CALT-68-1979 (Revised)
October 28, 1997

Interval Estimation using the Likelihood Function*

Frank Porter

Charles C. Lauritsen Laboratory of High Energy Physics
California Institute of Technology
Pasadena, California 91125

ABSTRACT

The general properties of two commonly-used methods of interval estimation for population parameters in physics are examined. Both of these methods employ the likelihood function: *i*) Obtaining an interval by finding the points where the likelihood decreases from its maximum by some specified ratio; *ii*) Obtaining an interval by finding points corresponding to some specified fraction of the total integral of the likelihood function. In particular, the conditions for which these methods give a confidence interval are illuminated, following an elaboration on the definition of a confidence interval. The first method, in its general form, gives a confidence interval when the parameter is a function of a location parameter. The second method gives a confidence interval when the parameter is a location parameter. A potential pitfall of performing a likelihood analysis without understanding the underlying probability distribution is discussed using an example with a normal likelihood function. The connection with Bayesian statistics is also noted.

(Somewhat expanded version of publication in Nucl. Inst. Meth. A **368** (1996) 793-803)

⋆ Work supported in part by the Department of Energy, grant DE–FG03–92–ER40701.

1. Introduction

We are often interested in measuring the value of some physical quantity. In statistical terms, a measurement corresponds to the sampling of a random variable from some parent probability distribution, where the quantity of interest is a parameter of this parent distribution. Based on the result of the sampling ("measurement"), we make an estimate for the value of the population parameter. An estimate of this form does not by itself provide us with an idea of how "close" we might be to the value of the underlying parameter. Thus, we usually go further, and make an estimate for an interval that has some probability of including the true value of the parameter. The sense of the term "probability" used here will require elaboration. Very often, in such measurements, this interval is estimated by one of two methods employing the likelihood function.

The purpose of this paper is to examine the properties of the estimated intervals obtained by these two likelihood function methods. We are especially interested in the question of whether these intervals are confidence intervals. Before addressing this question, a discussion of the definition of a confidence interval is presented, in order to clarify the confusion surrounding this notion. We will insist here on a rigorous usage of this term, and hence avoid the confusion which common casual usage engenders. Examples are used throughout to illustrate the ideas.

The two methods under consideration here are: (i) Evaluating a $100\alpha\%$ confidence interval [1] by finding the points where the logarithm of the likelihood function falls by an amount $d(\alpha)$ from its maximum value (*e.g.*, by $d(\alpha) = 1/2$ for a 68% interval). This may be referred to as the "likelihood ratio" method. (ii) Evaluating a $100\alpha\%$ confidence interval by finding points which contain $100\alpha\%$ of the area under the likelihood function, which may be referred to as the "likelihood integral" method.

Both techniques have the desired property in classical (meaning "frequentist" here) statistics if the data is sampled from a normal (Gaussian) distribution, where the parameter of interest is the mean. A question which often arises is whether either of these methods has the desired property in other cases. A third common method, based on evaluating the second derivative of the logarithm of the likelihood function ("parabolic errors"), has limitations which are widely understood, and is not considered here.

It should be noted at the outset that this paper is not about philosophical arguments concerning such matters as "Bayesian" *vs.* "classical or frequentist" statistics [2], or even about which confidence intervals are "best". However, the author has found that many such arguments among physicists are mixed up with misconceptions concerning the definitions and properties discussed in this paper, rather than deeper issues. It is thus hoped that this paper may serve the dual role of clarifying the properties of the interval estimation methods examined, and also of providing a better foundation for such philosophical considerations.

The results of this paper are, briefly, as follows: The problem of evaluating confidence intervals for a particular probability distribution is in general very difficult, but there are two simpler methods employing an analysis of the likelihood function which are widely used. Typically, these methods only give approximate confidence intervals, but it is possible to make some general conclusions concerning when they produce true confidence intervals: The likelihood ratio method, in its general form, gives a confidence interval when the parameter is a function of a location parameter. The likelihood integral method gives a confidence interval when the parameter is a location parameter. It is also pointed out that there are pitfalls in performing a likelihood analysis without understanding the underlying probability distribution.

2. Definitions, Confidence Intervals

The discussion in this note is restricted to one-dimensional distributions, that is, we consider sampling some random variable x from a probability distribution $P(x)$. The probability distribution is assumed to involve a parameter, denoted as θ. We would like to make inferences concerning this parameter, given a "measurement" x drawn from $P(x)$. The words "measurement" and "experiment" are used to mean a sample x from probability distribution $P(x)$. Most of the discussion will be further restricted to continuous probability distributions (the sample space for x consisting of continuous subsets of the real numbers). In this case, we use the abbreviations "pdf" for the probability density function $f(x;\theta)$ and "cdf" for the cumulative distribution function $F(x;\theta)$. The random variable will be treated as having a range $(-\infty,\infty)$, with the understanding that the pdf is defined to be zero outside of the allowed region for any given situation.

Def. *Likelihood function:* If an experiment has been performed resulting in a measurement x, drawn from some probability distribution with population parameter θ, the **likelihood function** for that experiment is defined as the probability (density) evaluated at the observed value of x.

The likelihood function is often denoted by $\mathcal{L}(\theta;x)$, and is treated as a function of θ in some algorithms for making inferences concerning θ. Since this is what we shall do, we choose to make the first argument θ for this discussion. Considerations of Bayesian *vs.* classical statistics are irrelevant in this definition of the likelihood function.

The notion of a location parameter will be useful in later discussions:

Def. *Location parameter:* If the pdf is of the form:

$$f(x;\theta) = f(x-\theta), \tag{2.1}$$

then θ is called a **location parameter** for x.

A confidence interval (or classical confidence interval) may be defined according to: [3]

Def. *Confidence interval:* A **confidence interval**, at the α confidence level, for a population parameter θ is an interval obtained by a prescription such that it will include θ a fraction α of the time such an experiment is performed and the prescription followed. This property must be independent of θ (which need not even be the same from one experiment to the next).

Note that the confidence interval may not be unique – there may be more than one interval (*i.e.*, more than one prescription) which satisfies this property. We are not concerned here with the question of the "best" choice of interval.

The meaning of the term "confidence interval" as given above, is not always well-understood, so it is worth elaborating further. It is useful to recognize that the confidence interval is purely a measure of the information content of the data. Questions concerning the "physicalness" or "non-physicalness" of whether the true value of the parameter could be in the resulting interval are meaningless in this context. Methods which attempt to constrain an interval to a "physical region" (*e.g.*, ref. [4]) generally do not yield confidence intervals – they often have the property that the quoted interval includes, in the frequency sense, the true value with a probability at the stated confidence level *or greater*, with the actual probability dependent on the value of the parameter. Because of this property, such intervals may be thought of as "conservative confidence intervals", but it should be understood that the interval may no longer be simply related to the information content of the measurement. Properties of such intervals are sometimes described in terms of "coverage" (*e.g.*, ref. [5]): Our "conservative confidence interval" is an interval with "over-coverage".

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

Def. *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. (A Bayes' distribution is a function of θ of the form $f(x;\theta)p(\theta)/\int_{-\infty}^{\infty} f(x;\theta)p(\theta)d\theta$, where $p(\theta)$ is a non-negative function called the prior distribution (or, simply, the "prior").)

The Bayesian interval and the confidence interval are often numerically identical, and often confused, but they are not the same in general, and have quite different interpretations. The "conservative confidence interval" referred to above may sometimes be interpreted as a Bayesian interval, where the physical constraints have been incorporated into the prior, although this interpretation is not required.

It is instructive to briefly address the question of existence of a confidence interval. Such an interval always exists in a trivial sense – for a 68% confidence interval, simply choose $(-\infty,\infty)$ 68% of the time, and an infinitely narrow interval 32% of the time. This prescription ignores the information in the data, and is not useful. It is more interesting to ask the question in terms of sufficient statistics. We consider this in the context of a particular distribution which is frequently encountered in physical measurements – the Poisson distribution,

$$g(n;\theta) = \frac{\theta^n e^{-\theta}}{n!}, \qquad n = 0, 1, 2, \ldots \tag{2.2}$$

In so doing, it is intended that the meaning of a confidence interval will be further elucidated.

It is not obvious that it is possible, in the case of discrete distributions such as the Poisson, to find a confidence interval which has the required property, independent of the value of θ. That is, with the Poisson distribution, there is at first sight no way to quote a 68% confidence interval for this distribution, based on the random variable n, which will include θ in 68% of the samples, *independent* of the value of θ.

In such cases we often quote intervals based on the "worst case", *i.e.*, which will include θ a fraction α of the time for some θ, and more than α of the time for other values of θ. For example, if we are interested in obtaining an upper limit on the mean of a Poisson distribution, given an observation n, at the ($\geq$) α confidence level, the "usual" technique is to solve the following equation for $\theta_1(n)$:

$$1 - \alpha = \sum_{k=0}^{n} g(k; \theta_1(n)). \tag{2.3}$$

This is just the probability that at most n would be observed *if* $\theta = \theta_1(n)$. It is easily demonstrated that $\theta_1(n)$ is the same quantity here as obtained by solving:

$$\alpha = \int_0^{\theta_1(n)} \frac{\theta^n e^{-\theta}}{n!} d\theta. \tag{2.4}$$

It can be demonstrated that Prob $[\theta_1(n) \geq \theta] \geq \alpha$, for all θ. For example, in the case $\theta = 0$, then $n = 0$ will always be observed, and $\theta_1(0) = -\ln(1-\alpha)$ will be greater than θ (*i.e.*, > 0) more than a fraction $0 < \alpha < 1$ of the time. As $\theta = 0$ is an extreme example, we will discuss the general case below.

Similarly with the procedure for an upper limit, the "usual" lower limit, $\theta_0(n)$, is defined by

$$1 - \alpha = \sum_{k=n}^{\infty} g(k; \theta_0(n)). \tag{2.5}$$

We define $\theta_0(0) = 0$. The corresponding integral form for this limit is:

$$\alpha = \int_{\theta_0(n)}^{\infty} \frac{\theta^{n-1} e^{-\theta}}{(n-1)!} d\theta. \tag{2.6}$$

These upper and lower limits are graphed in Fig. 1 as a function of n for $\alpha = 0.9$.

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

$$\begin{aligned} P(\theta_0 \leq \theta) &= \sum_{n=0}^{\infty} g(n;\theta) P(\theta_0(n) \leq \theta) \\ &= \sum_{n=0}^{n_0(\theta)} \frac{\theta^n e^{-\theta}}{n!}, \end{aligned} \tag{2.7}$$

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

$$P(\theta_0(n) \leq \theta) = \begin{cases} 1, & n \leq n_0; \\ 0, & n > n_0. \end{cases} \tag{2.8}$$

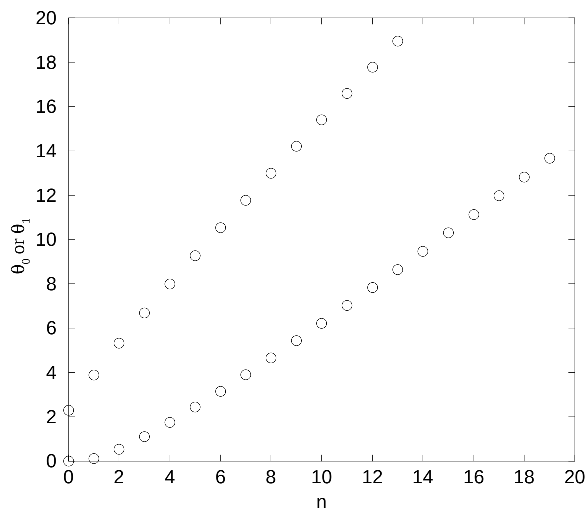

Figure 1. The "usual" upper and lower one-sided limits ($\alpha = 0.9$) for the mean of the Poisson distribution as a function of the observed number, n.

Similarly, for the case of an upper limit, the probability that $\theta_1 \geq \theta$ is:

$$P(\theta_1 \geq \theta) = \sum_{n=n_1(\theta)}^{\infty} \frac{\theta^n e^{-\theta}}{n!}, \tag{2.9}$$

where the critical value $n_1(\theta)$ is defined according to:

$$P(\theta_1(n) \geq \theta) = \begin{cases} 1, & n \geq n_1; \\ 0, & n < n_1. \end{cases} \tag{2.10}$$

These probabilities are shown as a function of θ in Fig. 2, for $\alpha = 0.9$, and for both the lower limit, θ_0, and the upper limit θ_1. It is seen that the probability that the lower (or upper) interval, defined by this method, contains θ is not independent of θ, but is always at least as large as the desired α. Thus, this method is biased in the "conservative" direction (over-covers) from a true confidence interval, that is, a quoted interval contains the value of the parameter with at least the quoted probability. This bias can be very substantial, depending on the value of θ.

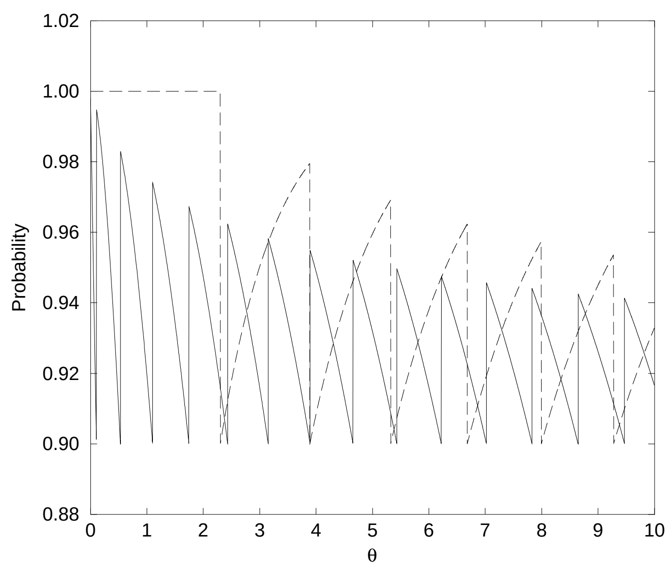

Figure 2. The probability that the "usual" lower limit for the mean of a Poisson distribution (solid curve), or upper limit (dashed curve), will be below, or above, respectively, the value of θ, as a function of θ (for $\alpha = 0.9$).

However, rigorous confidence intervals for discrete distributions such as the Poisson do exist. The method of obtaining such an interval employs a clever trick which eliminates the difficulty introduced by the discreteness of the sample space. The idea is to add an additional random variable to make the overall distribution continuous, in such a way that no loss in information is suffered (*i.e.*, sufficiency is maintained). By quoting a confidence interval based on the continuous distribution, an exact confidence interval for θ is obtained. Here is how it works [6], for the Poisson distribution:

Suppose we do an experiment where we sample a value n from a Poisson distribution with mean θ. Define variable $y \in (0, \infty)$ by $y = n + x$, where x is sampled from a uniform distribution:

$$f(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

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:

$$G(n; \theta) = \sum_{k=n}^{\infty} g(k; \theta). \tag{2.12}$$

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

$$\begin{aligned} \text{Prob}\{y > y_0\} &= \text{Prob}\{n > n_0\} + \text{Prob}\{n = n_0\}\text{Prob}\{x > x_0\} \\ &= G(n_0 + 1; \theta) + g(n_0; \theta)(1 - x_0) \\ &= x_0 G(n_0 + 1; \theta) + (1 - x_0) G(n_0; \theta). \end{aligned} \tag{2.13}$$

We may use this equation to derive exact confidence intervals for θ.

To accomplish this, for a lower limit, define "critical value" $y_c = n_c + x_c$

corresponding to probability $1-\alpha_\ell$ by:

$$\begin{aligned} \text{Prob}\{y > y_c\} &= 1-\alpha_\ell \\ &= x_c G(n_c+1;\theta) + (1-x_c)G(n_c;\theta) \end{aligned} \tag{2.14}$$

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

$$1-\alpha_\ell = x_0 G(n_0+1;\theta_\ell) + (1-x_0)G(n_0;\theta_\ell). \tag{2.15}$$

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, *i.e.*, θ_ℓ is a lower limit on θ at the α_ℓ confidence level. The interval (θ_ℓ, ∞) is a $100\alpha_\ell\%$ confidence interval for θ. Explicitly, the α_ℓ confidence level lower limit θ_ℓ is obtained by solving the following equation:

$$1-\alpha_\ell = -x_0\frac{\theta_\ell^{n_0}e^{-\theta_\ell}}{n_0!} + \sum_{k=n_0}^{\infty}\frac{\theta_\ell^k e^{-\theta_\ell}}{k!}. \tag{2.16}$$

We may see the similarity with Eq. (2.5), and how that equation is modified to obtain a confidence interval.

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

$$1-\alpha_u = (x_0-1)\frac{\theta_u^{n_0}e^{-\theta_u}}{n_0!} + \sum_{k=0}^{n_0}\frac{\theta_u^k e^{-\theta_u}}{k!}. \tag{2.17}$$

θ_ℓ and θ_u may be set to zero if solving these equations gives less than zero. The interval (θ_ℓ, θ_u) corresponds to a $\alpha_\ell + \alpha_u - 1$ two-sided confidence interval for θ.

This example, using the familiar Poisson distribution, has been discussed at some length, in order to elucidate the meaning of the term "confidence interval". It has been shown that the usual method for estimating an interval for the mean of this distribution does not yield a confidence interval. It has also been demonstrated that it is possible to construct a confidence interval for this parameter. It is important to recognize that information has been neither gained nor lost in this procedure involving the use of an additional random variable. Instead, a procedure has been found which eliminates the systematic overestimate of the interval which the usual procedure makes due to the discrete nature of the distribution. In spite of the existence of this exact procedure, there seems to this author to be insufficient practical advantage to warrant advocating its adoption over the simpler approximate method in widespread use.[7]

3. Method of likelihood ratios

A very commonly-used method for obtaining a 68% confidence interval is to find the value of the parameter for which the logarithm of the likelihood function achieves its maximum value, and then to find the points where the logarithm decreases by 1/2 from this maximum value. This is often referred to as a likelihood ratio method. For simplicity of discussion, we refer explicitly to a 68% confidence interval for a while – this is readily generalized to other confidence levels.

Let us first review the motivation behind this method [8]. Suppose that a sample is taken from a normal distribution. The logarithm of the likelihood function is

$$\ln \mathcal{L}(\theta; x) = -\frac{[x - w(\theta)]^2}{2\sigma^2} + \text{constant}, \tag{3.1}$$

where $w(\theta)$ is assumed to be an invertible function of θ. The standard deviation, σ, is assumed to be known. The constant term is unimportant, since we are considering only differences. The maximum of this function is at the value of $\theta = \theta^*$ which satisfies $w(\theta^*) = x$, assuming the solution exists. The points where the function decreases by 1/2 from the maximum are at $w(\theta_\pm) = x \pm \sigma$. Now let us see that this in fact corresponds to a 68% confidence interval: Since the sampling is from a normal distribution, the probability that 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. Hence, the interval for θ evaluated according to $w(\theta_\pm) = x \pm \sigma$ is a 68% confidence interval – the probability that the interval (θ_-, θ_+) contains θ is 0.68. Remember that $\theta_\pm$ are the random variables, so the probability statement is about $\theta_\pm$, not about θ.

It should be understood that the example above assumes that the data (x) is drawn from a normal distribution. The likelihood function (considered as a function of θ), 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 θ.

We thus see the motivation for this technique of evaluating confidence intervals. As long as the 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 of the distribution in a well-behaved manner, this method gives the desired confidence interval. It also has the merit of being relatively easy to calculate.

Now consider a simple non-normal distribution, and ask whether the method still works. A tractable example is to choose a "triangle" distribution:

$$f(x;\theta) = \begin{cases} 1 - |x-\theta| & \text{if } |x-\theta| < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

This distribution is shown in Fig. 3.

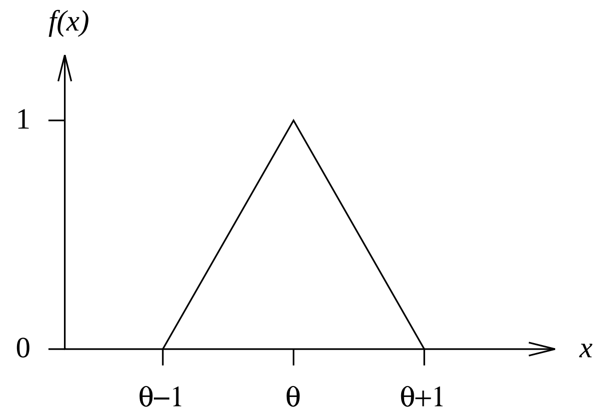

Figure 3. Triangle distribution of Eq. (3.2).

The corresponding likelihood function, given some measurement x, is:

$$\mathcal{L}(\theta; x) = \begin{cases} 1 - |x-\theta| & \text{if } |x-\theta| < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

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

$$\ln \mathcal{L}(\theta_\pm; x) = \ln\left(1 - |x - \theta_\pm|\right) = -1/2, \tag{3.4}$$

yields $\theta_\pm = x \pm 0.393$. Is this a 68% confidence interval for θ? That is, does this interval have a 68% probability of including θ? To answer this, the probability distribution for $\theta_\pm$ must be determined. This is easy here, since $\theta_\pm$ are linearly related to x. The problem is thus equivalent to asking if the probability is 68% that x is in the interval $(\theta - 0.393, \theta + 0.393)$. We find:

$$\text{Prob}\left(x \in (\theta - 0.393, \theta + 0.393)\right) = \int_{\theta-0.393}^{\theta+0.393} f(x;\theta)\, dx = 0.63, \tag{3.5}$$

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

A correct 68% confidence interval can be obtained for this distribution by evaluating:

$$\text{Prob}\,(x \in (x_-, x_+)) = 0.68 = \int_{x_-}^{x_+} f(x;\theta)\,dx. \tag{3.6}$$

This gives the result $x_\pm = \theta \pm 0.437$ (if we require a symmetric interval), so that the 68% confidence interval given result x is $(x - 0.437, x + 0.437)$, an interval which has a 68% probability of containing θ.

Note that the basic approach still works in this example, since we could use the points where the likelihood falls to a fraction 0.563 of its maximum, but it is wrong to use the points one would obtain for a normal distribution. When 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. However, there is no guarantee that even this procedure will give a correct confidence interval, because the appropriate fraction of the maximum likelihood will typically have some dependence 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. In the case that θ is a location parameter, or a function of a location parameter, the ratio corresponding to a given confidence level will be independent of θ, since in that case, the shape of the distribution does not depend on the parameter. This fact will be repeated below in the form of a theorem.

We briefly consider whether there are distributions besides the normal distribution for which the normal procedure works. The theorem given below may be used to see that it works for certain distributions which are related to the normal. However, it may be seen that the answer is otherwise no, at least for simple cases. For example, suppose that θ is a location parameter for x, that the distribution peaks at $x = \theta$, and that the distribution is symmetric about θ:

$$\begin{aligned} f(x;\theta) &= f(|x-\theta|) \\ \max_x f(|x-\theta|) &= f(0). \end{aligned} \tag{3.7}$$

Further assume that the prescription which is valid for the normal distribution is also valid here, for all confidence levels. In this case, we may argue that the distribution must therefore be normal:

Consider a normal distribution $n(x;\theta,\sigma)$ which has mean θ, and σ adjusted so that the peak value coincides with that of f: $n(\theta;\theta,\sigma) = f(0)$. The statement that the normal prescription works for f is equivalent to the statement that an interval defined by the points where f decreases from its maximum by any given ratio k contains the same integrated probability as the interval obtained by the same prescription for any normal distribution, in particular the normal distribution we have just specified above. For this to be true, f must in fact coincide with our normal distribution: Consider a small neighborhood of $x = \theta$. We can choose such a neighborhood sufficiently small that either $f(|x-\theta|) \leq n(x;\theta)$ or $f(|x-\theta|) \geq n(x;\theta)$ everywhere in the neighborhood. Only if strict equality holds will the desired probability content be achieved. For example, if $f(|x-\theta|) > n(x;\theta)$ in the neighborhood, then the probability of finding a value of x in a region (contained within the neighborhood) specified by the prescription will be greater for distribution f than for a normal distribution. This argument may be repeated for ever larger neighborhoods to find that f must be a normal distribution.

It is possible to make a statement about a general class of probability distributions for which the likelihood ratio method (with appropriately determined ratios) yields confidence intervals:

Theorem: Let x be a random variable with pdf $f(x;\theta)$. If there exists a transformation x to u, and an invertible transformation θ to τ, such that τ 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:

$$f(x;\theta) = g\left[u(x) - \tau(\theta)\right] \left|\frac{du}{dx}\right|, \tag{3.8}$$

then the likelihood ratio method yields confidence intervals.

Proof: Consider a pdf of the form $g(u-\tau)$, *i.e.*, where τ is a location parameter for u. Define "critical values", $u_\pm$, corresponding to probability α by:

$$\mathrm{Prob}(u_+ > u > u_-) = \int_{u_-}^{u_+} g(u-\tau)\,du = \alpha, \tag{3.9}$$

subject to the constraint:

$$g(u_+ - \tau) = \rho g(u_- - \tau), \tag{3.10}$$

where ρ is a given number, usually chosen to be $\rho = 1$ for a two-sided interval, or $\rho = 0$ or ∞ for an upper or lower limit, respectively. This may not be sufficient to uniquely determine $u_\pm$; for present purposes any $u_\pm$ pair satisfying the above properties may be selected. Also, let u_0 be a value of u for which the pdf is maximal:

$$g(u_0 - \tau) = \max_u g(u-\tau). \tag{3.11}$$

Define "ratios" $r_\pm$ according to:

$$r_\pm = g(u_\pm - \tau)/g(u_0 - \tau). \tag{3.12}$$

Given a sampling u, the likelihood function is given by $\mathcal{L}(\tau;u) = g(u-\tau)$. If τ^* is a value for τ where the likelihood function is maximal, then $\mathcal{L}(\tau^*;u) =$

$g(u_0 - \tau)$. The likelihood ratio method of finding an interval (τ_-, τ_+), corresponding to confidence level α, consists in solving for $\tau_\pm$ in the following:

$$r_\pm = \mathcal{L}(\tau_\mp; u)/\mathcal{L}(\tau^*; u). \tag{3.13}$$

Comparing with Eq. (3.12), we have $\mathcal{L}(\tau_\mp; u) = g(u - \tau_\mp) = g(u_\pm - \tau)$. Selecting one of these equations, $g(u - \tau_-) = g(u_+ - \tau)$, choose the solution for which $u - \tau_- = u_+ - \tau$. Note that, while τ and u_+ are unknown, the difference $u_+ - \tau$ is known, since the pdf is given. Hence, the prescription for calculating $\tau_- = u - (u_+ - \tau)$ is complete. Similarly, determine $\tau_+ = u + (\tau - u_-)$.

Now it can be seen that (τ_-, τ_+) is a confidence interval for τ, at the α confidence level: From $\tau_- = u - (u_+ - \tau)$, whenever $u > u_+$, then $\tau_- > \tau$, and whenever $u < u_+$, then $\tau_- < \tau$. From $\tau_+ = u + (\tau - u_-)$, whenever $u > u_-$, then $\tau_+ > \tau$, and whenever $u < u_-$, then $\tau_+ < \tau$. Since $\text{Prob}(u_+ > u > u_-) = \alpha$, then $\text{Prob}(\tau_+ > \tau > \tau_-) = \alpha$ also. Hence, (τ_-, τ_+) is a confidence interval for τ, at the α confidence level.

Consider now the generalization of this result to the form stated in the theorem:

$$f(x; \theta) = g\left[u(x) - \tau(\theta)\right] \left|\frac{du}{dx}\right|. \tag{3.8}$$

$u(x)$ is itself a random variable, and $\tau(\theta)$ is a location parameter for u. The pdf for u is just $g(u - \tau)$. This is of the form just considered above, so we may evaluate a confidence interval at the α confidence level for τ, (τ_-, τ_+), according to the likelihood ratio method. Since $\tau(\theta)$ is assumed to be invertible, we may define $\theta_\pm$ as the solutions to $\tau(\theta_\pm) = \tau_\pm$ (assume, for simplicity, that $\theta_+ > \theta_-$; the opposite case can be dealt with similarly), and (θ_-, θ_+) is an $\alpha\%$ confidence interval for θ, since $\text{Prob}(\theta_- < \theta < \theta_+) = \text{Prob}(\tau_- < \tau < \tau_+) = \alpha$.

Let τ^* be a value of τ for which the likelihood function is maximal. In terms of θ, the likelihood function is maximal at θ^*, defined by $\tau(\theta^*) = \tau^*$. For a given observation x, we have the values of the likelihood function at these points of interest:

$$\begin{aligned} f(x;\theta_\pm) &= g\left[u(x) - \tau(\theta_\pm)\right]\frac{du}{dx} \\ f(x;\theta^*) &= g\left[u(x) - \tau(\theta^*)\right]\frac{du}{dx}. \end{aligned} \tag{3.14}$$

Therefore:

$$\frac{f(x;\theta_+)}{f(x;\theta^*)} = \frac{g\left[u(x) - \tau(\theta_+)\right]}{g\left[u(x) - \tau(\theta^*)\right]} = r_-, \tag{3.15}$$

where r_- is the value of the ratio calculated with random variable u according to Eq. (3.12). This completes the proof.

4. Method of integrating the likelihood function

Another commonly employed method of estimating intervals involves integrating the likelihood function. For example, a 68% interval would be 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. It is necessary to comment immediately that this method is often interpreted as a Bayesian method. This connection will be made in section 6. However, it is simply a statement of an algorithm for finding an interval, and we are free to ask whether it yields a confidence interval, without reference to Bayesian statistics.

This method may be motivated by again considering the normal distribution, with mean θ, and standard deviation σ, where σ is known. If the likelihood function is integrated, as a function of θ, to obtain a (symmetric) interval containing 68% of the total area,

$$0.68 = \int_{\theta_-}^{\theta_+} \mathcal{L}(\theta; x)\, d\theta, \tag{4.1}$$

we obtain $\theta_\pm = x \pm \sigma$. There is a 68% probability that the interval given by random variables (θ_-, θ_+) will contain θ, and so this is a 68% confidence interval. Note that the $\ln \mathcal{L}_{\max} - 1/2$ method gives the same interval for this distribution.

With this motivation, 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 in the preceeding section. However, we may demonstrate that the answer in general is "no", with another simple example. We have actually seen one example of the failure of this method, as this is the "usual" method of obtaining interval estimates for the Poisson distribution (see Eq. (2.4)). Later, we shall state a theorem describing the realm of validity of this method.

Consider the following modified "triangle" distribution:

$$f(x;\theta) = \begin{cases} 1 - |x - \theta^2| & \text{if } |x - \theta^2| < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

This distribution is shown in Fig. 4 for $\theta = 1$. Sampling from this distribution provides no information on the sign of θ. Hence, let $\theta > 0$ stand for its magnitude.

Suppose we wish to obtain a 50% confidence level upper limit on θ. To apply the method of integrating the likelihood function, given an observation x, we must solve for $u(x)$ in the equation:

$$0.5 = \int_0^{u(x)} \mathcal{L}(\theta;x)\,d\theta \Big/ \int_0^{\infty} \mathcal{L}(\theta;x)\,d\theta. \tag{4.3}$$

Does this procedure give a 50% confidence interval? That is, does $\mathrm{Prob}(u(x) > \theta) = 0.5$?

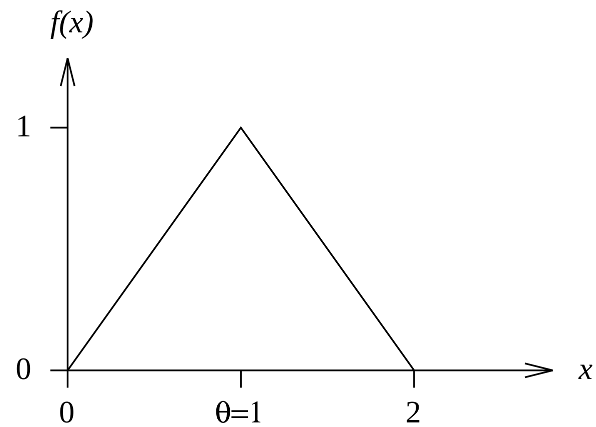

Figure 4. Modified triangle distribution of Eq. (4.2), for $\theta = 1$.

Since $\theta = 1$, 50% of the time we will observe $x < 1$, and 50% $x > 1$. Thus, the interval will be a 50% confidence interval if $u(1) = 1$. The likelihood function is graphed in Fig. 5 for $x = 1$. It is already clear from this figure that 50% of the area occurs for a value of θ which is less than one. In fact, $u(1) = 0.94$. Therefore, integration of this likelihood function does not give an interval with a confidence level equal to the integrated area. Integration to 50% of the area gives, for this case, an interval which includes $\theta = 1$ with approximately a 41% probability. This is still not a confidence interval, even at the 41% confidence level, however, because the probability is not independent of θ. As $\theta \to \infty$, the probability approaches 1/2, and as $\theta \to 0$, the probability approaches 1. Note that the likelihood ratio method, with an appropriately determined ratio (*e.g.*, 0.563 for a 68% confidence level) does yield a confidence interval for θ (with due attention to signs, since θ^2 is not strictly invertible).

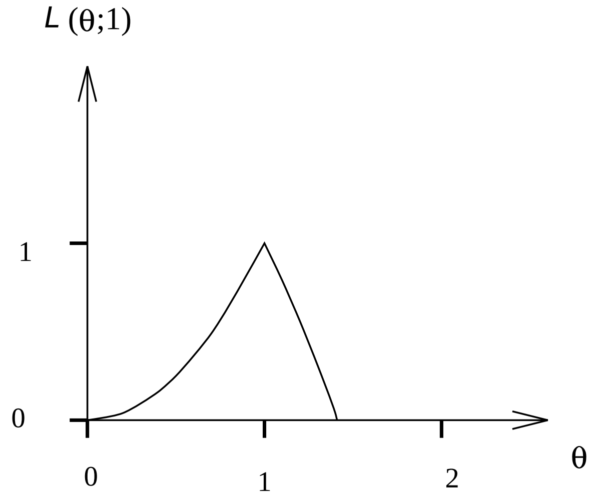

Figure 5. Likelihood function for modified triangle distribution with $x = 1$.

We may address the question in general of the necessary and sufficient conditions for the integrals of the likelihood function to yield confidence intervals. The theorem which we will prove makes use of the following theorem from Lindley: [9]

Theorem: (Lindley) "The necessary and sufficient condition for the fiducial distribution of θ, given x, to be a Bayes' distribution is that there exist transformations of x to u, and of θ to τ, such that τ is a location parameter for u."

Proof: (following Lindley) The fiducial distribution [10] is

$$\phi_x(\theta) = -\partial_\theta F(x;\theta), \tag{4.4}$$

The restrictions on cdf F are that the derivative exist, that $\lim_{\theta\to\infty} F(x;\theta) = 0$, and $\lim_{\theta\to-\infty} F(x;\theta) = 1$.

Assuming the pdf $f(x;\theta)$ exists, the Bayes' distribution corresponding to prior

$p(\theta)$ is:

$$f(x;\theta)p(\theta)/\rho(x) \tag{4.5}$$

where

$$\rho(x) = \int_{-\infty}^{\infty} f(x;\theta)p(\theta)\,d\theta. \tag{4.6}$$

Thus, we wish to find the condition for the existence of a solution to

$$-\partial_\theta F(x;\theta) = [\partial_x F(x;\theta)]\,p(\theta)/\rho(x). \tag{4.7}$$

Since the F=constant solution is not permitted, a solution exists only if F is of the form:

$$F(x;\theta) = G\left(R(x) - P(\theta)\right), \tag{4.8}$$

where G is an arbitrary function, and $R(x)$ and $P(\theta)$ are the integrals of $\rho(x)$ and $p(\theta)$.

If F is of the above form, the existence of a solution to (4.7) may be demonstrated by considering the random variable $u = R(x)$ and parameter $\tau = P(\theta)$. Then $F = G(u - \tau)$, and it can be verified by substitution that a solution to (4.7) exists with a uniform prior in the parameter τ: $p(\tau)$= constant. Since τ is a location parameter for u, this completes the proof.

We are ready to state the theorem of interest:

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

$$\alpha = \frac{\int_{a(x)}^{b(x)} f(x;\theta)\,d\theta}{\int_{-\infty}^{+\infty} f(x;\theta)\,d\theta}, \tag{4.9}$$

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

$$f(x;\theta) = g\left[v(x) - \theta\right]\left|\frac{dv(x)}{dx}\right|, \tag{4.10}$$

where g and v are arbitrary functions. Equivalently, a necessary and sufficient condition for I to be a confidence interval is that there exist a transformation $x \to v$ such that θ is a location parameter for v.

Proof: The proof consists mainly in showing that this is a special case of Lindley's theorem.

Consider pdf $f(x;\theta)$, with cdf $F(x;\theta)$: $f(x;\theta) = \partial_x F(x;\theta)$. It will be sufficient to discuss one-sided intervals, since other intervals can be expressed as combinations of these. Thus, we wish to find a confidence interval specified by random variable $u(x)$ such that:

$$\text{Prob}(u(x) > \theta) = \alpha. \tag{4.11}$$

That is, $u(x)$ is a random variable which is greater than θ with probability α. We will assume that u exists and is invertible (hence also unique). This corresponds to a value of x which is greater than (or possibly less than, a case which may be dealt with similarly) $x_\theta = u^{-1}(\theta)$ with probability α. If $p(u;\theta)$ is the pdf for u, we require:

$$\int_{-\infty}^{\theta} p(u;\theta)\,du = 1 - \alpha, \tag{4.12}$$

or, in terms of the pdf for x:

$$\int_{-\infty}^{x_\theta} f(x;\theta)\,dx = F(x_\theta;\theta) = 1 - \alpha. \tag{4.13}$$

Given a sample x, we use this equation by setting $x_\theta = x$, and solving $F(x;u(x)) = 1 - \alpha$ for $u(x)$. This procedure has the required property, since if $x < x_\theta = u^{-1}(\theta)$, then $u(x) < \theta$, and if $x > x_\theta$, then $u(x) > \theta$.

Now we wish to find the condition on $f(x;\theta)$ such that this interval is the same as the interval obtained by integrating the likelihood function. That is, we seek the condition such that $u(x) = u_b(x)$, where:

$$\frac{\int_{-\infty}^{u_b(x)} f(x;\theta)d\theta}{\int_{-\infty}^{\infty} f(x;\theta)\,d\theta} = \alpha. \tag{4.14}$$

The left-hand-side is the integral of a Bayes' distribution, with prior $p(\theta) = 1$. The $u(x) = u_b(x)$ requirement is thus:

$$\int_{-\infty}^{x} f(x';u)dx' = 1 - \int_{-\infty}^{u} f(x;\theta)\,d\theta/\rho(x), \tag{4.15}$$

with $\rho(x)$ as defined in Eq. (4.6). Differentiating with respect to u yields

$$-\partial_u F(x;u) = f(x;u)/\rho(x). \tag{4.16}$$

Since this must be true for any α we choose, hence for any u for a given x, this corresponds to the situation in Lindley's theorem with a uniform prior for the Bayes' distribution in θ. Thus, this condition is satisfied if and only if F is of the form $F(x;\theta) = G(\int\rho(x)dx - \theta)$, or $f(x;\theta) = G'(\int\rho(x)dx - \theta)\rho(x)$. With $v(x) = \int\rho(x)dx$ this is in the form as stated. If θ is a location parameter for a function of x, we may verify that (4.16) holds by substitution. This completes the proof.

The two methods, likelihood ratio and likelihood integral, are distinct approaches, yielding different intervals. However, 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. In this domain, they yield identical intervals, assuming that intervals with similar properties (*e.g.*, upper or lower limit, or interval with smallest extent) are being sought. 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.

The intuition should now be fairly clear: If the parameter is a function of a location parameter, then the likelihood function is of the form (for some random variable x):

$$\mathcal{L}(\theta; x) = f\left[x - h(\theta)\right]. \tag{4.17}$$

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 α. Hence, the quoted interval for θ according to this method will be a confidence interval (although it may be complicated if the inverse mapping of $h(\theta)$ is multi-valued). If the parameter is a location parameter for a function of x, then the likelihood function is of the form:

$$\mathcal{L}(\theta; v(x)) = g\left[v(x) - \theta\right]. \tag{4.18}$$

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

5. Discussion of case with normal likelihood function

Since the above methods both work (*i.e.*, give confidence intervals) for the case where θ is the mean of a normal distribution, it is interesting to ask the following question: Suppose the likelihood function, as a function of θ, 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.

Let us 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 he gets the expected result, and then quits. We consider a simple example of a distribution which could result from such a scenario.

Consider an experiment in which a measurement of a parameter θ corresponds to sampling from a Gaussian distribution of standard deviation one:

$$n(x;\theta,1)dx = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} dx. \tag{5.1}$$

Now suppose the experimenter has a prejudice that θ must be 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 θ.

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

$$f(m,n;\theta) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(m-\theta)^2}, & n = 1,\ m > 1 \\ 0, & n = 1,\ m < 1 \\ \frac{1}{\pi} e^{-(m-\theta)^2} \int_{-\infty}^{1} e^{-(x-m)^2}\, dx & n = 2. \end{cases} \tag{5.2}$$

A histogram of the sampling distribution for m is shown in Fig. 6. The fact that there are two random variables need not concern us here – if desired, we could map this onto a function of a single random variable without loss of information.

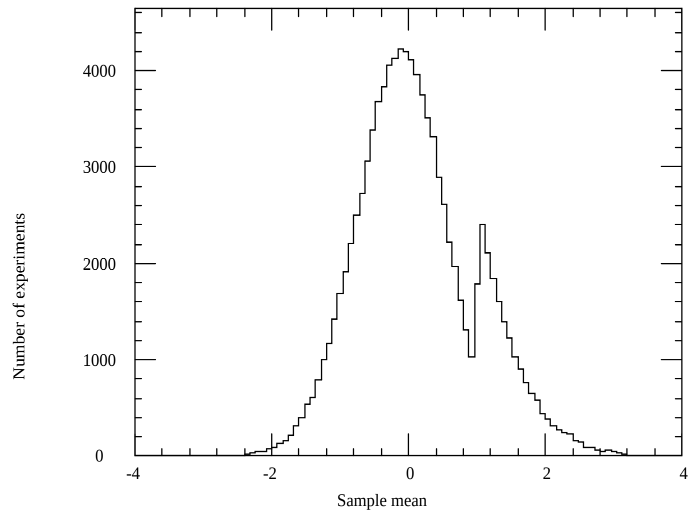

Figure 6. Histogram of sampling distribution for m, with pdf given by equation (5.2), for $\theta = 0$.

The interesting fact concerning this distribution, in the present discussion, is that the corresponding likelihood function, as a function of θ, has the shape of a normal distribution, given any experimental result. The peak of the distribution is at $\theta = m$, so m is the maximum likelihood estimator for θ. 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% confidence interval, as shown in Fig. 7.

Integrals of the likelihood function correspond to particular likelihood ratios for this distribution, and hence also do not give confidence intervals. For example, notice that m will be greater than θ 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 does not correspond to a 50% confidence interval, as illustrated in Fig. 8. Note that $f(m, n; \theta)$ is not of the form in Eq. (4.10).

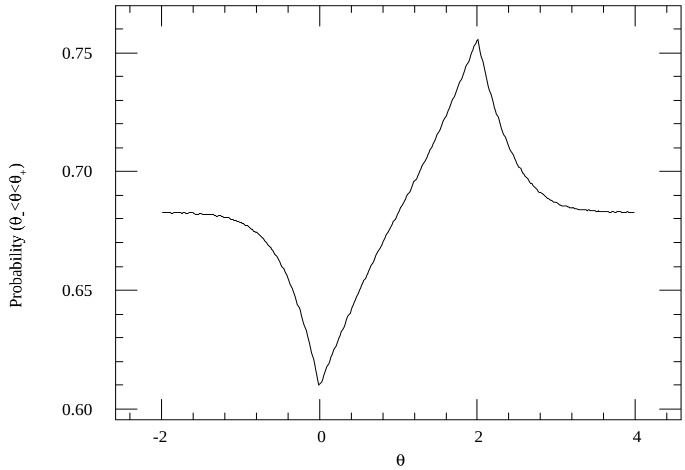

Figure 7. The probability that the likelihood ratio method (for a ratio $e^{-1/2}$) yields an interval which contains the true value of the parameter, for the distribution in Eq. (5.2).

The experimenter in the scenario of this example thinks he is merely taking n samples from a normal distribution, and uses one of these methods, in the knowledge that they work for a normal distribution. He gets an erroneous result because of the mistake in the distribution. If the experimenter realises that he has actually sampled from Eq. (5.2), he can do a more careful analysis by other methods to obtain more valid results. It should be noted that it would be incorrect to argue that since each sampling is from a normal distribution, it does not matter how the number of samplings was chosen.

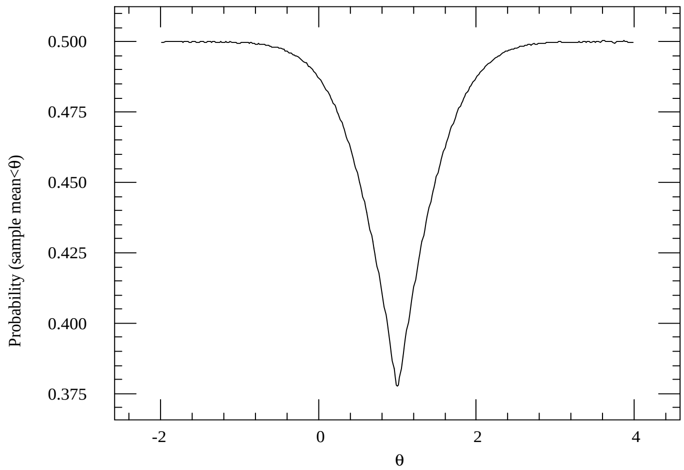

Figure 8. The probability that the likelihood integral method (for an integral of 1/2) yields a lower limit interval which contains the true value of the parameter, for the distribution in Eq. (5.2).

6. Bayesian interpretation

In this discussion, we have stayed within the regime of classical statistics, avoiding mention of Bayesian methods. This was done for two reasons: (*i*) It is recommended practice to quote one's results using classical statistics, leaving up to each reader, or to additional consideration, the option of whatever Bayesian interpretation appears appropriate (see, *e.g.*,[2]). (*ii*) It was desirable not to complicate the discussion with the ill-defined nature of the prior in Bayesian statistics. However, the method involving integrals of the likelihood function has a natural interpretation in terms of Bayesian statistics, which will be now made clear.

In Bayesian statistics, one deals with a probability distribution for the parameter(s) of interest. This probability distribution incorporates the experimental observations, and perhaps also information from previous experiments, "physical" constraints, and any personal ("theoretical") prejudices that seem desirable to include. Since the population parameter (call it θ) of interest is not a random variable, this probability distribution for θ is not a probability distribution in the usual frequency sense – *i.e.*, one can not sample from this distribution and obtain various values of θ. However, it can be defined with the formal properties of a probability, and the interpretation that it may be given is that it expresses the individual's relative "beliefs" concerning the true value of the parameter.

We remark that, in discussions of "classical *vs.* Bayesian" statistics, it should be kept clearly in mind what question is being asked. A classical statistic is an attempt to summarize relevant information content in the data, while a Bayesian statistic is an attempt to provide guidance on the value of the unknown parameter. If the distinction is understood, considerable confusion may be avoided.

A Bayesian analysis makes use of Bayes' theorem to take the likelihood function for an experiment, plus any "prior" information ($P_0(\theta)$), and turn it into a probability distribution for the parameter of interest:

$$P(\theta|x) = \frac{\mathcal{L}(\theta;x)P_0(\theta)}{\int \mathcal{L}(\theta;x)P_0(\theta)\,d\theta}. \tag{6.1}$$

For the sake of the current discussion, let us assume that the prior has been taken to be a constant in the parameter of interest (if the prior is interpreted as a probability function, this may be imagined as the limit of a uniform distribution within very wide limits), and that the likelihood function has been (re)normalized so that its integral over all values of the parameter is one. In this case, the desired probability distribution is precisely the likelihood function:

$$P(\theta|x) = \mathcal{L}(\theta;x). \tag{6.2}$$

Hence, intervals obtained by integrals of the likelihood function are identical to Bayesian intervals in which a uniform prior is assumed. It seems necessary to state the inverse as well: If a non-uniform prior is assumed, then such integrals of the likelihood function do not yield Bayesian intervals.

This can lead to apparently remarkable results from the frequentist perspective. For example, consider the example in the preceding section, in which the experimenter worked hard to get the "right" answer. A Bayesian analysis with uniform prior will quote a 50% confidence lower limit at $\theta = m$, even though the pdf can be analyzed to see that this interval would exclude the true value more than 50% of the time (see Fig. 8). This, however, is the notion of "confidence" in classical statistics, which need not coincide with the Bayesian notion. Indeed, the experimenter has in effect already folded his prior into the experiment design, and, as a Bayesian, is comfortable with the result.

7. Conclusions

Two commonly employed methods of evaluating intervals, based on the likelihood function, have been examined in the context of whether they produce confidence intervals. When one is sampling from an approximately normal distribution, the $\ln \mathcal{L} - 1/2$ method (for a 68% confidence interval) is typically pretty close. For non-normal distributions, the likelihood ratio method will still give valid confidence intervals if the parameter under study is a function of a location parameter, but the appropriate ratio must be determined for the actual distribution. For other cases, approximately valid results may still be obtained, for limited regions of parameter space, again by determining the appropriate ratio from the actual distribution, for the interesting region of parameter space.

The method of integrating the likelihood function may be interpreted as a Bayesian approach, with a flat prior. In certain cases it can yield a confidence interval, but in general it does not. The condition that it does is that the probability distribution must be of the form:

$$f(x;\theta) = g\left[v(x) - \theta\right] \frac{dv(x)}{dx}, \tag{7.1}$$

If the parameter is a location parameter for some function of x, then the distribution is of this form, and the method will give confidence intervals. This is a subset of the realm of validity of the ratio method.

A normal shape to the likelihood function suggests, but does not guarantee that these methods can be applied, as it is possible to have a normal likelihood function, without having sampled from a normal distribution.

Acknowledgements

I am grateful to Thomas Gottschalk and Robert Cousins for reading and commenting on the manuscript.

REFERENCES

[1] Typical (but not uniform) usage in physics, which is adopted here, is that the "confidence level" (α) corresponds to the probability that the confidence interval will include the true value of the population parameter. In other statistics literature, the opposite convention is often used, where this probability is $1 - \alpha$. To further complicate matters, the notation $1 - \alpha$ is often used for the "confidence level".

[2] A nice discussion touching on such matters may be found in: F. James and M. Roos, *Phys. Rev. D* **44**, 299 (1991).

[3] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Macmillan Pub. Co. Chapters 20-21 of volume 2 (fourth edition) are particularly relevant to our topic.

[4] G. Zech, *Nucl. Instr. and Meth.* **A277** (1989) 608. See also V. Innocente and L. Lista, *Nucl. Instr. and Meth.* **A340** (1994) 396.

[5] R. Cousins, *Am. J. Phys.* **63** (1995) 398.

[6] W. L. Stevens, "Fiducial limits of the parameter of a discontinuous distribution", *Biometrika* **37** (1950) 117. This reference develops the method using the example of a binomial distribution. A summary is given in ref. [3]. A table of limits for the Poisson distribution according to this method is given in C. R. Blyth and D. W. Hutchinson, "Table of Neyman-shortest unbiased confidence intervals for the Poisson parameter", *Biometrika* **48** (1961) 191.

[7] That is, and this is admittedly philosophical, the mistakes introduced by the approximate procedure do not appear to have sufficiently severe consequences to warrant a more exact treatment. However, there is a recent article, R. D. Cousins, *Nucl. Instr. and Meth.* **A337** (1994) 557, which advocates a method for obtaining upper limits based on the same idea. The method proposed therein does not yield a confidence interval, due to a difficulty in handling the case where $n = 0$ is observed, and the adoption of an *ad hoc* procedure for this case.

[8] For example, see W. T. Eadie, *et al.*, *Statistical Methods in Experimental Physics*, North-Holland Pub. Co. (1971), pp 203-205.

[9] D. V. Lindley, "Fiducial Distributions and Bayes' Theorem", *Journal of the Royal Statistical Society, Series B (Methodological)*, **20** (1958) 102.

[10] This may be regarded as the definition of the fiducial distribution. Further discussion concerning fiducial statistics may be found, *e.g.*, in ref. [3].