

# Statistics

## Lecture 4

August 9, 2000

Frank Porter

Caltech

The plan for these lectures:

1. The Fundamentals; Point Estimation
2. Maximum Likelihood, Least Squares and All That
3. What is a Confidence Interval?
4. Interval Estimation
5. Monte Carlo Methods

Additional topics will be covered by Roger Barlow and Norman Graf

## Methods of interval estimation

We have already dealt with method of error propagation in lecture 1.

1. Parabolic errors
2. Chi-square change
3. Likelihood ratios
4. Likelihood areas

## Parabolic Errors

Suppose we take a sample from a normal distribution to estimate the mean. The logarithm of the likelihood function is:

$$\ln L(x; \theta) = -(x - \theta)^2 / 2\sigma^2 + \text{constant}.$$

This describes a parabola. Let us take the second derivative, and see that:

$$\sigma^2 = -\frac{1}{\frac{\partial^2 \ln L(x; \theta)}{\partial \theta^2}}.$$

Thus, a method for estimating a 68% (e.g.) confidence interval is to take the second derivative of the logarithm of the likelihood function, plug into the above expression, and quote the interval:

$$(x - \sigma, x + \sigma).$$

## Parabolic Errors – Comments

- This is one of the easiest methods to apply, even if the differentiation must be done numerically.
- The method applies for a normal distribution, otherwise it is approximate. If the  $\log L$  distribution looks parabolic, chances are it is a good approximation (but this does not guarantee it!).
- If the second derivative differs on the high- and low-sides of the peak, it may be reasonable to quote an “asymmetric” error, using the appropriate value on each side. (but it also is a sign the estimates may not give a very precise confidence interval.)

## Chi-square Errors

Again, suppose we take a sample from a normal distribution to estimate the mean. We define a “chi-square” according to:

$$\chi^2 = (x - w(\theta))^2 / \sigma^2.$$

$w(\hat{\theta}) = x$  is our (LSE and MLE) estimator for the mean, and  $\chi^2 = 0$ , a minimum, for  $\theta = \hat{\theta}$ .

Note that the  $\chi^2$  increases from this minimum by one unit for  $w(\theta) = w(\hat{\theta}) \pm \sigma$ .

Hence, estimate the 68% (e.g.) confidence interval (for  $\theta$ ) by determining the points where the  $\chi^2$  increases by one unit from the minimum.

## Chi-square Errors – Comments

- This is another easily applied method, and is natural when the LSE method has been used to find the point estimator.
- It is equivalent, for the normal case, to the likelihood ratio method discussed below.
- For non-normal sampling distributions it yields approximate confidence intervals.
- Plotted as a function of  $\theta$ , the  $\chi^2$  may not look parabolic, and the errors may be asymmetric.

## Multi-dimensional problems, Error Ellipse

See the comments in our discussion of the LSE method for some hints on the extension of this to multi-dimensional parameter spaces.

In a two-dimensional parameter space, it is popular to plot the “**error ellipse**”, which is the contour where the  $\chi^2$  increase by one unit (or other value, as desired) from its minimum.

For example, consider the expression:

$$\chi^2 = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T H (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

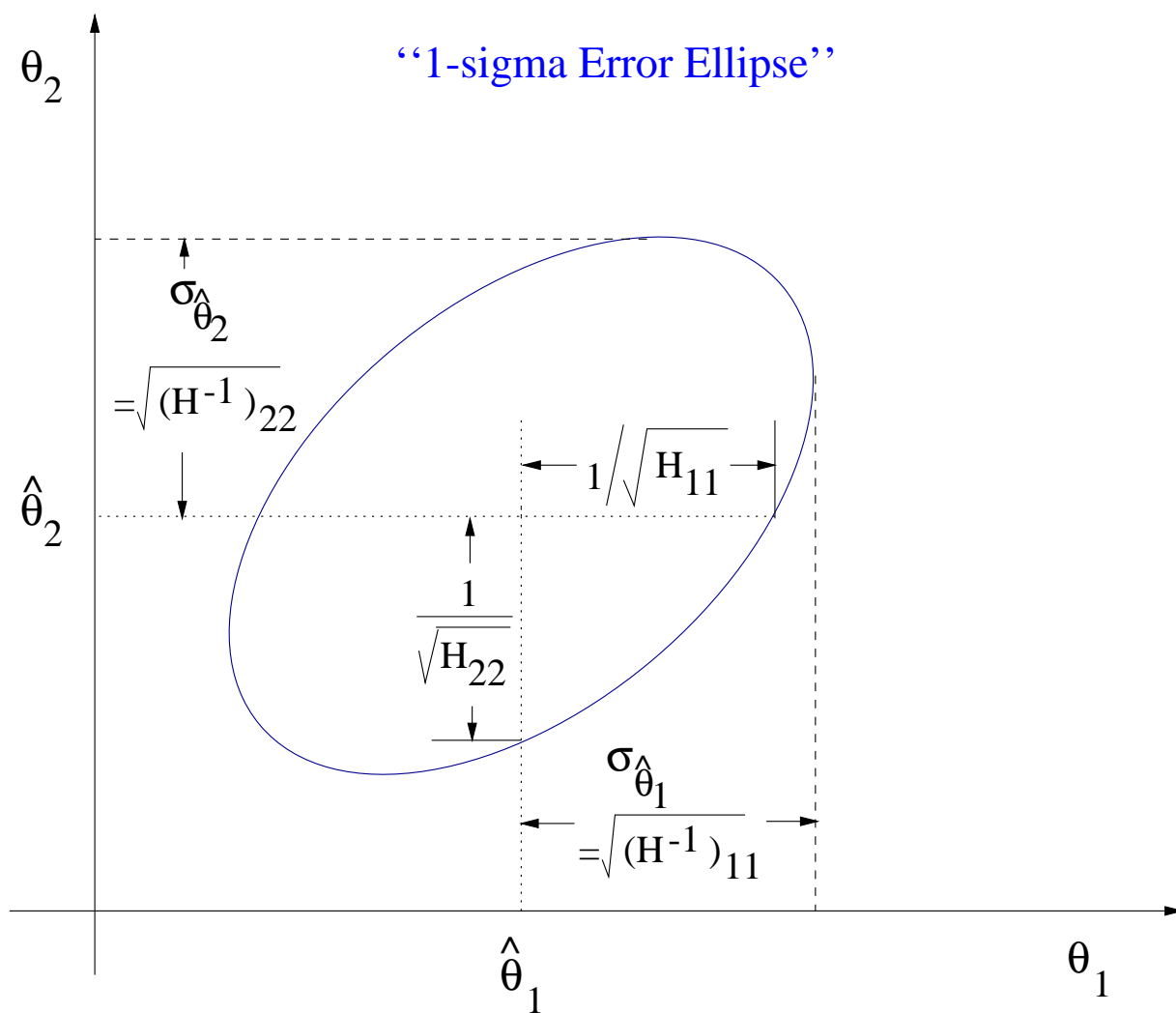
where (see LSE discussion):

$$H^{-1} = \langle (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \rangle,$$

and suppose we are in two dimensions ( $\boldsymbol{\theta} = (\theta_1, \theta_2)$ ).

Then the contour where  $\chi^2 = \chi_{\min}^2 + 1$  describes an ellipse in  $\theta$ -space. Let's look at it.

# The “1-sigma” Error Ellipse



## Error Contour – Comments

**Exercise:** Derive the above error ellipse figure.

- Note that the “1-sigma” ellipse is NOT a 68% confidence region.

**Exercise:** What is the confidence level of the 1-sigma ellipse, assuming the normal distribution?

- Projections of the 1- $\sigma$  ellipse are 68% confidence intervals (assuming normality, at least) for the one-dimensional subspaces.
- We can draw these ellipses using the likelihood ratio method (below) also.
- The illustration alerts us to the fact that, in order to find an error estimate for a single parameter, we must “re-minimize” with respect to all other parameters as we search for the  $\Delta\chi^2 = 1$  point.

## Method of likelihood ratios

A common method for obtaining a 68% “confidence interval” is to find the parameter values where the logarithm of the likelihood function decreases by 1/2 from its maximum value. This is referred to as a likelihood ratio method.

Motivation behind this method: Suppose a sample is taken from a normal distribution with known standard deviation:

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

where  $w(\theta)$  is assumed to be an invertible function of  $\theta$ . The logarithm of the likelihood function is

$$\ln \mathcal{L}(\theta; x) = -\frac{[x - w(\theta)]^2}{2\sigma^2} + \text{constant}.$$

## Method of likelihood ratios (continued)

The maximum likelihood is at  $w(\hat{\theta}) = x$ , assuming the solution exists. The points where the function decreases by  $1/2$  from the maximum are at  $w(\theta_{\pm}) = x \pm \sigma$ .

**This corresponds to a 68% confidence interval:** The probability that any sampling  $x$  will lie in the interval  $(w(\theta) - \sigma, w(\theta) + \sigma)$  is 0.68. Thus, in 68% of the times one makes a measurement (samples a value  $x$ ), the interval  $(x - \sigma, x + \sigma)$  will contain the true value of  $w(\theta)$ , and 32% of the time it will not. The probability that the interval  $(\theta_-, \theta_+)$  contains  $\theta$  is 0.68, and  $(\theta_-, \theta_+)$  is therefore a 68% confidence interval. **The probability statement is about random variables  $\theta_{\pm}$ , not about  $\theta$ .**

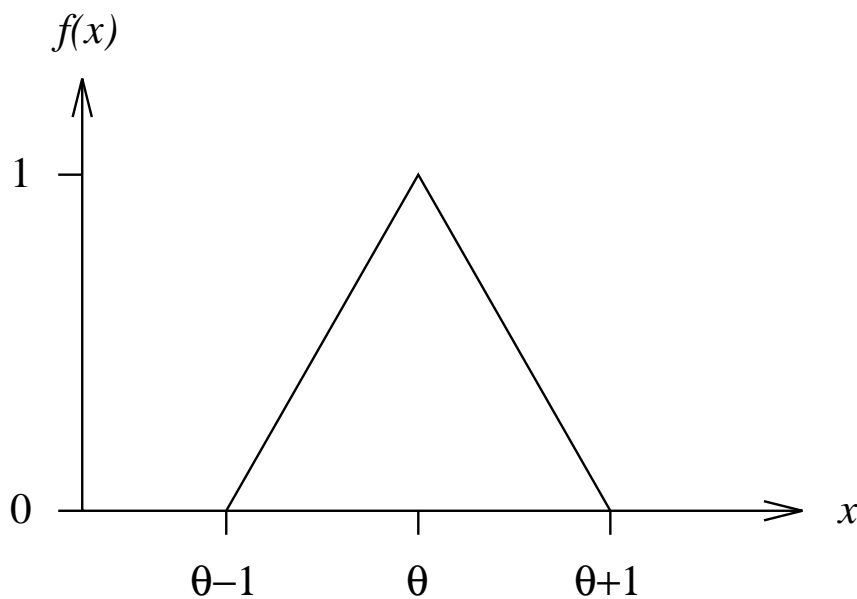
## Method of likelihood ratios (continued)

- The example assumes that the data ( $x$ ) is drawn from a normal distribution. The likelihood function (as a function of  $\theta$ ), on the other hand, is not necessarily normal. As long as  $w(\theta)$  is “well-behaved” (e.g., is invertible), the above method yields a 68% confidence interval for  $\theta$ .
- If data is sampled from a distribution which is at least approximately normal (as will be the case in the asymptotic regime if the central limit theorem applies), and the parameter of interest is related to the mean in a well-behaved manner, this method gives a confidence interval. It also has the merit of being relatively easy to calculate.

## Method of likelihood ratios (continued)

Now consider a simple non-normal distribution, and ask whether the method still works. For example, consider a “triangle” distribution:

$$f(x; \theta) = \begin{cases} 1 - |x - \theta| & \text{if } |x - \theta| < 1, \\ 0 & \text{otherwise.} \end{cases}$$



## Method of likelihood ratios (continued)

The peak of the likelihood function is  $\ln \mathcal{L}(\hat{\theta} = x; x) = 0$ . Evaluating the  $\ln \mathcal{L} - 1/2$  points:

$$\ln \mathcal{L}(\theta_{\pm}; x) = \ln(1 - |x - \theta_{\pm}|) = -1/2,$$

yields  $\theta_{\pm} = x \pm 0.393$ . Is this a 68% confidence interval for  $\theta$ ? That is, does this interval have a 68% probability of including  $\theta$ ?

Since  $\theta_{\pm}$  are linearly related to  $x$ , this is equivalent to asking if the probability is 68% that  $x$  is in the interval  $(\theta - 0.393, \theta + 0.393)$ :

$$P(x \in (\theta - 0.393, \theta + 0.393)) = \int_{\theta - 0.393}^{\theta + 0.393} f(x; \theta) dx = 0.63,$$

which is less than 68%. Thus, this method does not give a 68% confidence interval.

## Method of likelihood ratios (continued)

A correct 68% CI can be found by evaluating:

$$\text{Prob}(x \in (x_-, x_+)) = 0.68 = \int_{x_-}^{x_+} f(x; \theta) dx.$$

This gives  $x_{\pm} = \theta \pm 0.437$  (if require a symmetric interval), so that the 68% CI given result  $x$  is  $(x - 0.437, x + 0.437)$ , an interval with a 68% probability of containing  $\theta$ .

- The basic approach still works, if we use the points where the likelihood falls to a fraction 0.563 of its maximum, **but it is wrong to use the fraction which applies for a normal distribution.**
- **If the normal approximation is invalid, one can simulate the experiment (or otherwise compute the probability distribution) in order to find the appropriate likelihood ratio for the desired confidence level.**

## Method of likelihood ratios (continued)

- However, there is no guarantee that even this procedure will give a correct confidence interval, because the appropriate fraction of the maximum likelihood may depend on the value of the parameter under study. This dependence may be weak enough that the procedure is reasonable in the region of greatest interest.
- If  $\theta$  is a location parameter, or a function of a location parameter, the ratio corresponding to a given confidence level will be independent of  $\theta$ , since in that case, the shape of the distribution does not depend on the parameter. This fact can be expressed in the form of a theorem . . . .

## Method of likelihood ratios (continued)

**Theorem:** (**Likelihood Ratios**) Let  $x$  be a random variable with pdf  $f(x; \theta)$ . If there exists a transformation  $x$  to  $u$ , and an invertible transformation  $\theta$  to  $\tau$ , such that  $\tau$  is a location parameter for  $u$ , then the estimation of intervals by the likelihood ratio method yields confidence intervals. Equivalently, if  $f(x; \theta)$  is of the form:

$$f(x; \theta) = g[u(x) - \tau(\theta)] \left| \frac{du}{dx} \right|,$$

then the likelihood ratio method yields confidence intervals.

**Proof:** Exercise. See intuition ...

## Intuition (Ratio method theorem)

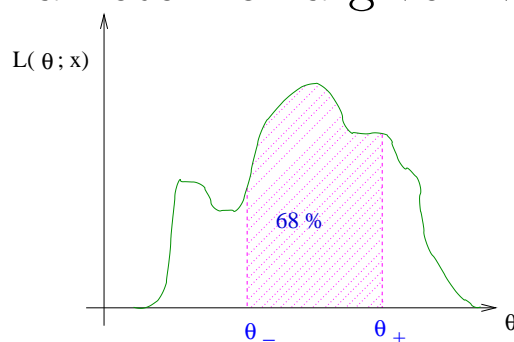
- If the parameter is a function of a location parameter, then the likelihood function is of the form (for some random variable  $x$ ):

$$\mathcal{L}(\theta; x) = f [x - h(\theta)].$$

- Finding the points according to the appropriate ratio to the maximum of the likelihood merely corresponds to finding the points in the pdf such that  $x$  is within a region around  $h(\theta)$  with probability  $\alpha$ .
- Hence, the quoted interval for  $\theta$  according to this method will be a confidence interval (possibly complicated if the inverse mapping of  $h(\theta)$  is multi-valued).

## Method of integrating the likelihood function

- Another common method of estimating intervals involves integrating the likelihood function. For example, a 68% interval is obtained by finding an interval which contains 68% of the area under the likelihood function, treated as a function of the parameter for a given value of the random variable.



- This method is often interpreted as a Bayesian method, since it yields a Bayesian interval if the prior distribution is uniform. However, it is simply a statement of an algorithm for finding an interval, and we may ask whether it yields a confidence interval, without reference to Bayesian statistics.

## Method of integrating the likelihood function

This method may be motivated by considering the normal distribution, with mean  $\theta$ , and (known) standard deviation  $\sigma$ :

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

The likelihood function, given a measurement  $x$ , is:

$$\mathcal{L}(\theta; x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x - \theta)^2}{2\sigma^2} \right].$$

If the likelihood function is integrated, as a function of  $\theta$ , to obtain a (symmetric) interval containing 68% of the total area,

$$0.68 = \int_{\theta_-}^{\theta_+} \mathcal{L}(\theta; x) d\theta,$$

we obtain  $\theta_{\pm} = x \pm \sigma$ . There is a 68% probability that the interval given by random variables  $(\theta_-, \theta_+)$  will contain  $\theta$ , and so this is a 68% confidence interval.

## Method of integrating the likelihood function

We may ask whether the method works more generally, i.e., does this method always give a confidence interval?

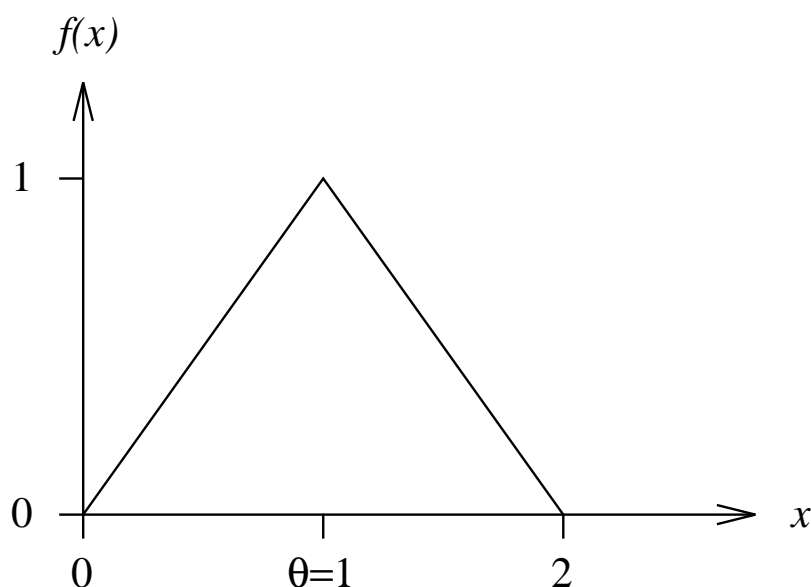
- It may be easily seen that the method also works for the triangle distribution considered earlier.
- However, we may demonstrate that the answer in general is “no”, with another simple example.

## Example: Modified “Triangle” Distribution

Consider the following modified “triangle” distribution:

$$f(x; \theta) = \begin{cases} 1 - |x - \theta| & \text{if } |x - \theta| < 1, \\ 0 & \text{otherwise.} \end{cases}$$

This distribution is shown for  $\theta = 1$ :



Sampling from this distribution provides no information on the sign of  $\theta$ . Hence, let  $\theta > 0$  stand for its magnitude.

## Modified “Triangle” Distribution (continued)

Suppose we wish to obtain a 50% confidence level upper limit on  $\theta$ , given an observation  $x$ .

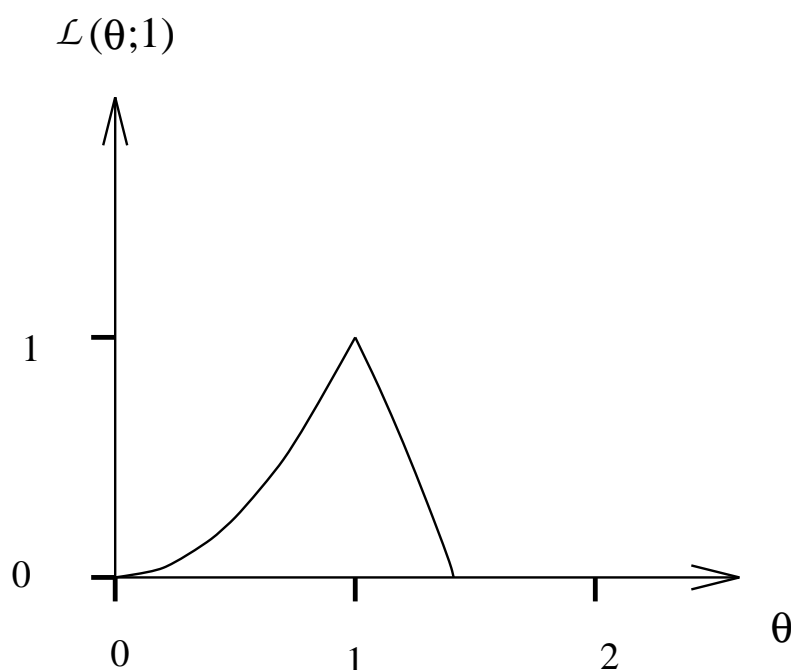
- To apply the method of integrating the likelihood function, given an observation  $x$ , solve for  $u(x)$  in the equation:

$$0.5 = \int_0^{u(x)} \mathcal{L}(\theta; x) d\theta / \int_0^{\infty} \mathcal{L}(\theta; x) d\theta.$$

- Does this procedure give a 50% confidence interval? That is, does  $\text{Prob}(u(x) > \theta) = 0.5$ ?
- If  $\theta = 1$ , 50% of the time we will observe  $x < 1$ , and 50%  $x > 1$ . Thus, the interval will be a 50% confidence interval if  $u(1) = 1$ .

## Modified “Triangle” Distribution (continued)

- From the graph of the likelihood function for  $x = 1$ , we see that 50% of the area occurs at a value of  $\theta < 1$ .



- In fact,  $u(1) = 0.94$ . Integration of this likelihood function does not give an interval with a confidence level equal to the integrated area.

## Modified “Triangle” Distribution (continued)

- Integration to 50% of the area gives an interval which includes  $\theta = 1$  with 41% probability.
- This is still not a confidence interval, even at the 41% confidence level, because the probability is not independent of  $\theta$ . As  $\theta \rightarrow \infty$ , the probability  $\rightarrow 1/2$ , and as  $\theta \rightarrow 0$ , the probability  $\rightarrow 1$ .
- The likelihood ratio method, with properly determined ratio (e.g., 0.563 for a 68% confidence level) does yield a confidence interval for  $\theta$  (with attention to signs, since  $\theta^2$  is not strictly invertible).

## Likelihood Integral Method → Confidence Interval?

**Theorem:** (**Likelihood Integral**) Let  $f(x; \theta)$  be a continuous one-dimensional probability density for random variable  $x$ , depending on population parameter  $\theta$ . Let  $I = (a(x), b(x))$  be an interval obtained by integrating the likelihood function according to:

$$\alpha = \frac{\int_{a(x)}^{b(x)} f(x; \theta) d\theta}{\int_{-\infty}^{+\infty} f(x; \theta) d\theta},$$

where  $0 < \alpha < 1$ . The interval  $I$  is a confidence interval if and only if the probability distribution is of the form:

$$f(x; \theta) = g[v(x) - \theta] \left| \frac{dv(x)}{dx} \right|,$$

where  $g$  and  $v$  are arbitrary functions.

**Proof:** Exercise.

## Likelihood Integral Theorem – Comments

- For a helpful hint towards proving, see D. Lindley, “Fiducial Distributions and Bayes’ Theorem”, Journal of the Royal Statistical Society, Series B (Methodological), 20 (1958) 102.
- Equivalently, a necessary and sufficient condition for  $I$  to be a confidence interval is that there exist a transformation  $x \rightarrow v$  such that  $\theta$  is a location parameter for  $v$ .

## Intuition (Integral method theorem)

If the parameter is a location parameter for a function of  $x$ , then the likelihood function is of the form:

$$\mathcal{L}(\theta; v(x)) = g[v(x) - \theta].$$

In this case, integrals over  $\theta$  correspond to regions of probability  $\alpha$  in  $v(x)$ , and hence in  $x$  if  $v$  is invertible.

## Comments

- The likelihood ratio and likelihood integral methods are distinct approaches, yielding different intervals.
- In the domain where the parameter is a location parameter, i.e., in the domain where the integral method yields confidence intervals, the two methods are equivalent: They yield identical intervals, assuming that intervals with similar properties (e.g., upper or lower limit, or interval with smallest extent) are being sought.

## Comments, Continued

- The ratio method continues to yield confidence intervals in some situations outside of this domain (in particular, the parameter need only be a function of a location parameter), and hence is the more general method for obtaining confidence intervals, although the determination of the appropriate ratios may not be easy.

## Discussion of case with normal likelihood function

Since the above methods both give confidence intervals for the case where  $\theta$  is the mean of a normal distribution, it is interesting to ask the following question:

- Suppose the likelihood function, as a function of  $\theta$ , is a normal function.
- Does this imply that either or both of the methods we have discussed will necessarily give confidence intervals?
- If a normal likelihood function implies that the data was sampled from a normal distribution, then this will be the case.
- However, there is no such implication, as we will demonstrate by an example.

## Normal likelihood function example

Motivate our example:

- It is often suspected (though extremely difficult to prove a posteriori) that an experimental measurement is biased by some preconception of what the answer “should be”. For example, a preconception could be based on the result of another experiment, or on some theoretical prejudice.
- A model for such a biased experiment is that the experimenter works “hard” until s/he gets the expected result, and then quits. Consider a simple example of a distribution which could result from such a scenario.

## Normal likelihood function example

- Consider an experiment in which a measurement of a parameter  $\theta$  corresponds to sampling from a Gaussian distribution of standard deviation one:

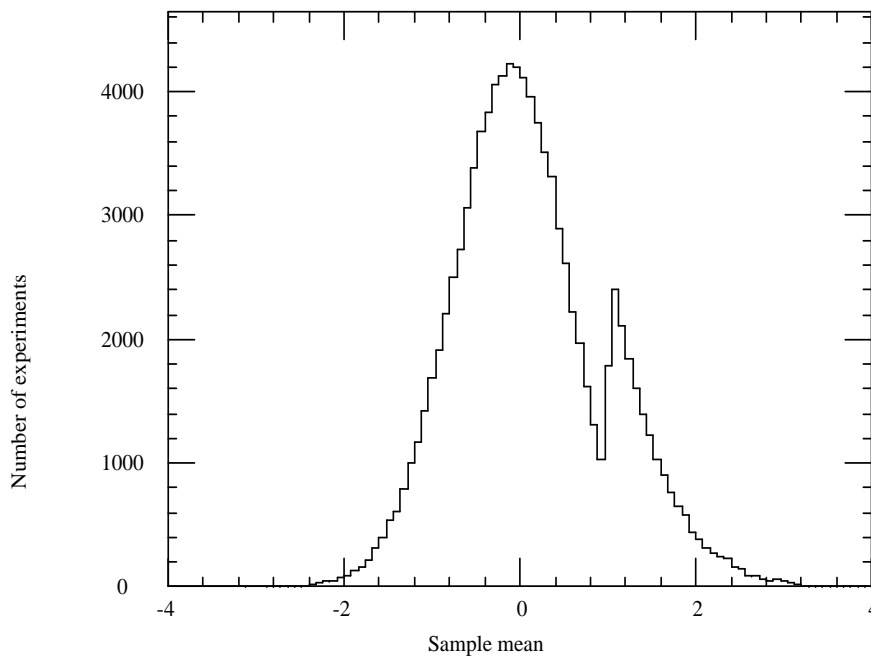
$$N(x; \theta, 1)dx = \frac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2}dx.$$

- Suppose the experimenter has a prejudice that  $\theta$  is greater than one.
- Subconsciously, he makes measurements until the sample mean,  $m = \frac{1}{n} \sum_{i=1}^n x_i$ , is greater than one, or until he becomes convinced (or tired) after a maximum of  $N$  measurements.
- The experimenter then uses the sample mean to estimate  $\theta$ .

## Normal likelihood function example

For illustration, assume that  $N = 2$ . In terms of the random variables  $m$  and  $n$ , the pdf is:

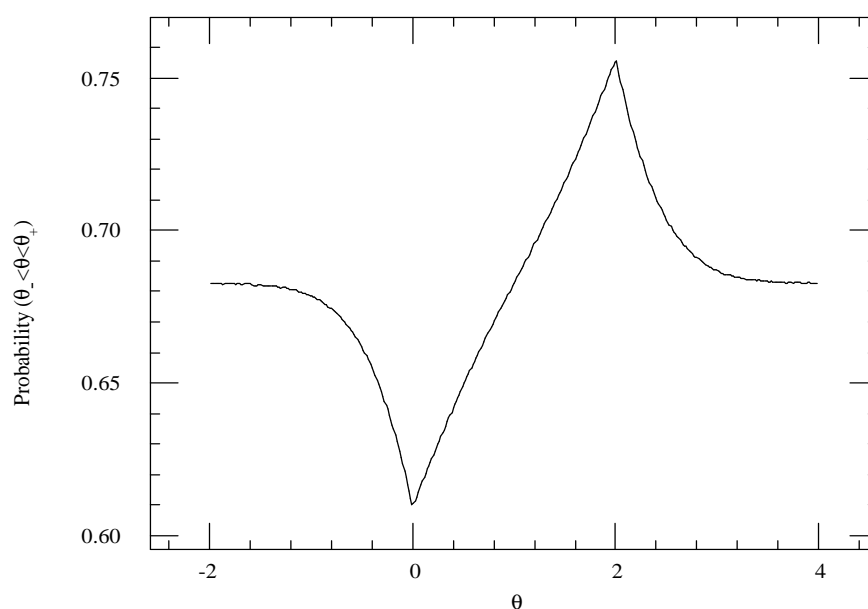
$$f(m, n; \theta) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(m-\theta)^2}, & n = 1, m > 1 \\ 0, & n = 1, m < 1 \\ \frac{1}{\pi} e^{-(m-\theta)^2} \int_{-\infty}^1 e^{-(x-m)^2} dx & n = 2 \end{cases}$$



Histogram of sampling distribution for  $m$ , with pdf given by above equation, for  $\theta = 0$ .

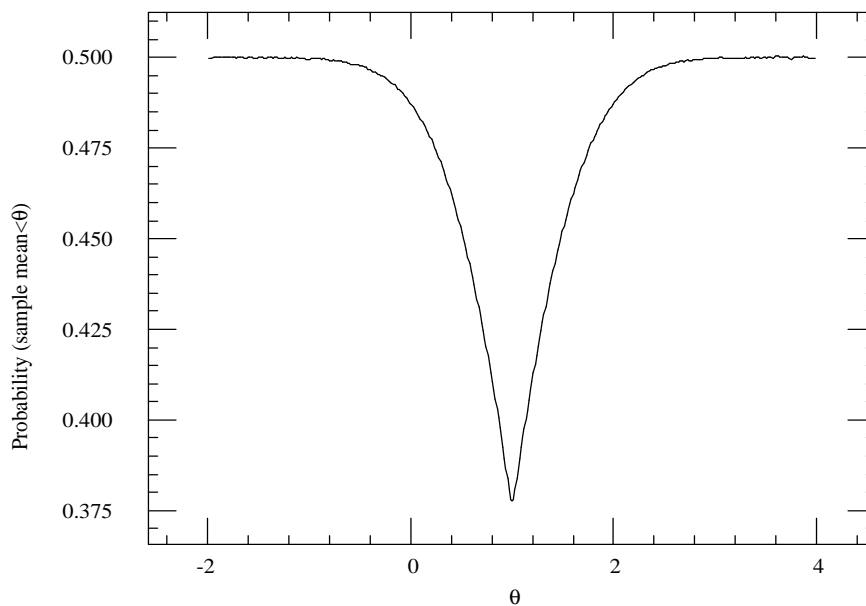
## Normal likelihood function example

The likelihood function, as a function of  $\theta$ , has the shape of a normal distribution, given any experimental result. The peak is at  $\theta = m$ , so  $m$  is the maximum likelihood estimator for  $\theta$ . In spite of the normal form of the likelihood function, the sample mean is not sampled from a normal distribution. The interval defined by where the likelihood function falls by  $e^{-1/2}$  does not correspond to a 68% CI:



## Normal likelihood function example

Integrals of the likelihood function correspond to particular likelihood ratios for this distribution, and hence also do not give confidence intervals. For example,  $m$  will be greater than  $\theta$  with a probability larger than 0.5. However, 50% of the area under the likelihood function always occurs at  $\theta = m$ . The interval  $(-\infty, m)$  thus obtained is not a 50% CI:



## Normal likelihood function example

- The experimenter in this scenario thinks he is taking  $n$  samples from a normal distribution, and uses one of these methods, in the knowledge that it works for a normal distribution.
- He gets an erroneous result because of the mistake in the distribution.
- If the experimenter realizes that sampling was actually from a non-normal distribution, he can do a more careful analysis by other methods to obtain more valid results.
- It is incorrect to argue that since each sampling is from a normal distribution, it does not matter how the number of samplings was chosen.

## Normal likelihood function example

### Philosophical Musings...

- Bayesian analysis may not care – use likelihood function
- However, Bayesian interprets his result as a degree of confidence in true answer  $\Rightarrow$  “betting odds”
- Depending on  $\theta$ , Bayesian apparently miscalculates the “betting odds”
- May be OK, because amounts to folding “prior distribution” into experiment design.

## Conclusions

- Important to understand what “confidence interval” means.
- Likelihood ratio method – Gives CI if a function of the parameter is a location parameter. But the ratio may not be  $e^{-1/2}$ .
- Likelihood integral method – Gives CI iff the parameter is a location parameter.
- Determination of CI depends on knowing the probability distribution – Not sufficient to just know the likelihood function.