Confidence intervals and limits

V. Blobel

Universität Hamburg

Zeuthen, May 9 – 11,2001

- Classical confidence intervals – Neyman construction
- Bayesian estimates and intervals
- Poisson with background
- Gaussian with boundary at origin

Probability and the normal distribution

Deviation of a single value x from the mean μ in case of a *normal* distribution:

$$\begin{array}{lcl} \text{2–sided} & P(|x-\mu| < 1.960\,\sigma) & = & 95\,\% \\ \text{1–sided} & P(x-\mu < 1.645\,\sigma) & = & 95\,\% \end{array}$$

This relation between probability P and standard deviations σ is valid only for the *normal* distribution!

Poisson process – upper limit

Assume $n_{\text{obs}} = 5$ events are observed.

What is the upper limit μ_b of the mean value at the 95 % C.L.?

Naive approach: use $1.645\,\sigma$ with $\sigma = \sqrt{n_{\text{obs}}}$.

$$\mu_b = 5 + 1.645 \cdot \sqrt{5} = 8.68$$

Is this correct?

Assume Poisson distribution with $\mu = 8.68$. What is the probability to observe $k = 5, 4, 3, 2, 1, 0$ events? $P = 13.6\,\%$ i.e. an upper limit of $\mu_b = 8.68$ is not correct (undercoverage).

Assume Poisson distribution with $\mu = 10.51$. What is the probability to observe $k = 5, 4, 3, 2, 1, 0$ events? $P = 5.00\,\%$

The limit $\mu_b = 10.51$ is a 95 % C.L. upper limit for the observed number $n_{\text{obs}} = 5$.

Poisson process – lower limit

Naive approach:

$$\mu_a = 5 - 1.645 \cdot \sqrt{5} = 1.32$$

Assume Poisson distribution with $\mu = 1.32$. What is the probability to observe $k = 5, 6, 3, \ldots$ events? $P = 1.1\,\%$ i.e. overcoverage.

Assume Poisson distribution with $\mu = 1.97$. What is the probability to observe $k = 5, 6, 3, \ldots$ events? $P = 5.000\,\%$

Zero observed events

Important special case: $n_{\text{obs}} = 0$
Upper limit for the confidence level $1 - \beta$:

$$\begin{aligned} \beta &= P(k=0;b) = P_b(0) = e^{-\mu_b} \\ \rightarrow \mu_b &= -\ln \beta \end{aligned}$$

Numerical example:	95 % C.L.	$\mu_b = 2.996$
	99 % C.L.	$\mu_b = 4.6$
	99.9 % C.L.	$\mu_b = 6.9$

If the true value is $\mu = 6.9$ or > 6.9, then the probability to observe $n = 0$ events is 0.1% or less.

Classical confidence intervals

A confidence interval is a member of a set, such that the set has the property that

$$P(\mu \in [\mu_a, \mu_b]) = 1 - \varepsilon$$

The limits μ_a and μ_b are functions of the measured n_{obs}, and the equation refers to the varying confidence intervals $[\mu_a, \mu_b]$ from a ensemble of experiments with fixed μ.

The intervals $[\mu_a, \mu_b]$ contain the fixed unknown μ_t in a fraction of $1 - \varepsilon$ experiments.

The set of confidence intervals has the correct *coverage*, if the equation is true for every allowed μ.

Basic principle of classical confidence intervals is *coverage*.

Example: if $n_{\text{obs}} = 5$ is observed, the confidence interval at the 90 % C.L. is [1.97, 10.51].

Poisson limits

The number n of events has been observed. What are limits for the underlying mean $\mu = E[n]$, while the observed values k follow the Poisson distribution? The maximum likelihood estimate for μ is $\widehat{\mu} = n$.

$$\begin{aligned}
\alpha &= P(k \geq n; \mu_a) = \sum_{k=n}^{\infty} P_a(k) = 1 - \sum_{k=0}^{n-1} P_a(k) = 1 - \sum_{k=0}^{n-1} \frac{\mu_a^k}{k!}\, e^{-\mu_a} \\
\beta &= P(k \leq n; \mu_b) = \sum_{k=0}^{n} P_b(k) = \sum_{k=0}^{n} \frac{\mu_b^k}{k!}\, e^{-\mu_b}
\end{aligned}$$

Lower and upper limits μ_a and μ_b, using a relation between the Poisson and the χ^2 distribution:

$$\mu_a = \frac{1}{2} F_{\chi^2}^{-1}(\alpha; n_d = 2n) \qquad \mu_b = \frac{1}{2} F_{\chi^2}^{-1}(1-\beta; n_d = 2(n+1))$$

Neyman construction

. . . for repeatable experiments based on the limiting frequency.

Basis: $p(x; a)$ = probability density of observing $\boldsymbol{x}$ if the true value of the parameter is a.

Classical construction first proposed by J. Neyman[a].

- Assume fixed value of parameter a and probability ε
- Construct from fixed value of a the limits $x_1(a)$ and $x_2(a)$ with

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} p(x; a)\, \mathrm{d}x = 1 - \varepsilon$$

  Repeated experiments would produce results x in $x_1 < x < x_2$ with frequency $1 - \varepsilon$.

- Repeat for many assumed values of the parameter a and draw curves $x_1(a)$ and $x_2(a)$ as a function of a.

[a] J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Phil. Trans. Royal Soc. London, Series A, **236** (1937) 333

Bayesian estimators

Knowledge about a parameter is summarized by a probability density function (pdf), which expresses the degree of belief for the parameter to take on a given value.

$$\begin{aligned}
\mu &= \text{parameter with true value } \mu_t \\
x &= \text{observable; } x_0 \text{ is a measurement of } x \\
P(x|\mu) &= \text{pdf for obtaining the value } x \\
\mathcal{L}(x_0|\mu) &= P(x_0|\mu) = \text{ likelihood function} \\
P(\mu_t) &= \text{"prior" pdf for the true value} \\
P(\mu_t|x_0) &= \text{"posterior" pdf for the true value } \mu_t
\end{aligned}$$

$P(\mu_t|x_0)$ is the conditional pdf for the unknown μ_t, *given* the result x_0 of the measurement, and represents *the degree of belief* about μ_t after the measurement, and is obtained by applying Bayes' theorem to pdfs for μ_t:

$$\boxed{P(\mu_t|x_0) = \mathcal{L}(x_0|\mu)\, P(\mu_t)/P(x_0)}$$

Prior and posterior probability

$$\begin{aligned}
\text{prior pdf} &= \text{state of knowledge before consideration of the data} \\
\text{posterior pdf} &= \text{state of knowledge given the data}
\end{aligned}$$

What to use for the prior? How to represent prior ignorance?

Naive choice is a uniform prior, constant in the physically allowed region e.g. restricted to $[0, \infty]$ ("non-informative" prior).

$$\text{Uniform in } \mu_t \text{ or in } 1/\mu_t \text{ or in } \ln \mu_t \text{ ?}$$

A prior uniform in μ_t transforms a likelihood function (function, not density) into a probability density about μ_t, which is non-zero only in the physically allowed region.

The Bayesian estimator is in general not invariant under a transformation of the parameter.

Bayesian intervals

A Bayesian interval $[\mu_1, \mu_2]$ (or "credible interval") corresponding to a certain confidence level can be constructed from the posterior pdf by

$$\int_{\mu_1}^{\mu_2} P(\mu_t|x_0)\,\mathrm{d}\mu_t = 1 - \varepsilon$$

For *making decision* the Bayesian description of inference is probably the way many scientists think: one person's subjective prior incorporates all personal beliefs as well as the results from previous experiments.

Poisson with background – Karmen

Karmen 2 experiment (1998): search for neutrino oscillation signal. For a signal like that reported by the LSND experiment 1 – 1.5 signal events were expected.

Result of Karmen 2:

expected background	$b = 2.88 \pm 0.13$
observed	$n = 0$

Can the effect claimed by LSND be excluded?

Poisson with background

Number n of observed events includes signal events with *unkwown* mean μ and background events with *known* mean B.

The number k of events observed in an experiment follows the Poisson distribution with mean $\mu + b$

$$P_\mu(k) = (\mu + b)^k \; \frac{\exp(-(\mu + b))}{k!}$$

Two problems:

- If the expected is for example $b = 3.0$ and no events are observed, the confidence intervall may be empty.
- If the observed number n is small, and if the data are used to decide either for an upper limit (one-sided) or a confidence interval (two-sided), then the resulting set can undercover.

Proposal to solve the problems by an ordering principle:

Gary J. Feldmann and Robert D. Cousins, A unified approach to the classical statistical analysis of small signals, HUTP-97/A096

Ordering principle based on likelihood ratios

Example: Poisson with *known* background $b = 3.0$

Assume in a Neyman construction a particular value of μ: $\mu = 0.5$

	$P_{0.5}(k)$	μ_{best}	$P_{\mu_{\text{best}}}(k)$	ratio R	rank
$k = 0$	0.030	0	0.050	0.607	6
$k = 1$	0.106	0	0.149	0.708	5
$k = 2$	0.185	0	0.224	0.826	3
$k = 3$	0.216	0	0.224	0.963	2
$k = 4$	0.189	1	0.195	0.966	1
$k = 5$	0.132	2	0.175	0.753	4
$k = 6$	0.077	3	0.161	0.480	7
$k = 7$	0.039	4	0.149	0.259	
...					

Value $k = 0$ has $P(0) = 0.030$, but what is the value of μ which maximizes $P_\mu(0)$? This is called μ_{best} and is given by $\mu_{\text{best}} = \max(0, k - b)$.

Values k are added to the acceptance region in *decreasing* order of the ratio R, until the sum meets the desired C.L. (here $k = 0 \ldots 6$ with sum 0.935).

Gaussian with boundary at origin

Assume only non-negative values for parameter μ are allowed (e.g. μ is a mass).

Observable x is the measured value of μ in an experiment with Gaussian resolution function with fixed standard deviation σ:

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

The same two problems as in the Poisson case can happen for small or negative x/σ) (undercoverage and empty confidence interval).

Ordering principle for Gaussian with boundary at origin

Assume μ physically bounded to non-negative values.

Let μ_{best} be the physically allowed value for which $P(x;\mu)$ is maximum. Then $\mu_{\text{best}} = max(0,x)$, and

$$P(x;\mu_{\text{best}}) = \begin{cases} 1/(\sqrt{2\pi}\sigma) & \text{if } x \geq 0 \\ 1/(\sqrt{2\pi}\sigma)\exp(-x^2/(2\sigma^2)) & \text{if } x < 0 \end{cases}$$

Neyman construction of confidence intervals with μ from grid with narrow spacing: the ratio R, given by

$$R(x) = \frac{P(x;\mu)}{P(x;\mu_{\text{best}})}$$

determines the order in which value of x are added to the acceptance region at a particular value of μ.

Goodness-of-fit and confidence interval

Standard classical analysis of a parameter measurement:

Goodness-of-fit C.L. Test reliability of the experiment: Parameter obtained by a poor fit? Measured value far in the unphysical region?

Confidence interval C.L. What is the confidence interval of the parameter at a certain confidence level? (e.g. determined from the curvature of a likelihood function)

These two questions may be coupled!

An advantage of the proposed ordering is a decoupling of the two questions.

Summary

There is no clear and unique solution to the problem of Confidence Limits in practice.

See opinions and discussion in:
Workshop on Confidence Limits, CERN 17-18 January 2000, ed. F. James, L. Lyons, Y. Perrin, CERN 2000-005 (yellow report)

"You see, a question has arisen,
about which we cannot come to an agreement,
probably because we have read too many books"
(Brecht: Leben des Galilei)