Comments on Statistics

Michael S. Briggs
at the Fermi School
2012 June 1

Statistics

Physicists and Astronomers invented much of statistics but in the 20th Century have fallen behind. There is a strong tendancy to take the “hammer” approach:

“I call it the law of the instrument, and it may be formulated as follows: Give a small boy a hammer, and he will find that everything he encounters needs pounding.” (Abraham Kaplan, 1964).

Physicists tend to have a small cookbook of “recipes”, and grab the one that cursorily seems best. *It might be sub-optimum or inapplicable!*

Statistics

What is a "statistic"?

It is a number (or small group of numbers) that summarizes the data.

Ideally the statistic is selected for useful properties, such as its ability to discriminate between hypotheses.

You can create your own statistic!

Statistical Tasks

There are basically two tasks for statistics:
hypothesis testing (comparison)
parameter estimation, including errors

Not: "goodness of fit", beloved of scientists, is not liked by statisticians.

Hypothesis Testing

Done by comparing two hypothesis, an alternative model H_1 that you wish to "prove" and the null hypothesis H_0. Conventional statistics uses a backwards approach: the alternative model H_1 is demonstrated by rejecting the null hypothesis H_0.

Limitation: You might show that H_1 is preferred over H_0, but actually H_2 that you never considered is much better than either!

Calculate the statistic. If you find that the value is highly improbable under H_0 you can reject H_0 and you now prefer the alternative. Otherwise you stick with the null hypothesis.

There are two possible errors: "demonstrating" H_1 when that hypothesis is false, or sticking to H_0 when H_1 is actually true, We guard against the first error by requiring a high confidence level (i.e., small chance probability) but concomitantly increase the probability of the second error.

Likelihood and Likelihood Ratio Tests

$$L = \prod P \tag{1}$$

$-2 \log L$ is frequently more convenient. The products become sums. If the probabilities are Gaussian, χ^2 is obtained.

Model comparison statistic:

$$\Lambda = \frac{L(\theta_0)}{\max L(\theta)} \tag{2}$$

Wilk's Theorem – 1

From Wikipedia:

"A convenient result, attributed to Samuel S. Wilks, says that as the sample size approaches ∞, the test statistic for a nested model will be asymptotically distributed with degrees of freedom equal to the difference in dimensionality of $theta_0$ and θ. This means that for a great variety of hypotheses, a practitioner can compute the likelihood ratio Λ for the data and compare $-2 \log L$ to the chi-squared value corresponding to a desired statistical significance as an approximate statistical test.

Wilk's Theorem – 2

That already has some important "ifs"!

But there is another one, brought forward by Protassov et al. in ApJ.:

"the null values of the additional parameters may not be on the boundary of the set of possible parameter values"

What to do?

What to do when the reference distribution is unknown?

Simulate!

Bayes' Theorem

Bayes' Theorem is a simple theorem of probability:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}. \tag{3}$$

No one can dispute Bayes' Theorem. The issue is its usefulness.

In conventional or frequentist statistics, probabilities are considered to be derived from set theory, or from long-run averages of measurements. From these definitions, it is nonsense to speak of the probability that a hypothesis is true. If a broader definition is allowed, Bayes' Theorem becomes much more useful:

Bayesian Inference – 1

$$P(H|D, I) = \frac{P(H|I)P(D|H, I)}{P(D|I)}, \tag{4}$$

where H is a Hypothesis, D is the Data and I is other Information.

Bayes' Theorem allows the "inversion" of the contents of the "P"s. Generally a hypothesis allows us to calculate the probability of observing particular data: $P(D|H, I)$. With Bayes' Theorem, we can convert this into a probability statement about the hypothesis: $P(H|D, I)$.

Bayesian Inference – 2

$$P(H|D,I) = \frac{P(H|I)P(D|H,I)}{P(D|I)}, \tag{5}$$

$P(H|I)$ is called the prior, the probability of the hypothesis before collecting the data.

$P(H|D,I)$ is the posterior probability – the probability updated based upon the data.

$P(D|H,I)$ is the likelihood.

$P(D|I)$ is a normalization factor.

Bayesian Inference – 3

Now if we take the ratio of Bayes's Theorem for two hypotheses the $P(D|I)$ term cancels:

$$\frac{P(H_1|D,I)}{P(H_2|D,I)} = \frac{P(H_1|I)P(D|H_1,I)}{P(H_2|I)P(D|H_2,I)} \tag{6}$$

Ockham factor

Epistemology

binary logic $\Rightarrow$ Bayesian Inference

References – 1

Statistical Methods in Experimental Physics, 1st ed., W. T. Eadie, et al., 1971.
Comprehensive but can only be understood by reading in order.

Statistical Methods in Experimental Physics, rev. 2nd ed., F. James, et al., 2006.
I haven't seen this version.

Statistics, Handle with Care: Detecting Multiple Model Components with the Likelihood Ratio Test, R. Protassov et al., ApJ, **571**, 545, (2002).

References – 2

Bayesian Logical Data Analysis for the Physical Sciences, P. Gregory, Cambridge Univ. Press, 2005.
A text book at the graduate level.

The Promise of Bayesian Inference for Astrophysics, T. Toredo, 1992. available from http://www.astro.cornell.edu/staff/loredo/bayes/tjl.html
Introductory article.

Data Analysis: A Bayesian Tutorial, D. S. Sivia, Oxford, 1996.
Introductory textbook.