Statistics of Small Signals

- A partial discussion of a paper "Unified Approach to the Classical Statistical Analysis of Small Signals," which I wrote with Bob Cousins. [Phys. Rev. D 57, 3873 (1988)]

A Simple Example (1)

- Suppose you are searching for a rare process and have a well-known expected background of 3 events, and you observe 0 events. What 90% confidence limit can you set on the unknown rate for this rare process?
- A classical (or frequentist) statistician makes a statement about the probability of data given theory. That is, given a hypothesis for the value of an unknown true value μ, he or she will give you the probability of obtaining a set of data x, $P(x \mid \mu)$.
- A classical confidence interval (Jerzy Neyman, 1937) is a statement of the form: The unknown true value of μ lies in the region $[\mu_1,\mu_2]$. If this statement is made at the 90% confidence level, then it will be true 90% of the time, and false 10% of the time.

A Simple Example (2)

- Poisson statistics $P(x = 0 \mid \mu = 2.3) = 0.1$. Therefore, in the "standard" classical approach, $\mu < 2.3$ at 90% C.L. Since $\mu = s + b$, and $b = 3.0$, $s < -0.7$ at 90% C.L.
- Thus, we are led to a statement that we know is a priori false.

Bayesian Statistics

- A Bayesian takes the opposite position from a classical statistician. He or she calculates the probability of theory given data. That is, given a set of data x, he or she will calculate the probability that the unknown true value is μ, $P(\mu \mid x)$.
- This appears attractive because it is what you really want to know. However, it comes at a price:

Bayes's Theorem

- $P(x \mid \mu)$ and $P(\mu \mid x)$ are related by Bayes's Theorem, which in set theory is the statement that an element is in both A and B is

  $$P(A \mid B)\, P(B) = P(B \mid A)\, P(A)$$

  which for probabilities becomes

  $$P(\mu \mid x) = P(x \mid \mu)\, P(\mu)/P(x).$$

  $P(x)$ is just a normalization term, but Bayes's Theorem transforms $P(\mu)$, the prior distribution of "degree of belief" in μ, to $P(\mu \mid x)$, the posterior distribution.

  A "credible interval" or "Bayesian confidence interval" is formed by

  $$\int_{\mu_1}^{\mu_2} P(\mu \mid x)\, d\mu = 90\%.$$

An Example (1)

- Suppose you have a large number of marbles, which are either white or black, and you wish information on the fraction that are white, μ. You draw a single marble, and it is white. What can you say at 90% confidence?

	Prior	
Classical:		$\mu \geq 0.1$
Bayesian:	flat	$\mu \geq 0.316$
	μ	$\mu \geq 0.464$
	$1/\mu$	$\mu \geq 0.1$
	$(1-\mu)$	$\mu \geq 0.196$
	$1/(1-\mu)$	unnormalizable

An Example (2)

- Notice that most of the Bayesian priors do not cover, i.e., they are not true statements the stated fraction of the time (90% in this case). There is no requirement that credible intervals cover. However, Bob Cousins warns [Am. J. Phys. 63, 398 (1995)],

  "…if a Bayesian method is known to yield intervals with frequentist coverage appreciably less than the stated C.L. for some possible value of the unknown parameters, then it seems to have no chance of gaining consensus acceptance in particle physics."

The Role of Bayesian Statistics (1)

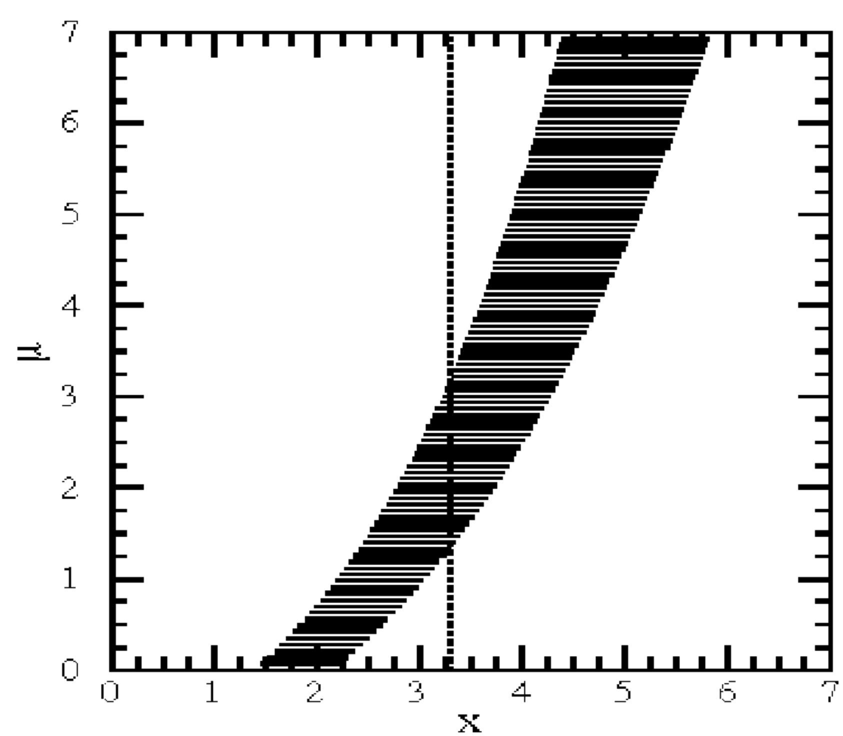
- Harrison Prosper [Phys. Rev. D 37, 1153 (1988)] argues for a $1/\mu$ prior based on a scaling argument. I found it unsatisfactory for two reasons:
  - It fails for $x = 0$. (Unnormalizable)
  - In general, it undercovers.
- To quote Bob's prose from our paper:
  - "In our view, the attempt to find a non-informative prior within Bayesian inference is misguided. The real power of Bayesian inference lies in its ability to incorporate 'informative' prior information, not 'ignorance.'"

The Role of Bayesian Statistics (2)

- Prosper wrote that he was using a Bayesian approach because
  - "…we are merely acknowledging the fact that a coherent solution to the small-signal problem is more easily achieved within a Bayesian framework than one which uses the methods of 'classical' statistics."
- Through this talk, I hope to convince you that this is no longer true.

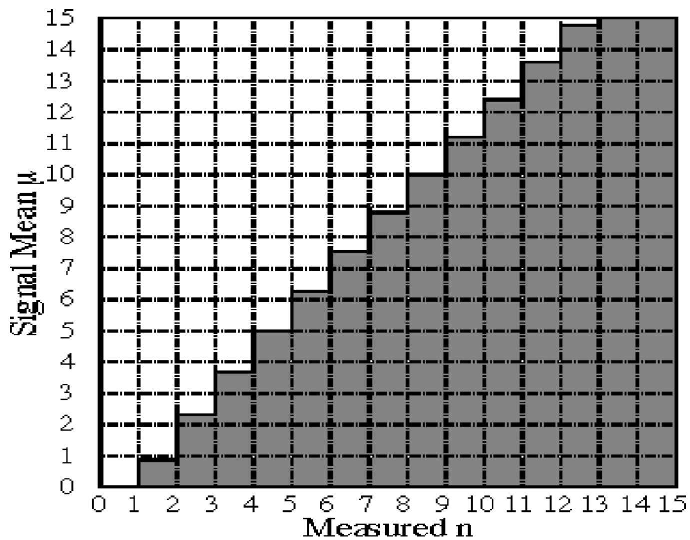
Construction of Confidence Intervals

- Neyman's prescription: **Before doing an experiment,** for each possible value of theory parameters determine a region of data that occurs C.L. of the time, say 90%. **After doing the experiment,** find all of values of the theory parameters for which your data is in their 90% region. This is the confidence interval.
- **Notice that there is complete freedom of choice of which 90% to choose.** This will be the key to ou[r] solution.

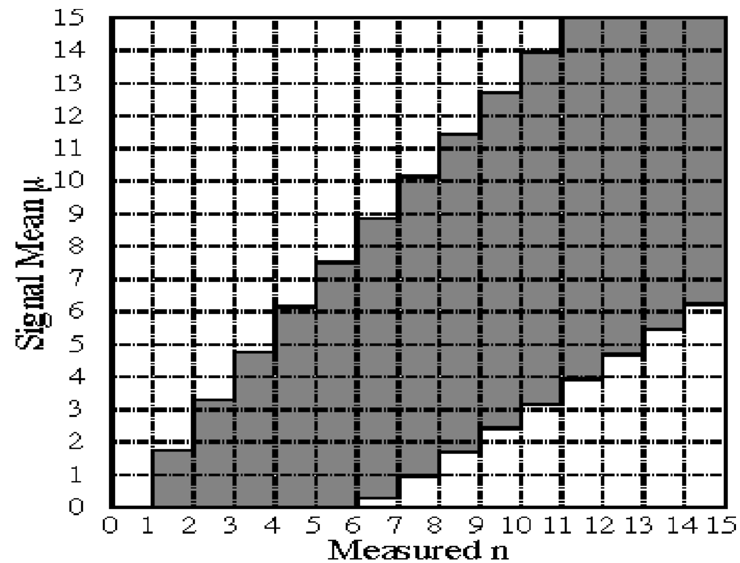


Examples of Poisson Confidence Belts

- For our example: 90% C.L. limits for Poisson μ with background = 3



Upper limits



Central limits

The Solution

- **For both the upper limit and central limit, $x = 0$ excludes the whole plane. But consider the problem from the point of view of the data. If one measures no events, then clearly the most likely value of μ is zero. Why should one rule out the most likely scenario?**
- **Therefore, we proposed a new ordering principle based on the ratio of a given μ to the most likely μ:**

$$R = \frac{P(x \mid \mu)}{P(x \mid \mu^*)}$$

where μ* is the most likely value of μ given x.

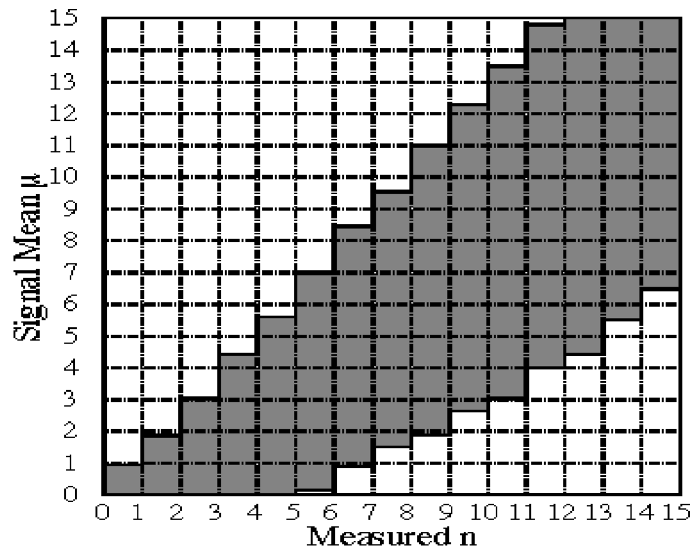
An Example (1)

- Example for $\mu = 0.5$ and $b = 3$:

x	$P(x\|\mu)$	μ^*	$P(x\|\mu^*)$	R	rank	U.L.	C.L.
0	0.030	0.0	0.050	0.607	6		
1	0.106	0.0	0.149	0.708	5	x	x
2	0.185	0.0	0.224	0.826	3	x	x
3	0.216	0.0	0.224	0.963	2	x	x
4	0.189	1.0	0.195	0.966	1	x	x
5	0.132	2.0	0.175	0.753	4	x	x
6	0.077	3.0	0.161	0.480	7	x	x
7	0.039	4.0	0.149	0.259		x	x
8	0.017	5.0	0.140	0.121		x	

Unified Poisson Limits

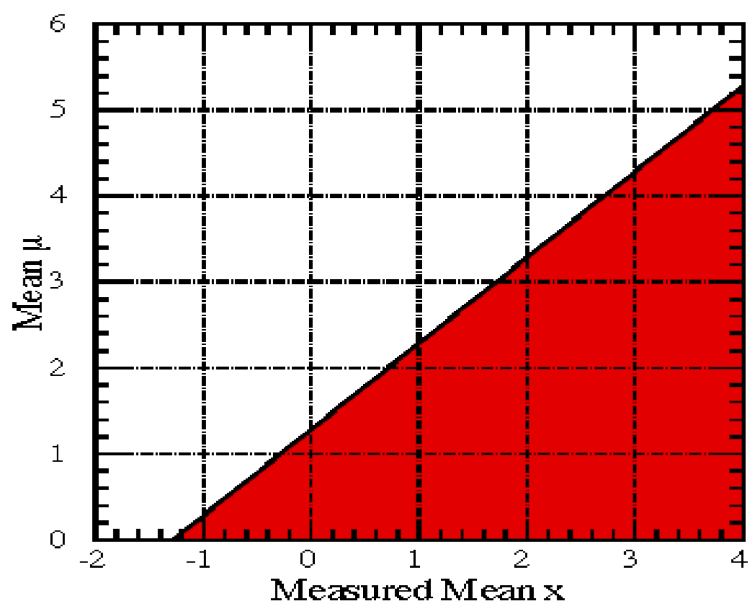
- 90% C.L. unified limits for Poisson μ with background = 3



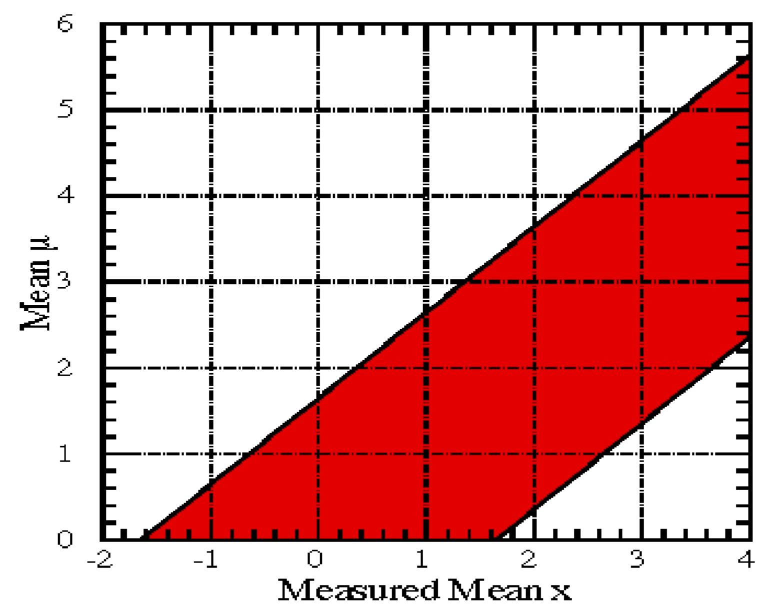
- **Solution to our original problem: μ < 1.08 at 90% C.L.**

Examples of Gaussian Confidence Belts

- 90% C.L. limits for Gaussian $\mu \geq 0$ vs. x (total – background) in □



Upper Limits



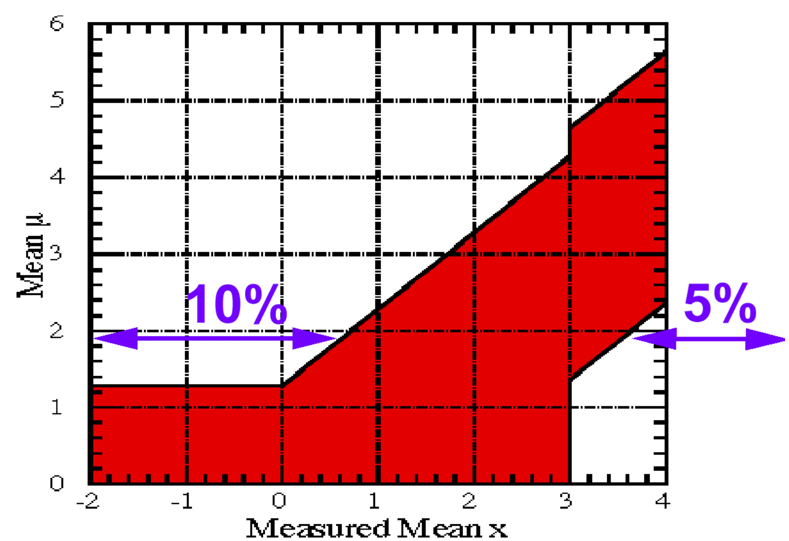
Central Limits

Flip-Flopping (1)

- How does a typical physicist use these plots?
  - "If the result $x < 3\sigma$, I will quote an upper limit."
  - "If the result $x > 3\sigma$, I will quote a central confidence interval."
  - "If the result $x < 0$, I will pretend I measured zero."

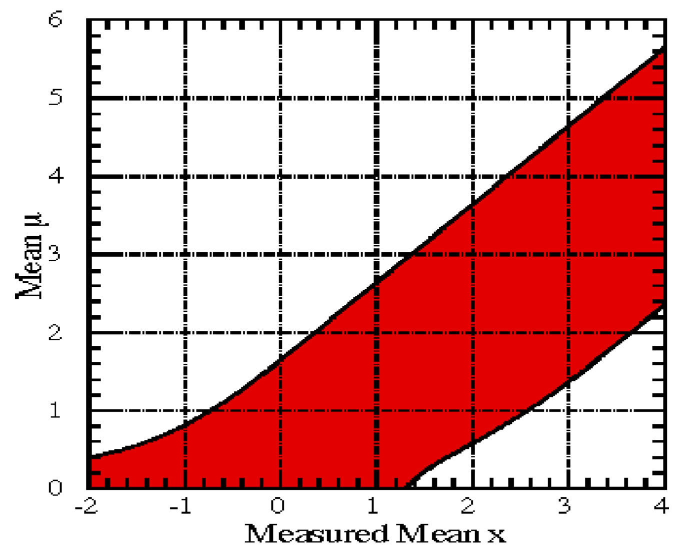
Flip-Flopping (2)

- This results in the following:



- In the range $1.36 \le \mu \le 4.28$, there is only 85% coverage!
- Due to flip-flopping (deciding whether to use an upper limit or a central confidence region based on the data) these are not valid confidence intervals.

Unified Solution for the Gaussian Case (1)



- **Notes:**
  - **This approaches the central limits for x >>1**
  - **The upper limit for x = 0 is 1.64, the two-sided rather than the one-sided limit.**

Unified Solution for the Gaussian Case (2)

- **Notes (continued):**
  - **From the defining 1937 paper of Neyman, this is the only valid confidence belt, since there are 4 requirements for a valid belt:**

    (1) It must cover.

    (2) For every x, there must be at least one μ.

    (3) No holes (only valid for single μ).

    (4) Every limit must include its end points.

Sensitivity

- **The main objection to this work has been that an experiment that observes fewer events than the expected background may report a lower upper limit than a (better designed ?) experiment that has no background.**
- **To address this problem and to provide additional information for the reader's assessment of the significance of the results, we suggested that experiments that have fewer counts than expected background also report their sensitivity, which we defined as the average* upper limit that would be obtained by an ensemble of experiments with the expected background and no true signal.** *Should be median
- **We did this in the NOMAD experiment and other experiments have been doing the same thing.**

Visit to Harvard Statisticians (1)

- Towards the end of this work, I decided to try it out on some professional statisticians whom I know at Harvard.
- They told me that this was the standard method of constructing a confidence interval!
- I asked them if they could point to a single reference of anyone using this method before.
- They could not.

Visit to Harvard Statisticians (2)

- **Their logic:**
  - **In statistical theory there is a one-to-one correspondence between a hypothesis test and a confidence interval. (The confidence interval is a hypothesis test for each value in the interval.)**
  - **The Neyman-Pearson Theorem states that the likelihood ratio gives the most powerful hypothesis test.**
  - **Therefore, it must be the standard method of constructing a confidence interval.**

Kendall and Stuart (1961)

- **So I started reading about hypothesis testing.**
- **At the start of chapter 24 of Kendall and Stuart's *The Advanced Theory of Statistics* (chapter 23 of Stuart and Ord), I found 1 1/4 cryptic pages that propose this method and its extension to errors on the background.**
- **We were able to include a reference to Kendall and Stuart in a note added in proof to our paper.**

Extensions

- This technique is more general than the simple examples described here.
- The paper discusses the application to neutrino oscillations, in which limits are set on two parameters, $\sin^2 2\theta$ and Δm^2, simultaneously.
- It can also be extended to cases in which the backgrounds are not precisely known (but we have not yet published this).
- In fact, I have yet to find a problem in the construction of classical confidence intervals and regions that is not solvable by the ordering principle suggested here.