

## Testing Nested Models

F. Masci, March 3, 2006

Two models are considered “nested” if one is a subset or an extension of the other. For example, the two models  $f_1(x) = mx + c$  and  $f_2(x) = ax^2 + mx + c$  are nested. A question often asked is: how do we decide if a more complex model is “better” at predicting the data than a simpler one? If the more complex model does not significantly contribute more information, then there’s no reason to accept the parameters estimated there from. The simpler, “more parsimonious” model is preferred on all grounds.

The problem boils down to testing if the residual sum of squares from fitting a simple model (or  $\chi^2$  if the squared residuals are weighted by their inverse variance priors) is significantly larger than that of the more complex model. Let’s label the *minimum* chi-square values (after parameter estimation) for the simple and complex models by  $\chi_s^2$  and  $\chi_c^2$  respectively. The only difference is in the assumed model  $f_s(x,p)$  or  $f_c(x,q)$ , with number of parameters  $p, q$  respectively where  $p < q$ .

$$\chi_s^2 = \sum_{i=1}^N \frac{[y_i - f_s(x_i, p)]^2}{\sigma_i^2}$$
$$\chi_c^2 = \sum_{i=1}^N \frac{[y_i - f_c(x_i, q)]^2}{\sigma_i^2}$$

Under the null hypothesis that each model “generates” the data, these quantities will follow chi-square distributions with  $N - p$  and  $N - q$  degrees of freedom respectively for fits to repeated realizations of the data with errors drawn from  $N(0, \sigma_i)$ . This can be used to check for model plausibility given some maximum tolerable probability  $\alpha$  (e.g., 0.05) below which an experimenter declares it’s unlikely the model could have generated the data. I.e., the chance that it did is  $< \alpha$ . However, in the spirit of Occam’s Razor, can we do better when given two competing (nested) models?

R.A. Fisher (1938) worked out the distribution of the relative difference between  $\chi_s^2$  and  $\chi_c^2$  under the null hypothesis that the simpler (restricted) model is correct, which is equivalent to the statement that some parameters in the complex model are zero. This distribution assumes the fit residuals are normally distributed or approximately so. We construct a test statistic known as the “*F*-ratio” (where I presume *F* stands for Fisher):

$$F = \frac{(\chi_s^2 - \chi_c^2)/(p - q)}{\chi_c^2/(N - q)}$$
$$= \frac{N - q}{p - q} \left( \frac{\chi_s^2}{\chi_c^2} - 1 \right).$$

If the simpler model is *correct*, this will follow an *F*-distribution with degrees of freedom  $(\nu_1, \nu_2) = (p - q, N - q)$ . We reject the simpler model if *F* satisfies  $1 - \text{CDF}(F) < \alpha$  where

CDF is the cumulative ( $F$ -)distribution function evaluated at  $F$  and  $\alpha$  is some maximum tolerable probability (e.g., 0.05) below which it's unlikely the simple model generated the data. If so, the more complex model is favored. In practice, it's faster to have a pre-computed look-up-table of critical values  $F_{crit}$  corresponding to different combinations of  $(\alpha, \nu_1, \nu_2)$ . For a given  $F$ , one then rejects the null hypothesis if  $F > F_{crit}$ .

Note that if the models are not nested, one cannot use the  $F$ -ratio or any likelihood-ratio test for that matter. In this case, one may use Akaike's Information Criterion (AIC) after testing for plausibility of each model using e.g., a chi-square (for normally distributed errors) or some generalized likelihood method.

## Further reading

<http://en.wikipedia.org/wiki/F-test>

<http://en.wikipedia.org/wiki/F-distribution>

[http://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](http://en.wikipedia.org/wiki/Akaike_information_criterion)

[http://www.indiana.edu/~clcl/Q550/Papers/Zucchini\\_JMP\\_2000.pdf](http://www.indiana.edu/~clcl/Q550/Papers/Zucchini_JMP_2000.pdf)

<http://www.ime.unicamp.br/~lramos/mi667/ref/19kadane04.pdf>

<http://vserver1.cscs.lsa.umich.edu/~crshalizi/notabene/model-selection.html>

<http://www.cis.upenn.edu/~mkearns/papers/ms.pdf>