Hypothesis Testing

David A. van Dyk[1]

[1]Department of Statistics
University of California, Irvine

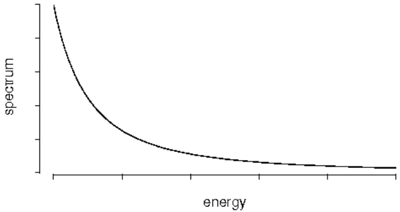
2008 HEAD Meetings

Outline

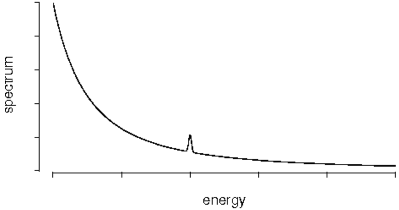
1. Hypothesis Testing
    - Basic Framework
    - Test Statistics
2. Mathematical Computations
    - Asymptotics
    - Assumptions
3. Numerical Computations
    - Monte Carlo
    - Bootstrap and Posterior Predictive P-values

Hypothesis Testing

- The Null Hypothesis
  - H_0: Supposed interesting feature doesn't exist in the data.
- The Alternative Hypothesis
  - H_A: Supposed interesting feature does exist in the data.



H_0: No emission line.



H_1: Emission line.

The null is a special case of the alternative:
Line intensity equals zero.

Test Statistics

Test Statistics are used to measure the evidence for null and alternative hypotheses.

Assuming the null hypothesis is true, how likely are we to see a value of the test statistics as extreme or more extreme than the observed value?

1. The distribution of the the *test statistic* must be known under the null hypotheis.
2. The *test statistic* must behavior differently under the alternative hypothesis.
3. For example, large value of the *test statistic* may give evidence for the alternative and agaist the null hypothesis.

How large must the test statistic be?

P-values

Assuming the null hypothesis is true, how likely are we to see a value of the test statistics as extreme or more extreme than the observed value?

$$\Pr(T \geq t_{\text{obs}} | H_0) = \text{p-value}$$

Unfortunately, these probability calculation are intractable in all but the simplest situations.

Solution: "Large sample" approximations.

Likelihood Ratio Test Statistics

$$R = \frac{\sup_{\theta \in \Theta_0} L(\theta|Y)}{\sup_{\theta \in \Theta} L(\theta|Y)},$$

where

1. Θ is the parameter space under the alternative (dim $= d$).
2. $\Theta_0 \in \Theta$ is the parameter space under the null (dim $= d_0$).
3. L is the Likelihood

Fit model with and without the line and compare the best fits.

Under certain assumptions, the distribution of $-2\log(R)$ under H_0 approaches $\chi^2_{(d-d_0)}$ as the sample size (or counts) increases.

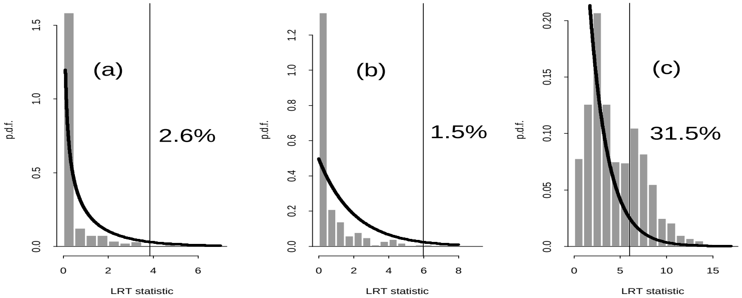
BUT... Assumptions include:

1. The null hypothesis must be a special case of the alternative hypothesis: $\Theta_0 \in \Theta$.
2. The null hypothesis must be in the interior of the alternative hypothesis, more precisely Θ_0 must be in the interior of Θ.

The second assumption fails when testing for a spectral line:

1. When there is no line, the line intensity is zero, it may not be negative.
2. Further, the the location and width of the line *do not exist* when there is no line. They have *no* values.

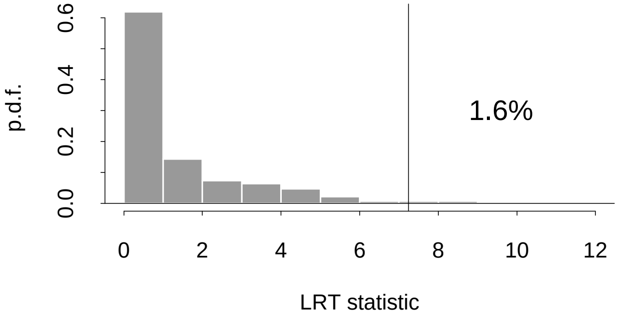
The F-test is similarly inappropriate for testing for a line.



- The actual distribution of the LRT statistic (histogram) is compared with its nominal distribution (line).
- Three cases: fitting a narrow line (fixed location), fitting a wide line (fit location), testing for an absorption line.
- The nominal cut off for 5% false positives is shown along with the simulated false positive rates.

Monte Carlo Calibration

1. We do not know the true (sampling) distribution of the *test statistic*.
2. We can evaluate the distribution numerically using Monte Carlo simulation.
3. Simulate L data sets under H_0 and compute the *test statistic* for each of the L data sets.
4. A histogram of the simulated test statistics approximates the sampling distribution of the test statistic.



Computing the p-value:

$\Pr(T \geq t_{\rm obs}|H_0) =$ the proportion of simulated test statistics larger than $t_{\rm obs}$.

Bootstrap and Bayesian Posterior Predictive Sampling

A complication: If there are unknown parameters in null the model, we can not directly simulate data.

Solutions:

1. Fit the real data under the null model. Compute fitted parameters and error bars.
2. Parametric Bootstrap suggests resampling data sets with unknown parameters set accounting for these error bars.
3. Bayesian Posterior Predictive modeling simulates unknown parameters from their posterior distribution, which are in turn used to simulate data sets.

For Further Reading I

van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2001).
Analysis of energy spectra with low photon counts via Bayesian posterior simulation.
The Astrophysical Journal **548**, 224–243.

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2002).
Statistics: Handle with care – detecting multiple model components with the likelihood ratio test.
The Astrophysical Journal **571**, 545–559.

van Dyk, D. A. and Kang, H. (2004).
Highly structured models for spectral analysis in high energy astrophysics.
Statistical Science **19**, 275–293.

van Dyk, D. A., Connors, A., Esch, D. N., Freeman, P., Kang, H., Karovska, M., and Kashyap, V. (2006).
Deconvolution in High Energy Astrophysics: Science, Instrumentation, and Methods (with discussion).
Bayesian Analysis, **1**, 189–236.