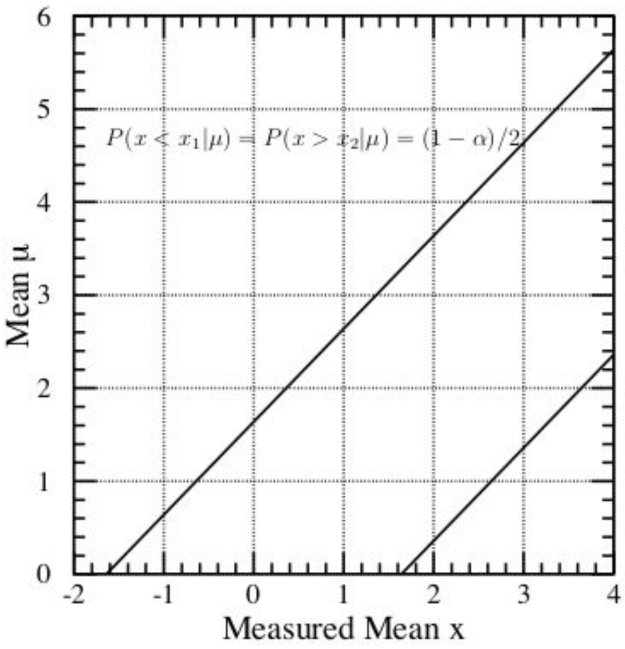


FIG. 2. Standard confidence belt for 90% C.L. upper limits for the mean of a Gaussian, in units of the rms deviation. The second line in the belt is at $x = +\infty$.

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

When should you give a central interval and when an upper limit?

Let us suppose, for example, that Physicist X takes the following attitude in an experiment designed to measure a small quantity: "If the result x is less then 3σ, I will state an upper limit from the standard tables. If the result is greater than 3σ, I will state a central confidence interval from the standard tables." We call this policy "flip-flopping" based on the data. Furthermore, Physicist X may say, "If my measured value of a physically positive quantity is negative, I will pretend that I measured zero when quoting a confidence interval", which introduces some conservatism.

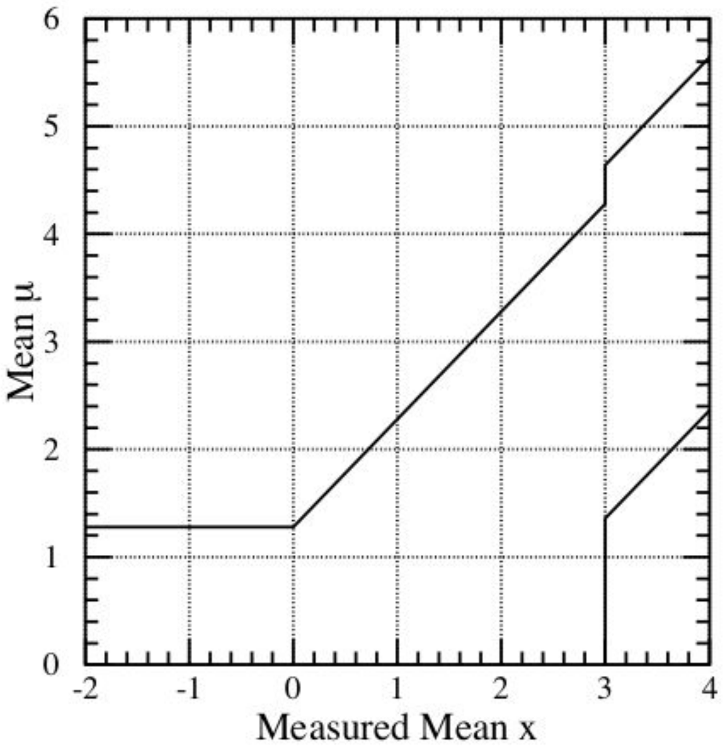


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping Physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence level. For $1.36 < \mu < 4.28$, the coverage (probability contained in the horizontal acceptance interval) is 85%.

We need an ordering principle

Feldman & Cousins ordering principle :

$$R = P(n|\mu)/P(n|\mu_{best})$$

Poisson with Background

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		√	

$$P(n|\mu) = (\mu + b)^n \exp(-(\mu + b))/n!$$

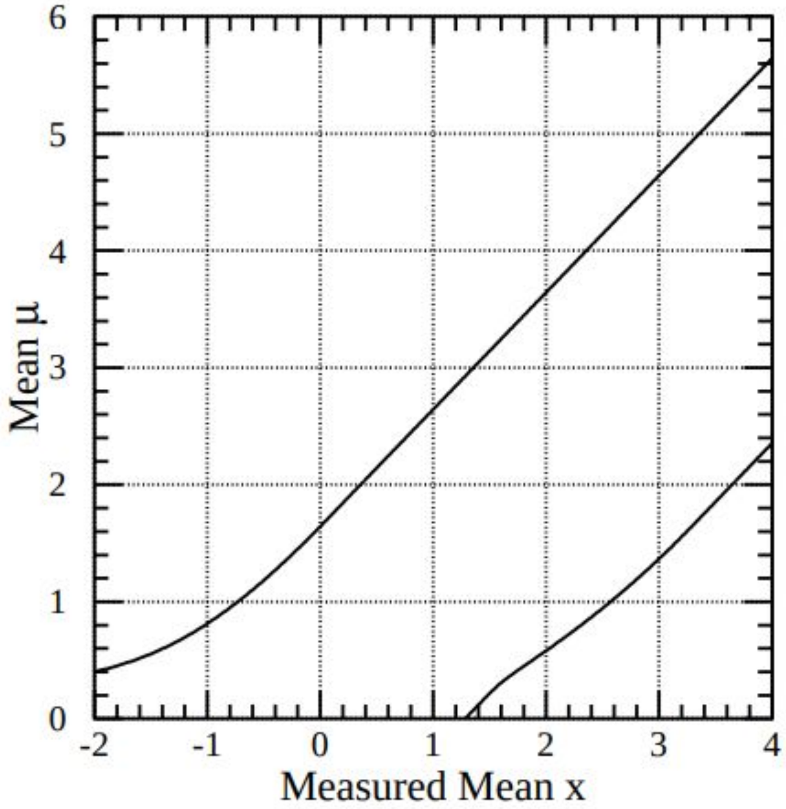


FIG. 10. Plot of our 90% confidence intervals for mean of a Gaussian, constrained to be non-negative, described in the text.

Summary

A Neyman construction is the most technically straightforward frequentist way to provide a confidence interval.

Profile likelihoods are the currently best accepted frequentist techniques for handling nuisance parameter uncertainties.

However it requires an ordering principle to ensure perfect coverage for small signals(FC Confidence intervals)

Comparison of Frequentist and Bayesian statistics

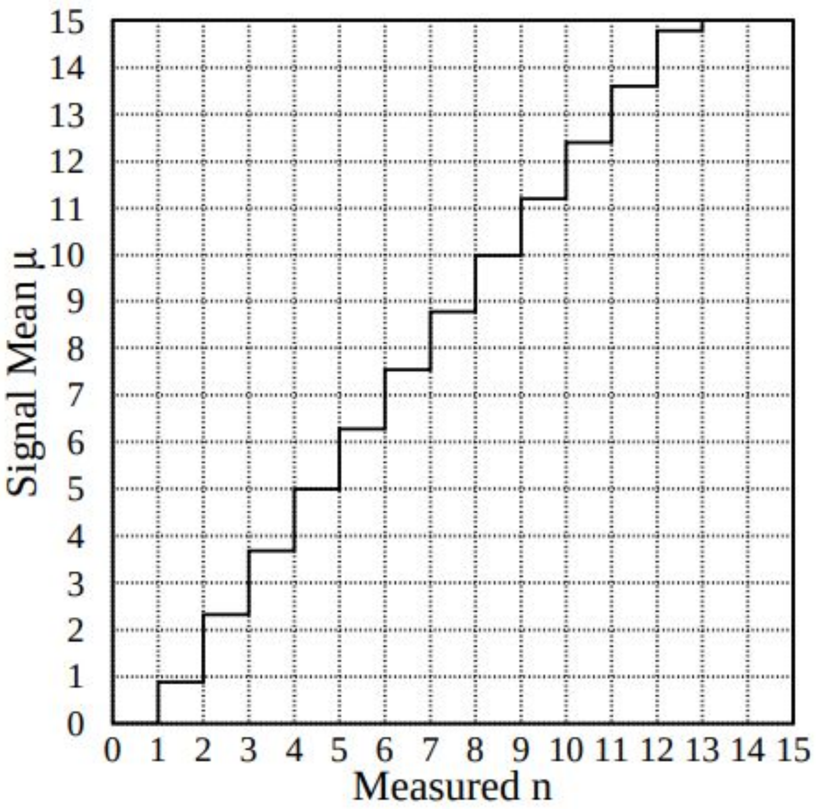


FIG. 5. Standard confidence belt for 90% C.L. upper limits, for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$. The second line in the belt is at $n = +\infty$.

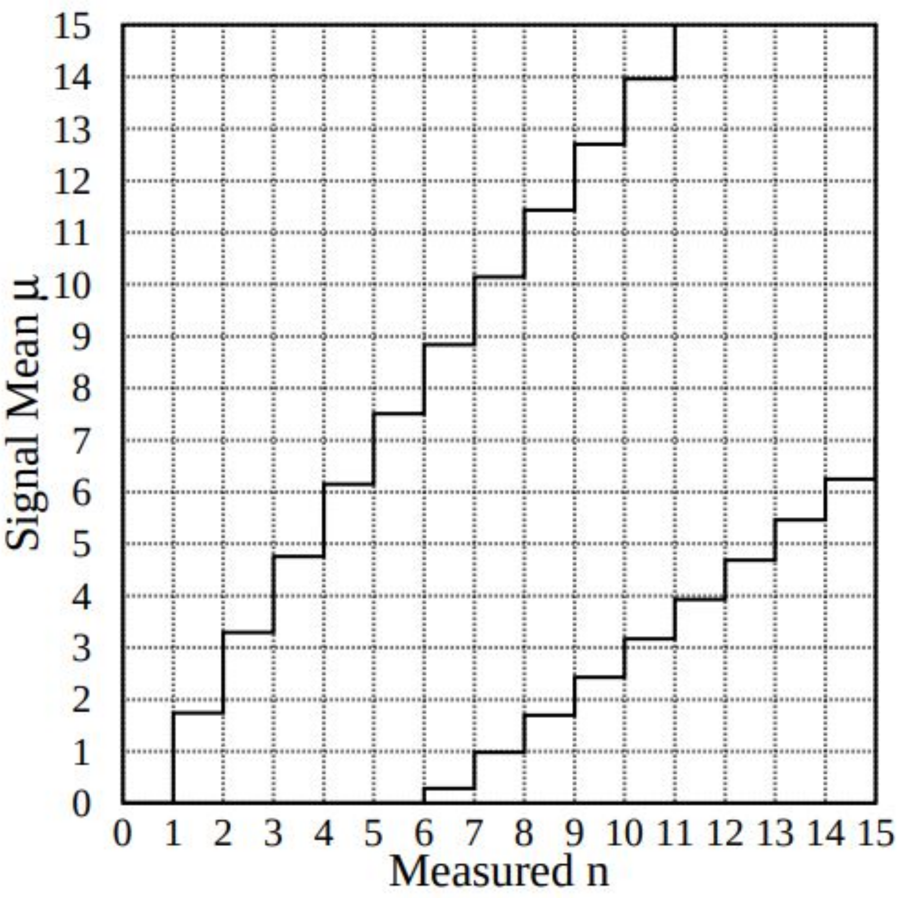


FIG. 6. Standard confidence belt for 90% C.L. central confidence intervals, for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$.

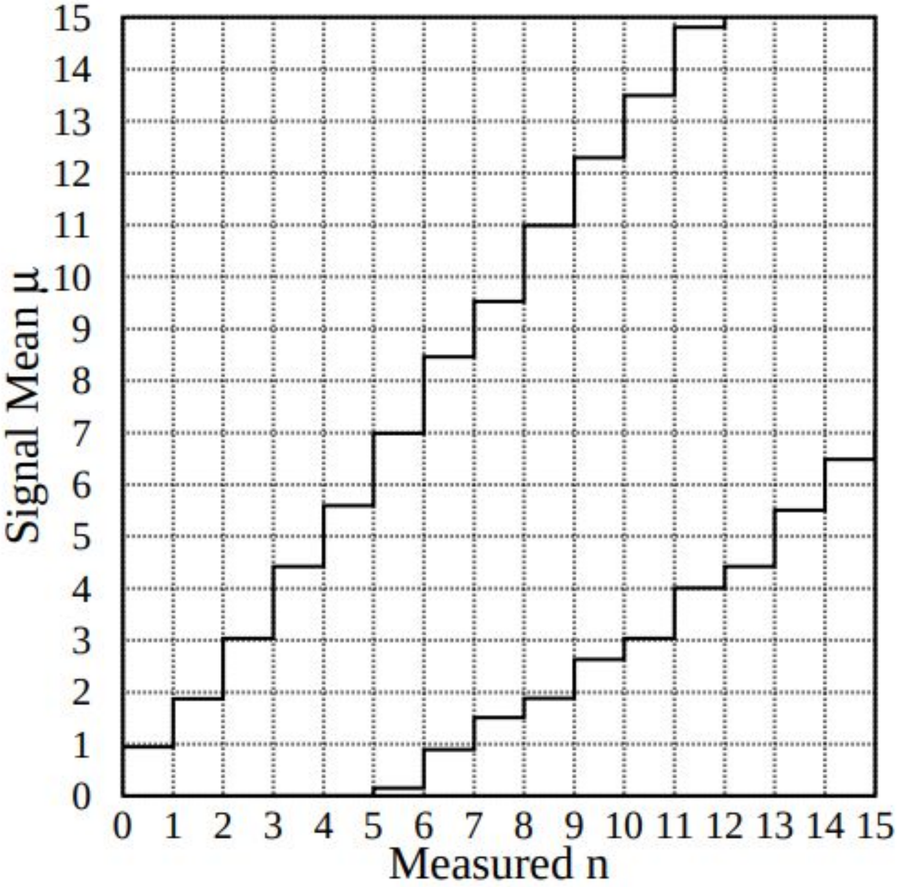


FIG. 7. Confidence belt based on our ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean μ in the presence of Poisson background with known mean $b = 3.0$.

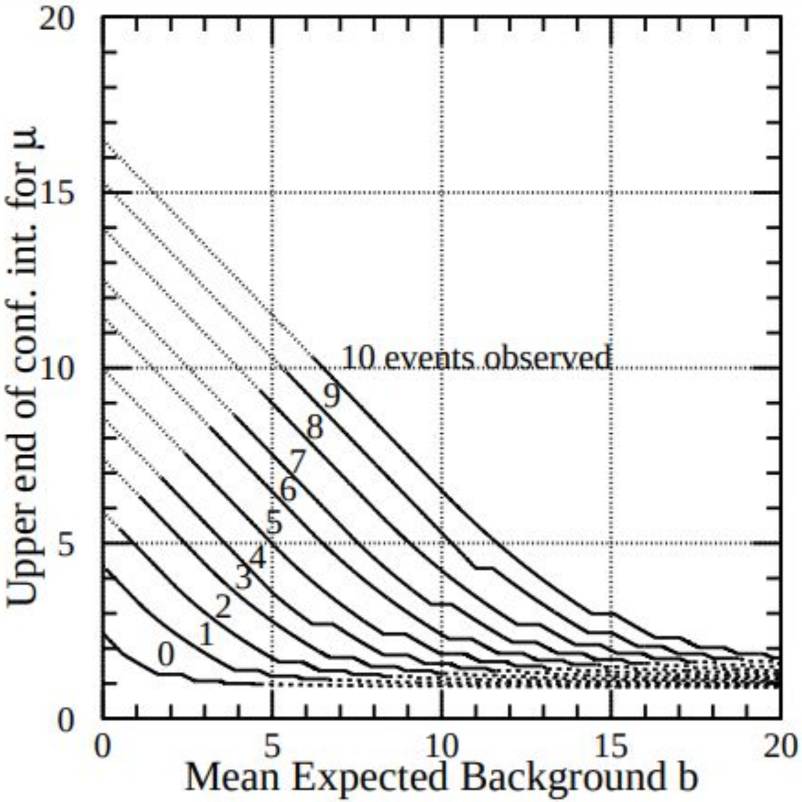


FIG. 8. Upper end μ_2 of our 90% C.L. confidence intervals $[\mu_1, \mu_2]$, for unknown Poisson signal mean μ in the presence of expected Poisson background with known mean b. The curves for the cases n_0 from 0 through 10 are plotted. Dotted portions on the upper left indicate regions where μ_1 is non-zero (and shown in the following figure). Dashed portions in the lower right indicate regions where the probability of obtaining the number of events observed or fewer is less than 1%, even if $\mu = 0$.

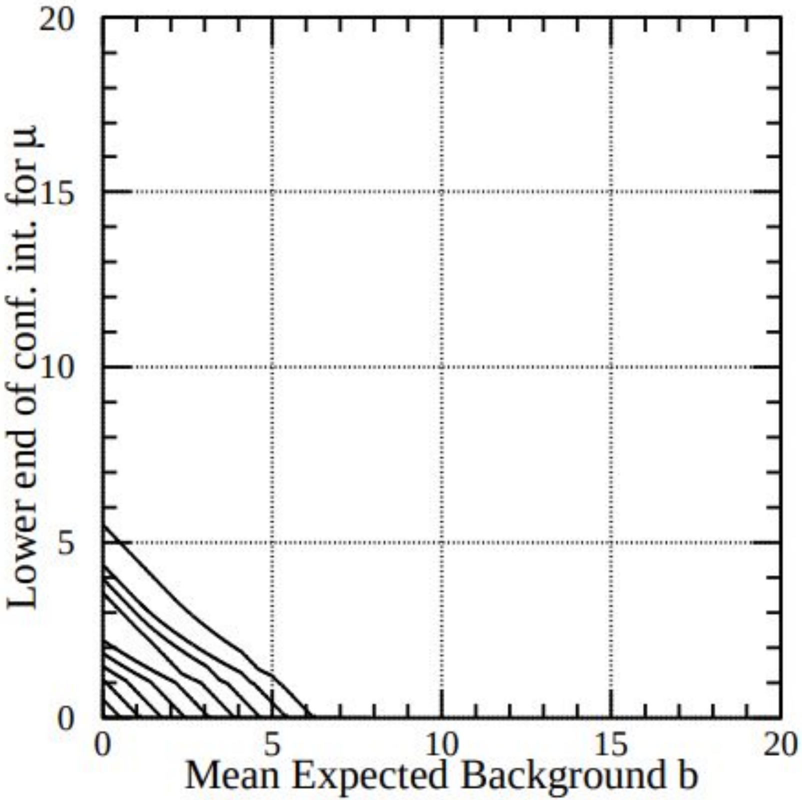


FIG. 9. Lower end μ_1 of our 90% C.L. confidence intervals $[\mu_1, \mu_2]$, for unknown Poisson signal mean μ in the presence of expected Poisson background with known mean b. The curves correspond to the dotted regions in the plots of μ_2 of the previous figure, with again $n_0 = 10$ for the upper right curve, etc.

NEW INTERVALS FROM AN ORDERING PRINCIPLE BASED ON LIKELIHOOD RATIOS

- Poisson with Background
- Gaussian with Boundary at Origin

Application to Neutrino Oscillation searches

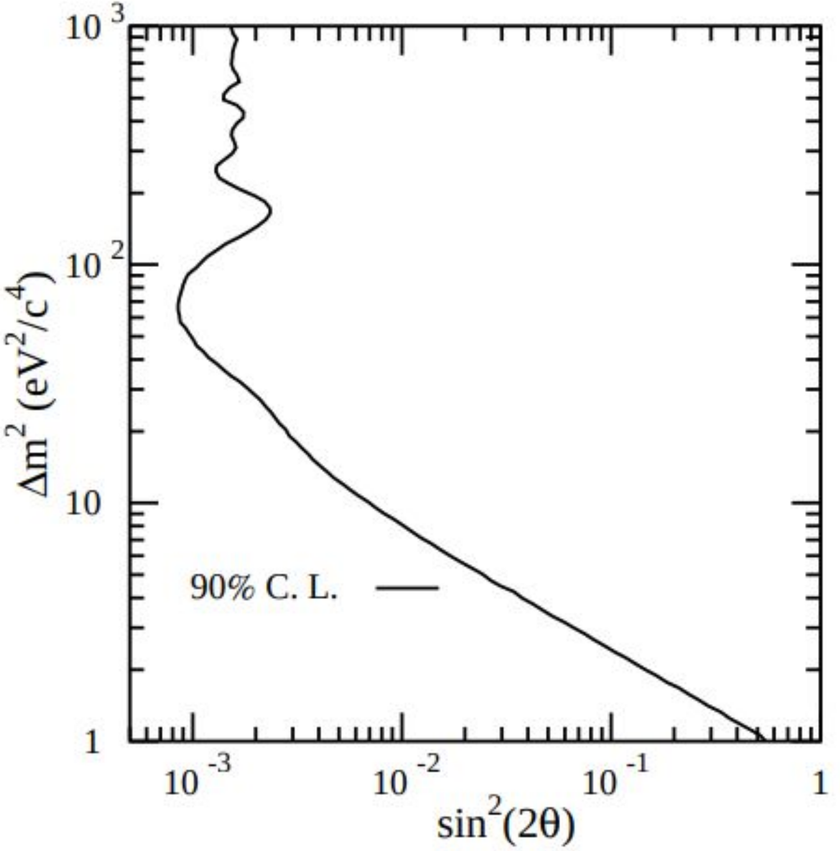


FIG. 11. Calculation of the confidence region for an example of the toy model in which $\sin^2(2\theta) = 0$. The 90% confidence region is the area to the left of the curve.

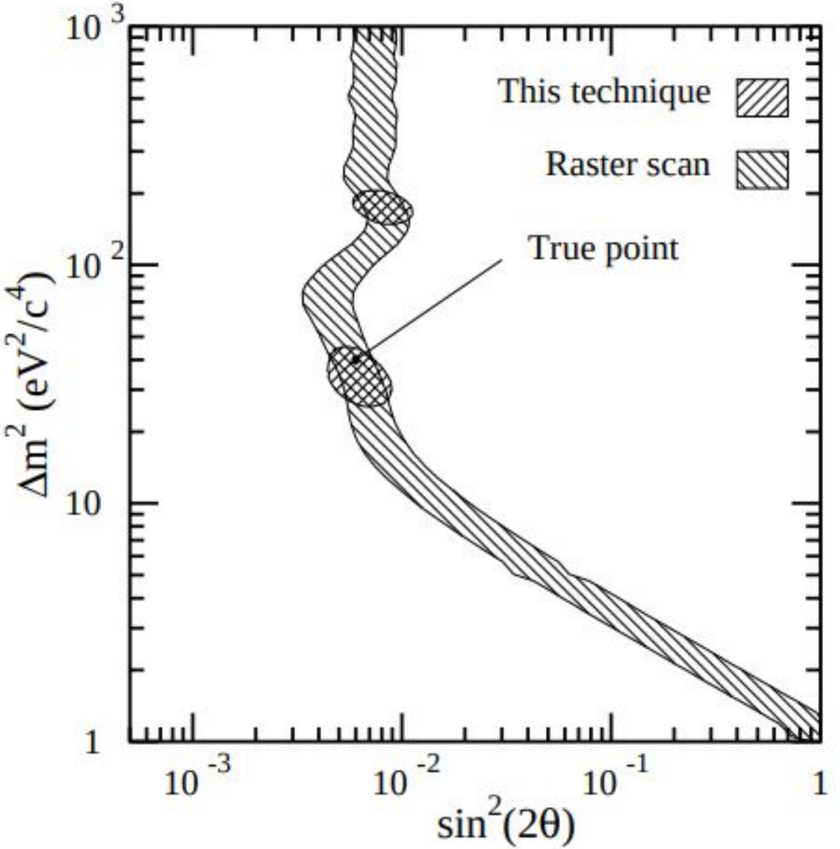


FIG. 12. Calculation of the confidence regions for an example of the toy model in which $\Delta m^2 = 40\ (\mathrm{eV}/c^2)^2$ and $\sin^2(2\theta) = 0.006$, as evaluated by the proposed technique and the Raster Scan.

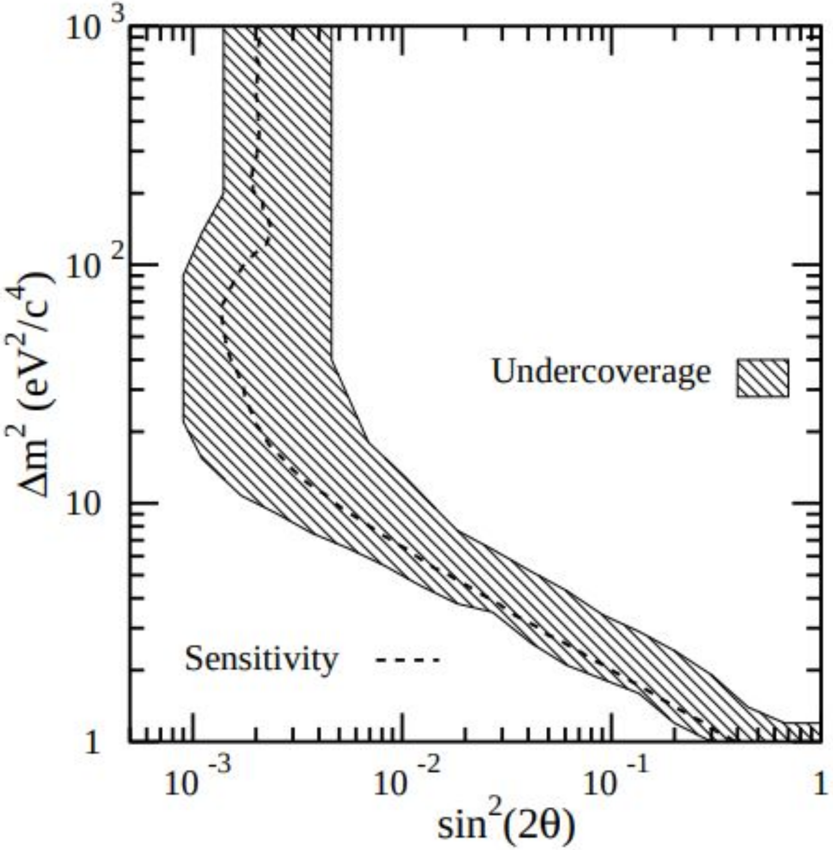


FIG. 13. Region of significant undercoverage for the Flip-Flop Raster Scan.

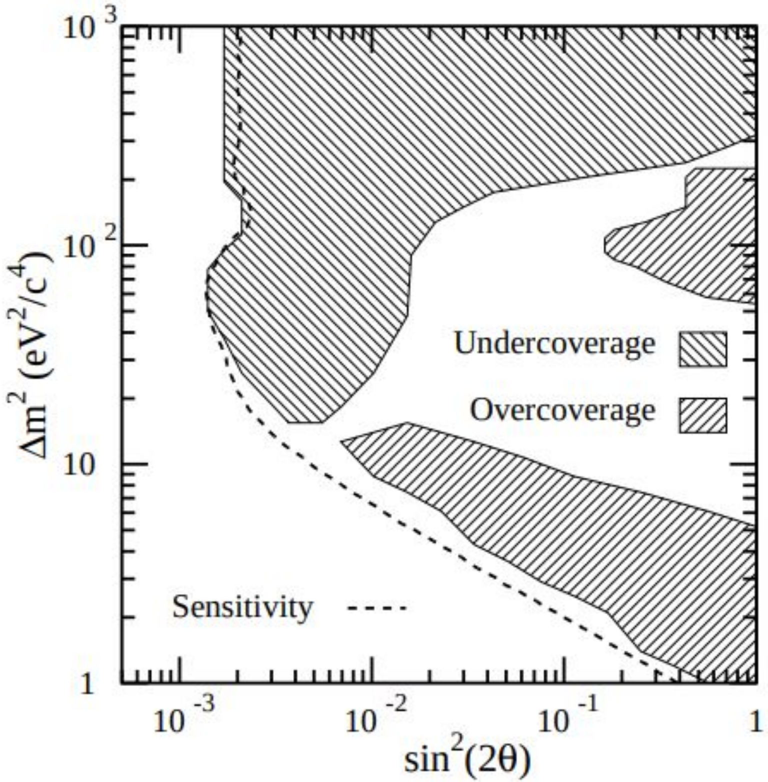


FIG. 14. Regions of significant under- and overcoverage for the Global Scan.

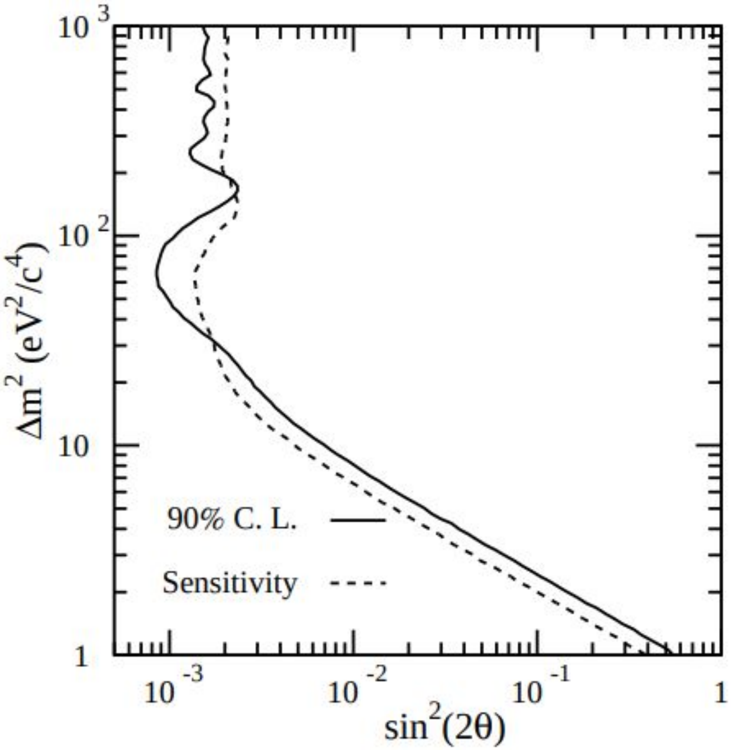
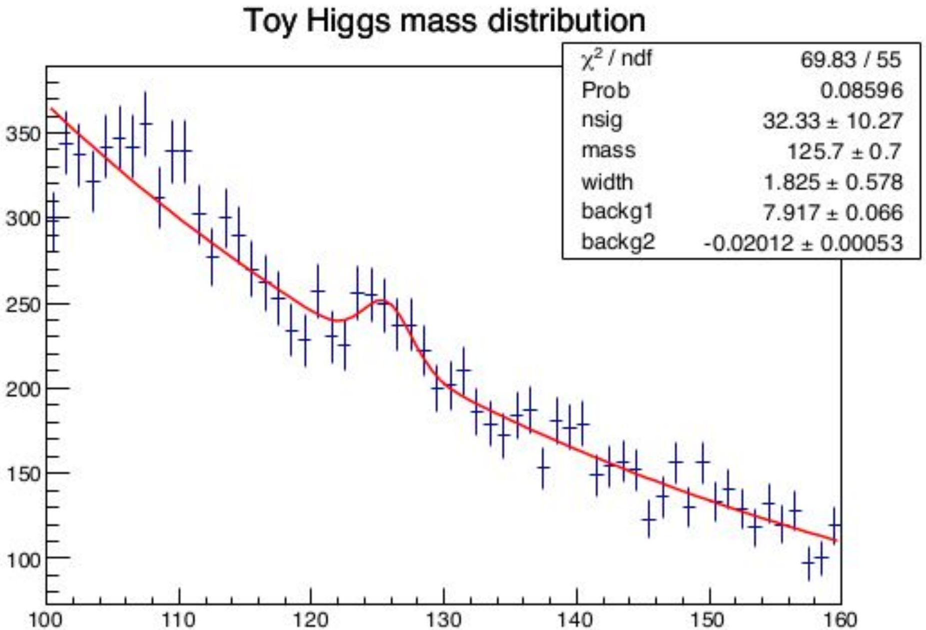


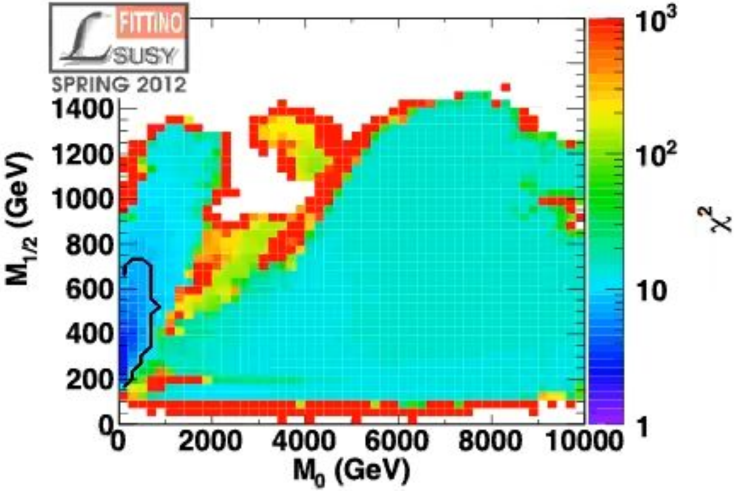
FIG. 15. Comparision of the confidence region for an example of the toy model in which $\sin^2(2\theta) = 0$ and the sensitivity of the experiment, as defined in the text.

The Profile Likelihood Technique in a fit

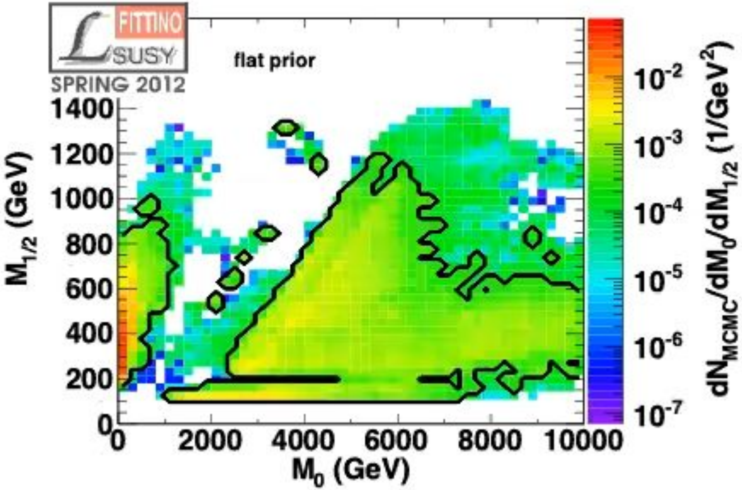
- In a fit to measurements $\vec{x}$, you vary the parameters $\vec{a}$ and either maximize the Likelihood $\ln \mathcal{L}(\vec{x};\vec{a})$ (or minimize the χ^2)
- In special cases: (and no correlations)

$$-2\ln\mathcal{L} = \chi^2 = \sum_i \frac{(x_i - \bar{x}_i(\vec{a}))^2}{\sigma_i^2}$$





Frequentist Profile Likelihood



Bayesian, Flat prior

- Quantify the agreement between each model point and the data:

$$\chi^2 = \sum_{i=1}^{n_{Obs}} \frac{(M_i - O_i(\vec{P}))^2}{\sigma_i^2} + Constraints$$

- Advanced MCMC scans with automatically adapting proposal density width