Class 0x0D: Confidence regions

What is a confidence region?

Given a parameter or parameters fit to data according to a model:

- The *confidence region* is an interval or area around the best fit point which has a certain probability of containing the true value, assuming the model is correct.
- The probability that the confidence region will include (or "cover") the true value is called the *coverage*. Different authors may also call it the *coverage level* or the *confidence level* or the *confidence limit* or simply *CL*.
- A confidence region in a single dimension is also called a *confidence interval*.
- A confidence region is determined from the data according to some procedure.

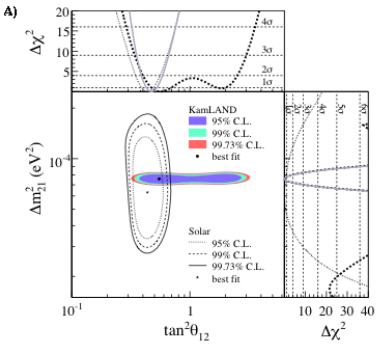
Examples

Upper limit: "The hard disk failure rate is less than 0.01/year (95%CL)."

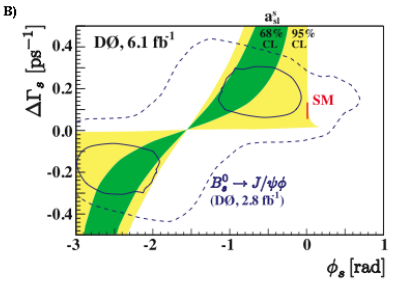
Lower limit: "The expected probability of failure is greater than 0.01/mission (95%CL)."

Two-sided limit: "The allowed 1σ (3σ) CL is $\theta_{12} = 34.4 \pm 1.0 \left(\begin{array}{c} +3.2 \\ -2.9 \end{array} \right)^\circ$."

Multi-dimensional confidence region:



Allowed region for neutrino oscillation parameters from KamLAND and solar neutrino experiments. The side-panels show the $\Delta\chi^2$-profiles for KamLAND (dashed) and solar experiments (dotted) individually, as well as the combination of the two (solid).



The 68% and 95% C.L. regions of probability for $\Delta\Gamma_s$ and ϕ_s values obtained from this measurement, considering the experimental constraints on a^d_{sl} [23]. The solid and dashed curves show respectively the 68% and 95% C.L. contours from the $B^0_s \to J/\psi\phi$ measurement [25]. Also shown is the standard model (SM) prediction for ϕ_s and $\Delta\Gamma_s$.

Figure 1: Two-parameter confidence regions for fitted parameters of neutrino oscillations (A) and CP violation (B), taken from [KamLAND2008] and [DZero2010], respectively.

General features of confidence regions

- The endpoints of a confidence interval (or the boundaries of a region) are determined by data, and therefore are random variables.
- The p.d.f.s of the data in the model hypothesis together with the procedure used to choose the region determine the coverage level.
- In general, confidence regions for parameter(s) θ are defined by

$$P(\theta \text{ in region}) \geq 1 - \alpha$$

  where $1 - \alpha$ is the confidence level. [Larson]
- An okay procedure for can have a CL that depends on the true values of the unknown parameter(s) of the model, but still meets the inequality above. A very clever procedure can have a CL that is equal to $1 - \alpha$ regardless of the true model parameter(s). One such procedure is due to Neyman, described in [PDG-Stat].
- There are shortcuts for gaussian statistics, and some texts only describe those, but I'll keep it more general.

Desirable features

- A sensible confidence region should ...
  - contain the best fit point.

  - generally have points inside more consistant the points outside.
  - have a known coverage level.
  - be consistent and efficient, contracting around the true value as more data is obtained.

- It's possible to come up with procedures that result in regions without these properties, but that is undesirable.
  - Trivial "stupid" example: transform the χ^2 of the data into a uniform random variable u. If it is less than 0.1, define an infinitely thin confidence region; otherwise, define the region to include all possible parameters. This is an interval, and it has 90% coverage -- but it's not the kind of 90% CL interval we're looking for.

Illustration of confidence region coverage

Consider an estimator $\hat{\theta}$ for some constant θ. Suppose our model tells us that the $\hat{\theta}$ is a gaussian random variable with mean θ and standard deviation $\sqrt{\theta}$.

What is the difference between the rms of the estimator and the region with 67% coverage?

(Draw picture.)

Confidence intervals are not unique

The following intervals could all have the same coverage and contain the best fit:

- $-\infty < \theta \leq \theta_u$
- $\theta_1 \leq \theta \leq \theta_2$
- $\theta_l \leq \theta < \infty$

How do we choose confidence intervals (Neyman construction)

(Following section 32.3.2.1 in [PDG-Stat].)

- Find intervals for each value of the parameter θ such that

$$P(\hat{\theta}_1 < \hat{\theta} < \hat{\theta}_2) = 1 - \alpha = \int_{\hat{\theta}_1}^{\hat{\theta}_2} f(\hat{\theta}; \theta) d\hat{\theta}.$$

  Here $f(\hat{\theta}; \theta)$ is the p.d.f. of the estimator $\hat{\theta}$.

- $\hat{\theta}_1$ and $\hat{\theta}_2$ should depend on θ monotonically. The functions are invertible: $\hat{\theta} = \hat{\theta}_1(\theta)$ implies $\theta_1 = \theta_1(\hat{\theta})$.
- Then (see figure)

$$1 - \alpha = P(\hat{\theta}_1 < \hat{\theta} < \hat{\theta}_2) = P(\theta_1(\hat{\theta}) < \hat{\theta} < \theta_2(\hat{\theta}))$$

Example 1: coin toss

Suppose we toss a coin N times. The coin may or may not be fair. We have an unknown probability p of the coin coming up heads, $1 - p$ tails.

The maxmimum likelihood estimator for p is

$$\hat{p} = \frac{k}{N},$$

where k is the number of times the coin came up heads.

The probability distribution of k is known:

$$P_k = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}.$$

Example 1a: upper limit

We can find the sum

$$P(k \leq k_2) = \sum_{k=k_2}^{N} P_k(k).$$

Use that to define the upper limit according to the Neyman procedure. (The lower limit of the interval is set to 0.)

(...picture...)

Example 1b: lower limit

We can find the sum

$$P(k \geq k_1) = \sum_{k=0}^{k_1} P_k(k).$$

Use that to define the lower limit according to the Neyman procedure. (The lower limit of the interval is set to N.)

(...picture...)

Dangers of one-sided limits

Suppose our procedure (perhaps not consiously decided) were to report a "95% CL" upper limit for p if $\hat{p}$ turned out very small, and a "95% CL" lower limit for p if the $\hat{p}$ is near 1.

- E.g., if we got tails all N times, we'd report $p < p_1$ at "95% CL", and do the opposite if we got heads all N times.

What's wrong with that? Suppose the coin is fair, and $p = 0.5$ really.

- The 1-sided upper limit excludes this 5% of the time.
- The 1-sided lower limit excludes this 5% of the time.

10% of the time, we make report implying that the fair coin is unfair "with 95% CL".

Example 1c: two-sided limit

One common approach: use $P(k \geq k_1) = \alpha/2$ and $P(k \leq k_2) = \alpha/2$ to define the intervals.

The only problem is that sometimes this isn't possible. E.g., if we get tails all N times, $k = 0$, the probability at one end is constrained. Then we have to adjust somehow,

The "unified" method (aka "Feldman-Cousins")

Very similar to the two-sided limit, except we pick the edges of the region to have a given log-likelihood. If the parameter space has a boundary, no problem, just adjust the log-likelihood level to encompass enough space.

In general, the p.d.f. of the log-likelihood is generated via MC for each parameter.

Assignment

Read sections 32.3.2.1 and 32.3.2.2 in [PDG-Stat].

Next Assignment

Build the 90%-CL and 99%-CL confidence regions for the same exponential + background of the assignment from class 11 (aka class 0x0B), with the restriction that the background parameter b must be in the range $0 \leq b < 1$ and the mean μ must be positive.

- Note this is a good example of where the full Feldman-Cousin's treatment is usually necessary: non-Gaussian model, limits on parameters.
- Just follow the Feldman-Cousins procedure.

[KamLAND2008] KamLAND Collaboration, "Precision Measurement of Neutrino Oscillation Parameters with KamLAND", Phys.Rev.Lett.100:221803,2008; arXiv:0801.4589v3 [hep-ex].

[DZero2010] D0 Collaboration, "Evidence for an anomalous like-sign dimuon charge asymmetry", Submitted to Phys. Rev. D, 2010; Fermilab-Pub-10/114-E; arXiv:1005.2757v1 [hep-ex].

[PDG-Stat] "Statistics", G. Cowan, in *Review of Particle Physics*, C. Amsler et al., PL B667, 1 (2008) and 2009 partial update for the 2010 edition (http://pdg.lbl.gov/2009/reviews/rpp2009-rev-statistics.pdf).