Computing Confidence Limits for Small Datasets

Or

How to Calculate Error Bars When You Only Have a Few Observations

Tim Haines

py-astro-stat Discussion

June 24, 2014

Representing Measurements

- How do we make a measurement meaningful?
  - Units
  - Error Bars
  - Comparison To Models
- We are going to focus on the last two.
- But first...

The Central Limit Theorem

- As the number of observations becomes sufficiently large, the distribution of the measured values tends to the normal (Gaussian) distribution- independent of the true distribution governing the measured quantity.
- How many observations are “sufficient” to use the normal distribution?
  - There is not a hard-and-fast rule for this, but a rough guide is ~60. I prefer >100.
- How do we know what the “true” distribution is?
  - This requires a careful understanding of the definition of each distribution.

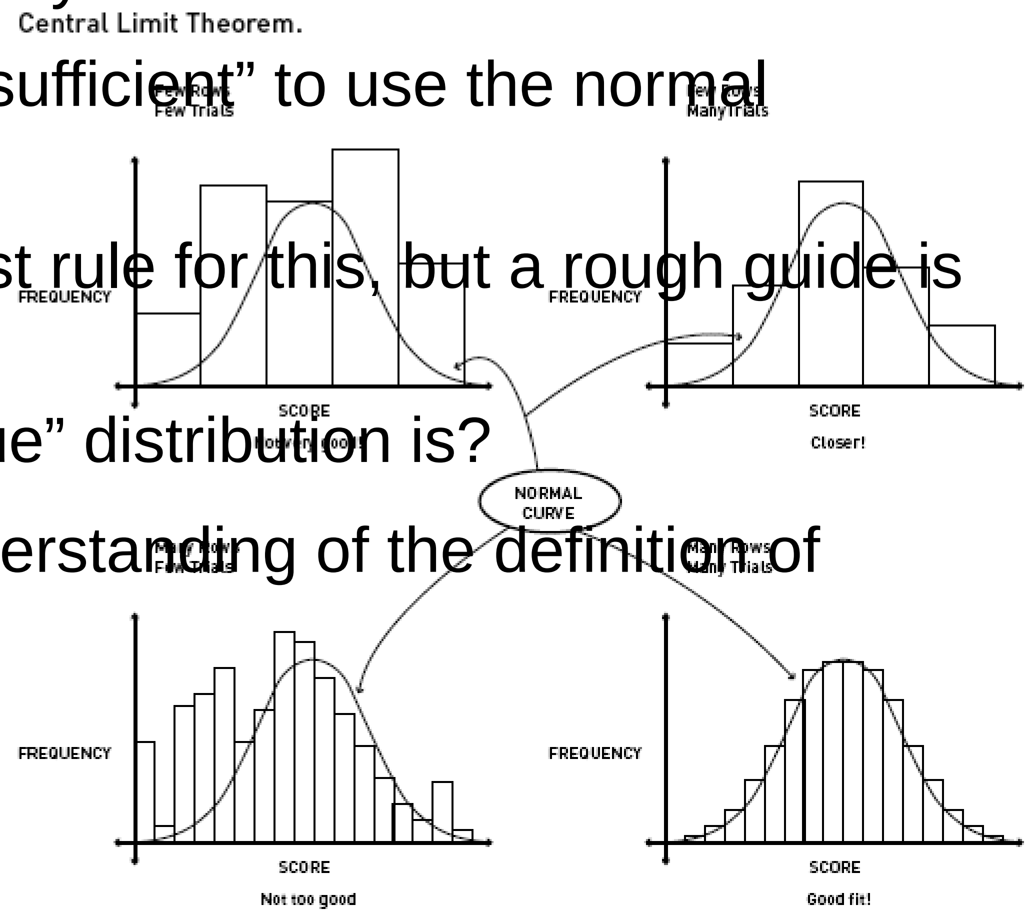
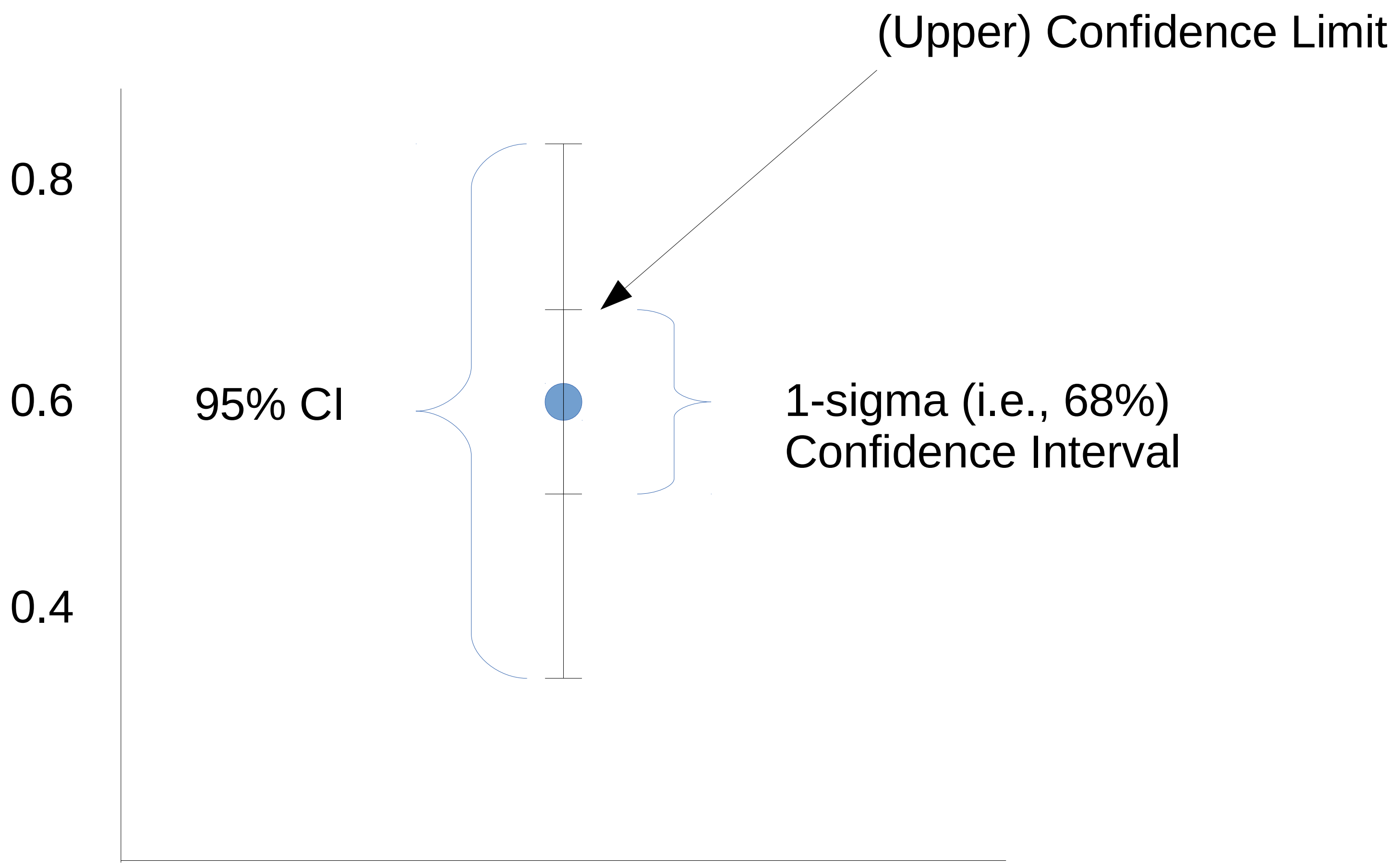


Image © Annenberg Foundation 2013

Some Terminology

- Confidence Level
  - How confident do I want to be in my observation? (68%, 95%, 99.99999999%?)
- Confidence Limit
  - The minimum/maximum value of the measurement at the given confidence level
- Confidence Interval
  - This is the error bar you are trying to put on your measured data point. It lies between the two confidence limits.

Some Terminology (cont.)



A Note On Confidence Levels

- Every distribution has its own rules for computing confidence *limits*, but confidence levels are **always** defined in terms of the Normal distribution.

$$CL(S) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{S} \exp\{-z^2/2\} dz$$

- S is the number of standard deviations desired. The more famous CLs are the 1-sigma level at 84%, the 2-sigma level at 97%, and the 3-sigma level at 99%. Please note these are the one-sided CLs which differ from the two-sided variants you are probably accustomed to (i.e., the 68%, 97%, and 99% levels).

Small Number Statistics

- What do you do when you don't have enough measurements to satisfy the Central Limit Theorem?
- You need to know which distribution governs your measured data. This can be hard. There are dozens of distributions. But we will focus on the two most important: the Poisson and the Binomial.

The Poisson Distribution

- Describes any process having events that are discrete, random, and independent.
- This covers a wide range of processes, but we usually use it to describe stochastic space- or time-rate events.
  - In astronomy, Poisson statistics are often referred to as “counting statistics.”

Binomial Distribution

- Describes performing a measurement as executing a trial of some particular process having only two possible outcomes. Each measurement must be performed in exactly the same way, and the outcomes between any two trials are assumed to be equally likely.
- Formally, it describes “the output of performing N Bernoulli trials.”
- The two possible outcomes are usually referred to as “success” and “failure.”
- The binomial is used when, essentially, the outcome can be determined by the flip of a coin. Specifically, it is used when computing a fraction.

An Example

- We have two hypothetical initial mass functions (IMF; the number of stars at a given mass) which differ only in the predicted fraction of nebulae containing more than 30% A stars.
- The first IMF predicts the fraction of nebulae will never exceed 20%. The second one allows more than 20% of the nebulae to contain more than 30% A stars.
- Let us use the prediction of the first IMF as the null hypothesis.
- We go to the telescope and measure the number of A stars in 50 different nebulae in the Milky Way.
- We find that only 5 (10%) of the nebulae are comprised of more than 30% A stars.

An Example (cont.)

- How do we report our findings?
- We need to compute the confidence limits!
- But which method do we use?
- Let's try them all and see what we get.

An Example (cont.)

- Normal (Gaussian) 1-sigma Confidence Limits

$$\bar{x} \pm \frac{s}{\sqrt{N}}$$

$$0.1 \pm \frac{1}{\sqrt{50}}$$

$$0.1 \pm 0.14$$

Since we can't have negative values, we say that the upper confidence limit is 24.1% and the lower confidence limit is 0%.

An Example (cont.)

- Poisson 1-sigma Confidence Limits

```
CL = scipy.special.ndtr(1.0)  # remember, all confidence levels are
                              # determined from the normal distribution
upper = scipy.special.pdtri(5, 1- CL) / 50
lower = scipy.special.pdtri(5-1,CL) / 50
```

- Scary stuff! But don't worry, Scipy does all the heavy lifting. You just need to remember two important rules:
  - 1. Confidence levels are determined by the normal distribution.
  - 2. The Poisson distribution is based on the average event rate, so you need to divide by the total number of observations to get your confidence limits.
- The upper and lower confidence limits are then 16.8% and 5.7%, respectively.

An Example (cont.)

- Binomial 1-sigma Confidence Limits

```
CL = scipy.special.ndtr(1.0)
upper = scipy.special.bdtri(5,50,1-CL)
lower = 1 - scipy.special.bdtri(50-5,50,1-CL)
```

- The upper and lower confidence limits are then 16.2% and 5.7%, respectively.

An Example (cont.)

	Lower Limit	Upper Limit	Reject H_0?
Normal	0.0%	24.1%	Yes
Poisson	5.7%	16.8%	No
Binomial	5.7%	16.2%	No

- Do we reject H_0? (i.e., can the fraction of nebulae be greater than 20%?)
- Which one is correct?
  - The binomial
- How do we know?
  - We computed a fraction!
- It is clear, then, that our choice of distribution can strongly affect the outcome of our analysis.