Statistics

Lecture 3

August 8, 2000

Frank Porter

Caltech

The plan for these lectures:

- The Fundamentals; Point Estimation
- Maximum Likelihood, Least Squares and All That
- What is a Confidence Interval?
- Interval Estimation
- Monte Carlo Methods

Additional topics will be covered by Roger Barlow and Norman Graf

Interval estimation

In physics applications, we usually aren't satisfied with merely quoting a "best estimate" for the value of a parameter. We would also like to relay some idea of how precise the experiment is, or, alternatively, give some feeling for how close the point estimate may be expected to be to the true value of the unknown parameter.

This brings us to the emotionally charged topic of interval estimation.

We'll spend today trying to understand it, tomorrow looking at common techniques.

Confidence Intervals

Neyman's definition of Confidence Interval:

"If the functions θ_ℓ and θ_u possess the property that, whatever be the possible value ϑ_1 of θ_1 and whatever be the values of the unknown parameters $\theta_2, \theta_3, \ldots, \theta_s$, the probability

$$P\{\theta_\ell \le \vartheta_1 \le \theta_u | \vartheta_1, \theta_2, \ldots, \theta_s\} \equiv \alpha,$$

then we will say that the functions θ_ℓ and θ_u are the lower and upper confidence limits of θ_1, corresponding to the confidence coefficient α."

The interval $\left(\theta_\ell, \theta_u\right)$ is called the Confidence Interval for θ_1.

Note: Some people use $1 - \alpha$.

Example: Mean of a Normal Distribution

Suppose we sample a value x from an $N(\theta, 1)$ distribution.

We could form a 68% confidence interval for θ as follows: Throw a random number, r, uniform in $(0, 1)$. If $r < 0.68$, quote the interval $(-\infty, \infty)$. Otherwise, quote the null interval.

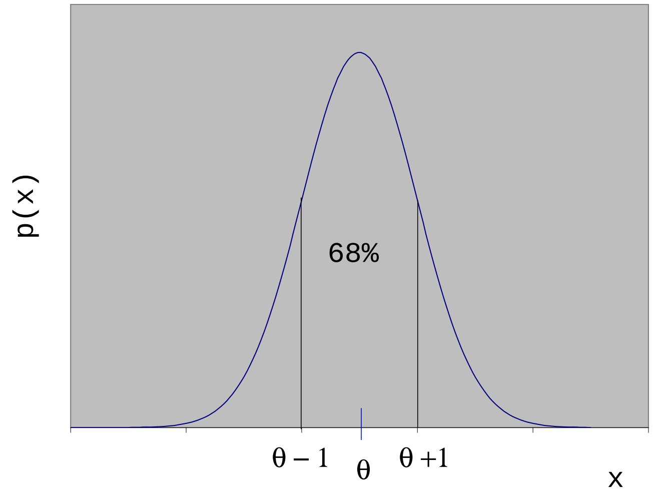
This is clearly a valid confidence interval, including the exact value of θ, independent of θ, with a frequency of precisely 68%.

However, it is a useless exercise – it has nothing to do with the measurement! To be useful, then, we must require more from our interval estimation; we should strive for sufficiency!

We could also ask for other properties, such as equal distances on each side of a point estimator.

Example Continued; Try Again

Notice that 68% of the time we make a sample x from an $N(\theta, 1)$ distribution, the value of x will be within 1 unit of θ.



Thus, if we quote the interval $(x-1, x+1)$, we will have a valid 68% confidence interval for θ – The quoted interval will include θ with a frequency of precisely 68%. Of course, for any given sample, the quoted interval either includes θ or it doesn't. We might even know that it doesn't, e.g., if the interval is outside some "physical" boundary on allowed values of θ. This is irrelevant!

Note: these statistics are sufficient.

The Mind Set Problem

Neyman goes on to say:

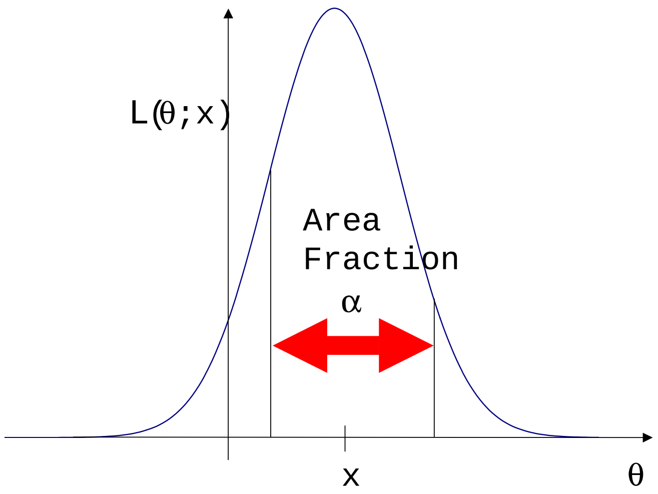
“In spite of the complete simplicity of the above definition, certain persons have difficulty in following it. These difficulties seem to be due to what Karl Pearson (1938) used to call routine of thought. In the present case the routine was established by a century and a half of continuous work with Bayes's theorem. . . ”

Physicists have the same difficulty, even without the 150 years of Bayes' theorem

Bayesian Intervals

The definition of a confidence interval may be contrasted with the "Bayesian interval":

Bayesian Interval: A Bayesian Interval, at the α confidence level, for a population parameter θ is an interval which contains a fraction α of the area under a Bayes' distribution.



Bayes' Distribution

Bayes' Distribution: A Bayes' Distribution is a function of θ of the form:

$$p(x;\theta)P(\theta)/\int_{-\infty}^{\infty} p(x;\theta)P(\theta)d\theta,$$

where $P(\theta)$ is a non-negative function called the prior distribution, and $p(x;\theta)$ is the pdf evaluated at the observed value x.

Note the appearance of Bayes' theorem in the Bayes' distribution. It's just being used in a new way…

Confidence Intervals – Comments

1. A confidence interval tells us about our data, e.g., a small interval indicates a more precise measurement than a wide interval.

2. A confidence interval tells us nothing about the true value of the parameter. It is up to the reader to make any inferences concerning the closeness of the unknown parameter to the quoted interval. But that is a Bayesian step, outside of frequentist statistics.

3. The utility of the confidence interval is in summarizing, objectively, the quality of the measurement. The true value of the parameter is irrelevant.

Bayesian Intervals – Comments

4. The motivation for the Bayesian interval is that we really are interested in making some statement about the value of the parameter.

5. The interpretation of the Bayes' distribution is that it expresses our "degree of belief" in where the parameter lies.

6. The Bayes' distribution is mathematically a probability distribution, but it does not have a frequency interpretation. We cannot sample from it.

7. The utility of the Bayes' distribution/interval is that it gives us a formalism with which to make decisions, based on available knowledge.

The Problem of the Prior Distribution

Suppose we are faced with the problem of forming a Bayes' distribution, after making some measurement x:

$$b(\theta; x) = p(x; \theta)P(\theta) / \int_{-\infty}^{\infty} p(x; \theta)P(\theta)d\theta.$$

What do we use for the "prior", $P(\theta)$?

If we already had a Bayes' distribution, prior to the present experiment, that becomes our new prior.

But suppose we don't; suppose we consider ourselves "completely ignorant" of the parameter value prior to the experiment? What is the distribution representing complete ignorance?

Expressing Ignorance

We might suppose that complete ignorance is simply expressed as a flat distribution, since we have no reason to prefer one value over another.

Two worries about this:

1. May not be normalizable if range of possible parameters is infinite. Not a problem in practice, just set $P(\theta) = 1$, since normalization divides out anyway.
2. More serious: Complete ignorance (flat distribution) in θ is not complete ignorance in all functions of θ, e.g., θ^2. Ignorance is parameterization-dependent!

Does it Matter?

There is a vast literature on dealing with the issue of the prior distribution.

My (philosophical!) perspective:

Go ahead and use a flat, or other smooth, prior for whatever parameter you care about. But if it makes much difference in any decisions you're going to make, Be Warned!

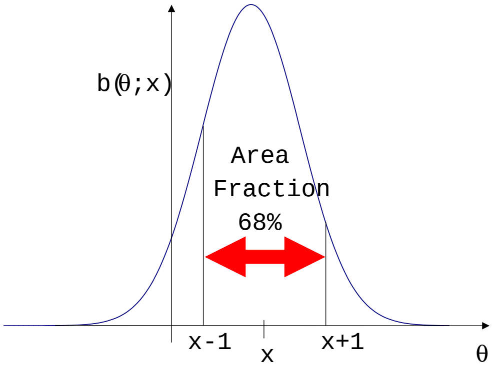
Bayes' Treatment of Normal Example

Having discussed the prior, we are ready to give the Bayesian treatment of our $N(\theta, 1)$ example.

With a flat prior, the Bayes' distribution is:

$$b(\theta; x) = \frac{1}{\sqrt{2\pi}} \exp\left[-(x-\theta)^2/2\right],$$

which is just the pdf for sampling $x(!)$, except now interpreted as a probability function for θ.



A 68% Bayes' interval for θ is a region containing 68% of the area under this distribution; e.g., for a central interval about x, it is just $(x-1, x+1)$.

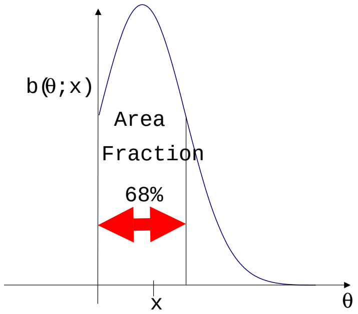
This is the same as our confidence interval! But the interpretation is much different...

Bayes' Treatment of Normal Example, Continued

Let us suppose now that θ is a particle mass, i.e., the physically allowed region is > 0. We know then that the probability that the parameter is < 0 must be zero, so our Bayes' distribution should contain this knowledge.

This constraint is readily incorporated in the prior distribution, setting $P(\theta < 0) = 0$. Thus, if our prior is otherwise flat, the Bayes' distribution becomes:

$$b(\theta; x) = \begin{cases} 0 & \theta < 0, \\ p(x;\theta)/\int_0^\infty p(x;\theta)dx & \theta \geq 0. \end{cases}$$



If $x >> 1$, this doesn't affect our quoted interval much, but if $x \lesssim 1$ it matters, and in particular we will never quote an interval that goes negative.

Meanwhile, Back to Confidence Intervals. . .

Let's try to achieve a more complete understanding by considering another example:

Suppose we sample a value n from a Poisson distribution:

$$p(n;\theta) = \frac{\theta^n e^{-\theta}}{n!}, \quad n = 0, 1, 2, \ldots$$

How do we obtain confidence intervals on θ for this distribution?

Poisson – Usual Approach to Upper Limits

Let's consider the "usual" method for obtaining upper limits with a Poisson sampling distribution:

The prescription, given an observation n, is to solve the following equation for θ_1:

$$1-\alpha = \sum_{k=0}^{n} p(k;\theta_1(n)).$$

That is, $\theta_1(n)$ as an upper limit on θ at the α confidence level is obtained as the value of the mean for which we would observe n or less counts a fraction $1-\alpha$ of the time.

Poisson – Usual Upper Limits, Comments

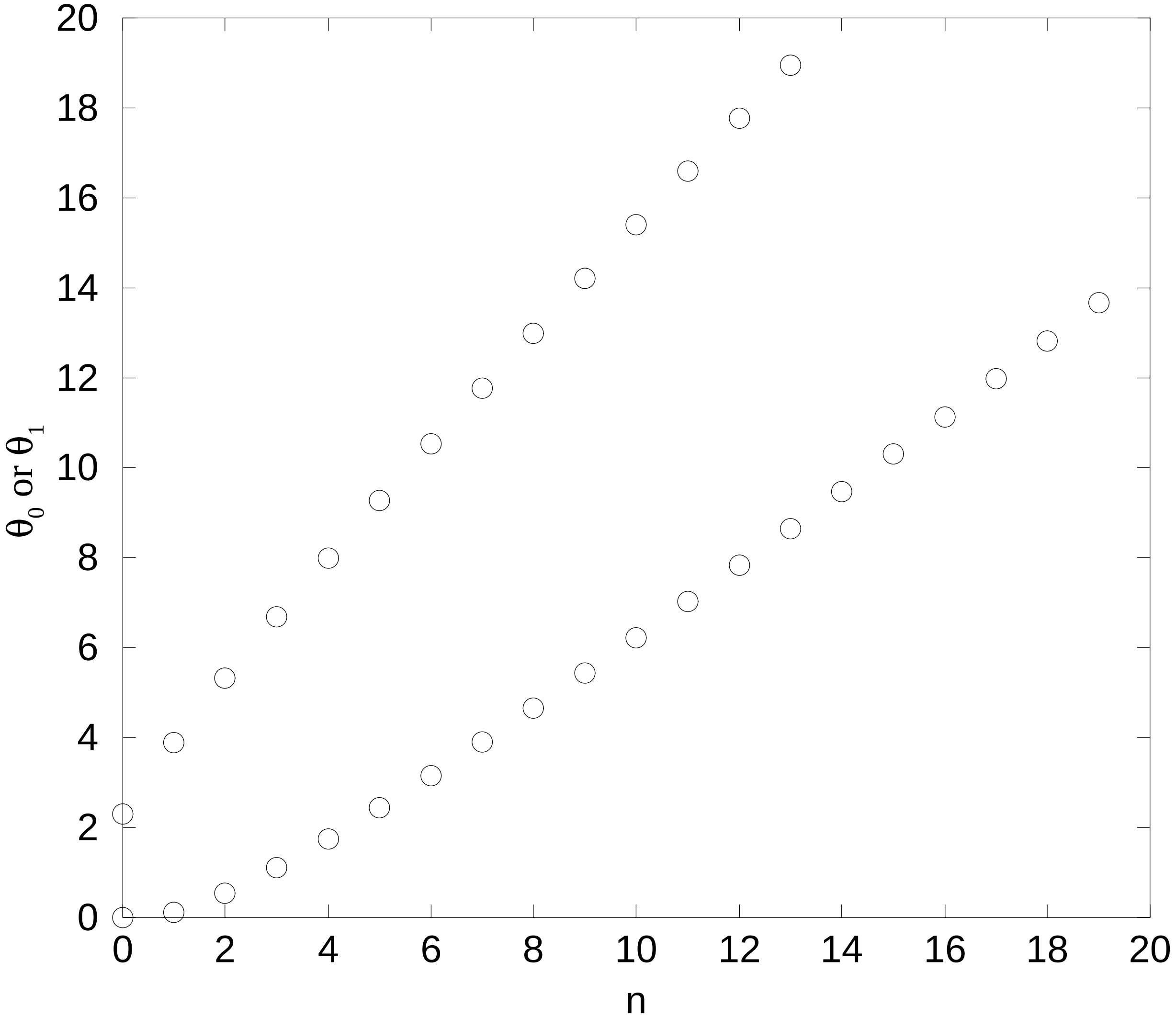
- Similarly, the "usual" lower limit, $\theta_0(n)$, is defined by
  $$1-\alpha = \sum_{k=n}^{\infty} g(k;\theta_0(n)).$$
  We define $\theta_0(0)=0$.

- Note that the prescription for θ_1 is equivalent to letting:
  $$\alpha = \int_0^{\theta_1(n)} \frac{\theta^n e^{-\theta}}{n!} d\theta,$$
  i.e., to the Bayes' interval for a uniform prior.

- Note also that this prescription doesn't give confidence intervals! The quoted limit has the property that it will include the true value of θ at least α of the time, but depending on the value of θ, it may be more often. For example, consider the case when $\theta = 0$.

Usual Poisson Upper/Lower limits

These "usual" limits are graphed as a function of n for $\alpha = 0.9$:



Poisson Example (continued)

Sketch to see how we show that the "Usual" Method does not give (Neyman) Confidence Intervals:

To see how the results of this prescription compare with the desired confidence level, we calculate for the lower limit the probability that $\theta_0 \le \theta$:

$$
\begin{aligned}
P\big(\theta_0 \le \theta\big) &= \sum_{n=0}^{\infty} p\big(n;\theta\big) P\big(\theta_0(n) \le \theta\big) \\
&= \sum_{n=0}^{n_0(\theta)} \frac{\theta^n e^{-\theta}}{n!},
\end{aligned}
$$

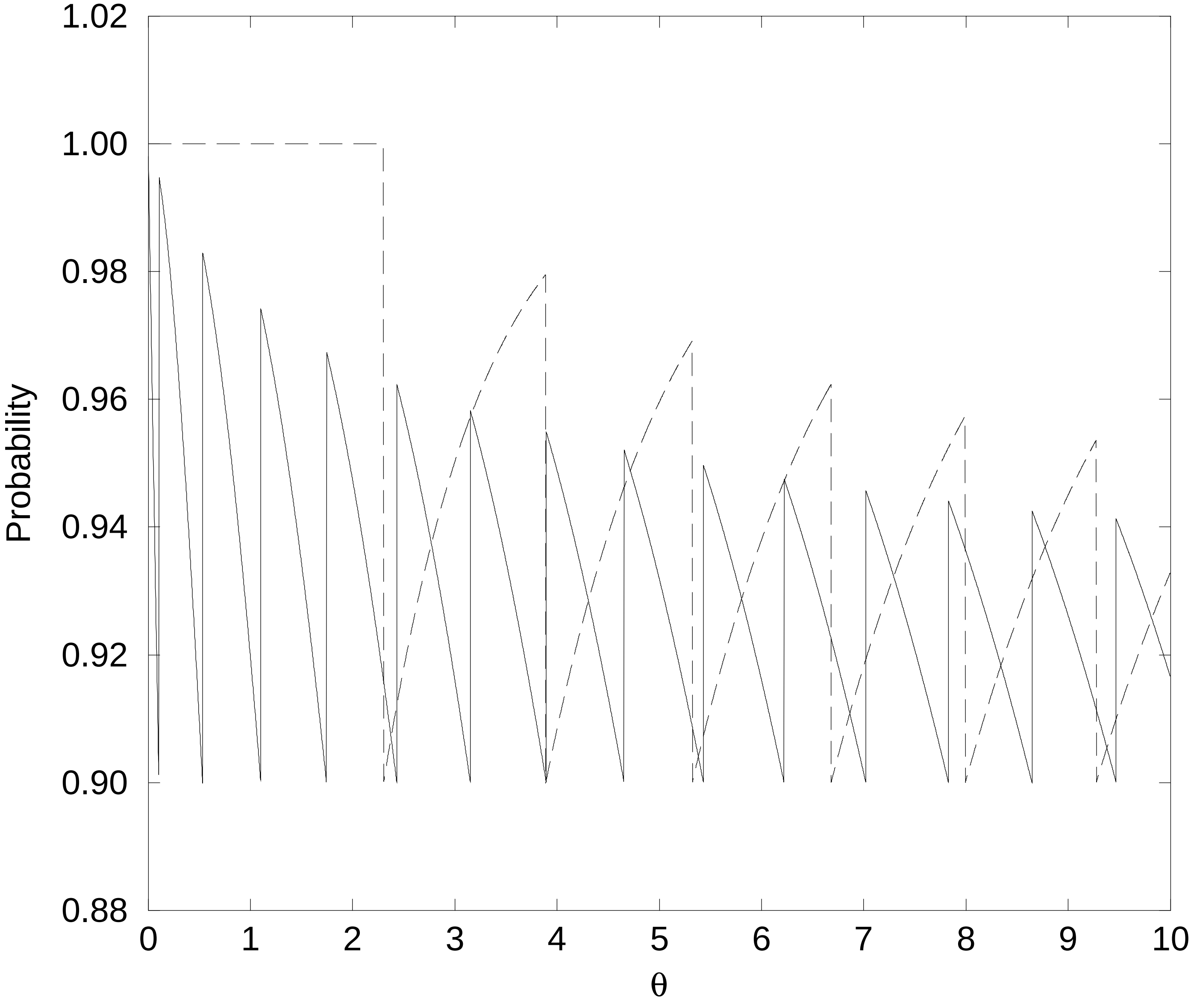
where the critical value $n_0\big(\theta\big)$ is defined according to:

$$
P\big(\theta_0(n) \le \theta\big) = \begin{cases} 1, & n \le n_0; \\ 0, & n > n_0. \end{cases}
$$

Poisson Example (continued)

Solid curve: $P\big(\theta_0 \leq \theta\big)$.

Dashed curve: $P\big(\theta_1 \geq \theta\big)$.



Poisson – Comments, Continued

- Conclusion: "Usual" prescription does not yield confidence intervals. But they may be regarded as approximate CIs, with the "conservative" property that:

$$P\big(\theta_0 \le \theta\big) \ge \alpha;$$

$$P\big(\theta_1 \ge \theta\big) \ge \alpha.$$

- We are generally willing to live with this as an approximate (and "conservative") confidence interval.

- However, it seems useful to our understanding to ask whether exact (and sufficient!) confidence intervals can, in fact, be constructed for the Poisson distribution? Remarkably, the answer is yes!

Exact Poisson Confidence intervals

Sample a value n from a Poisson distribution with mean θ.

Define variable $y = n + x$, where x is sampled from a uniform distribution:

$$f(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

y uniquely determines both n and x, hence, y is sufficient for θ, since n is.

Define $G(n;\theta)$ as the probability of observing n or more:

$$G(n;\theta) = \sum_{k=n}^{\infty} p(k;\theta).$$

Poisson Confidence Intervals (continued)

Let $y_0 = n_0 + x_0$, for some x_0 and n_0.

$$\begin{aligned} P\{y > y_0\} &= P\{n > n_0\} + P\{n = n_0\}P\{x > x_0\} \\ &= G(n_0 + 1; \theta) + g(n_0; \theta)(1 - x_0) \\ &= x_0 G(n_0 + 1; \theta) + (1 - x_0) G(n_0; \theta). \end{aligned}$$

We use this equation to derive exact confidence intervals for θ:

For a lower limit, define "critical value" $y_c = n_c + x_c$ corresponding to probability $1 - \alpha_\ell$ by:

$$\begin{aligned} P\{y > y_c\} &= 1 - \alpha_\ell \\ &= x_c G(n_c + 1; \theta) + (1 - x_c) G(n_c; \theta). \end{aligned}$$

For an observation $y_0 = n_0 + x_0$, define $\theta_\ell(y_0)$ according to:

$$1 - \alpha_\ell = x_0 G(n_0 + 1; \theta_\ell) + (1 - x_0) G(n_0; \theta_\ell).$$

Poisson Confidence Intervals (continued)

Similarly, an upper limit, θ_u, at the α_u confidence level is obtained by solving:

$$1 - \alpha_u = (x_0 - 1)\frac{\theta_u^{n_0} e^{-\theta_u}}{n_0!} + \sum_{k=0}^{n_0} \frac{\theta_u^k e^{-\theta_u}}{k!}.$$

θ_ℓ and θ_u may be set to zero if solving these equations gives less than zero.

Poisson Confidence Intervals (continued)

Whenever $y_0 > y_c$, then $\theta_\ell > \theta$, and whenever $y_0 < y_c$, then $\theta_\ell < \theta$. Since the probability of sampling a value $y_0 < y_c$ is α_ℓ, the probability that θ_ℓ is less than θ is α_ℓ. Therefore, the interval $\big(\theta_\ell, \infty\big)$ is a $100\alpha_\ell\%$ confidence interval for θ.

Can we have it all?

Physicists have long been confused/irritated by confidence intervals, especially when they give "unphysical results".

Actually, there is nothing "unphysical" going on.

- Some people just abandon classical statistics, and use Bayesian methods.

  Hence, giving up, or at least polluting, the notion of summarizing information content.

- Others have tried to have it all, by clever constructions which maintain classical validity while keeping the interval "physical"…

Example: Cousins and Feldman Intervals

Reference: "Unified Approach to the Classical Statistical Analysis of Small Signals", Feldman and Cousins ("FC"), Phys. Rev. D 57 (98) 3873.

The Method:

Construct a table of confidence intervals as function of x, for desired confidence level α:

1. For any given θ, form the ratio

   $$R(x) = p(x;\theta)/p(x;\hat{\theta}),$$

   where $\hat{\theta}$ is the maximum likelihood estimator given x, but restricted to the desired ("physical") region.

2. Find x_1, x_2 such that $R(x_1) = R(x_2)$, and:

   $$\int_{x_1}^{x_2} p(x;\theta)dx = \alpha.$$

FC Method, Continued

3. The confidence interval for θ, for a given measurement x, is given by the inverse image: $(\theta_1(x_2 = x), \theta_2(x_1 = x))$.

Exercise: Show that this procedure gives intervals with $\text{Prob}(\theta_1 < \theta < \theta_2) = \alpha$, where the probability statement is in the frequency sense concerning the random variables θ_1, θ_2.

FC Treatment of Normal Example

Suppose we know that $\theta > 0$ in our $N(\theta, 1)$ example:

In the "Usual Approach" (as above), we have

$$\begin{aligned} R(x) &= p(x;\theta)/p(x;\hat{\theta} = x) \\ &= \exp\left[-(x-\theta)^2/2\right], \end{aligned}$$

leading to 68% confidence interval $(x-1, x+1)$.

In the FC approach, this is modified to:

$$\begin{aligned} R(x) &= p(x;\theta)/p(x;\hat{\theta} = \max(x,0)) \\ &= \begin{cases} \exp\left[-(x-\theta)^2/2\right], & x \geq 0 \\ \exp\left[x\theta - \theta^2/2\right], & x < 0. \end{cases} \end{aligned}$$

FC – Normal Example Intervals

We may compare the 68% intervals from these two methods:

x	UA	FC	Bayes UL
-3	(-4, -2)	(0, 0.04)	0.33
-2	(-3, -1)	(0, 0.07)	0.44
-1	(-2, 0)	(0, 0.27)	0.64
0	(-1, 1)	(0, 1)	0.99
1	(0, 1)	(0.24, 2)	1.6
2	(1, 3)	(1, 3)	2.5

FC Example – Comments

1. The algorithm has the nice feature that it decides for you when to quote an upper limit, and when to quote a two-sided interval, helping perhaps to avoid the well-known problem that biases may be introduced when an experimenter bases his methodology on the result. More on this tomorrow.

FC Example – Comments, Continued

2. The algorithm avoids "unphysical regions" while retaining a frequency interpretation. Frequentist: "So what?" – Yes it has the frequency interpretation, but is a bit obscure as a summary of the information content. Bayesian: "I'm worried" – This seems to correspond to some prior that is more restrictive than I like. For a uniform prior, if $x = -1$, I would quote $\theta < 0.64$.

3. The assertion is also made that FC decouples the "goodness-of-fit" CL from the "confidence interval" CL. Since the information content of the FC interval is precisely the same as for the simple "usual approach", this is mysterious.

Final Remarks – Lecture 3

- We have given and examined definition of confidence interval.
- We have also defined Bayesian interval.
- They are different in interpretation, even when identical in number.
- Both have their raisons d'être.

Tomorrow – Examine some methods of interval estimation.