Model Choice

Say measure $(Y, X_1, \ldots, X_m)$ for n subjects.

Data aquisition may have been "fishing expedition" - don't necessarily believe all m predictors are relevant.

May seek a final model including only m^* of the predictors (or equivalently, set $m - m^*$ of the regression coefficients to zero).

First, a more focussed question. How to compare *two* models.

Likelihood Ratio Test (LRT)

Data D, Model M_0 (p_0 params) nested within M_1 (p_1 params).

If M_0 true,

$$2\left\{l_1(\hat{\theta}_1; D) - l_0(\hat{\theta}_0; D)\right\} \quad \overset{approx}{\sim} \quad \chi^2_{p_1 - p_0}$$

Usual hypothesis testing implementation and interpretation.

An Information Criterion (AIC)

No requirement that competing models be nested.

Choose the one maximizing

$$l_i(\hat{\theta}_i; D) - p_i,$$

i.e., notion of *complexity penalty*.

Motivated as a measure of *predictive performance*.

Bayesian Information Criterion (BIC)

Again no nesting requirement.

Choose the model maximizing

$$l_i(\hat{\theta}_i; D) - \{(1/2)\log n\}p_i,$$

i.e., bigger complexity penalty than AIC, especially for large samples.

Rationale: Bayesian - somewhat crude approximation to choosing the model for which Pr(Model i is true|Data=D) is largest.

Practical Difference - YES

Say comparing M_0 and M_1, nested, with $p_1 = p_0 + 1$. Choose M_1 if $2\{l_1(\hat{\theta}_1; D) - l_0(\hat{\theta}_0; D)\} > c$.

LRT: $c = \chi_1^2$ quantile, i.e.,

$$c = \begin{cases} 2.71 & 10\% \text{ sig.}, \\ 3.84 & 5\% \text{ sig.}, \\ 6.63 & 1\% \text{ sig.} \end{cases}$$

AIC: $c = 2$.

BIC: $c = \log n$, i.e.,

$$c = \begin{cases} 4.6 & \text{if } n = 100, \\ 6.2 & \text{if } n = 500, \\ 6.9 & \text{if } n = 1000. \end{cases}$$

Comparing all possible models

?IC can compare any collection of models.

There are 2^m subsets of m predictor variables.

Fitting 2^{10} models may be tolerable.

Fitting 2^{20} models may not be.

Situation worse if want to consider possibilities of ‘curved’ effects and/or interactions.

e.g. interactions: m physical variables, but $m + m(m-1)/2$ possible predictors for inclusion/exclusion.

Motivates **stepwise** procedures. Search for models with high values of criterion function without evaluating all possible models.

stepAIC() in R (part of MASS library)

Iterative scheme.

From current model, consider all possible ‘one-term deletions‘ (backward) AND/OR ‘one-term additions (forward).’

Of these, the new model is the one with the best improvement in AIC (or BIC).

Iterate this scheme until no such changes improve AIC.

Practical, but no guarantee of global max.

Some Comments

Developing ?IC's is an active research area: recently DIC, FIC.

Lack of agreement between methods more pronounced for model choice than for within-model estimation. Reflection of an inherently harder problem?

Partial insight: Criterion function over $\{0,1\}^m$ less well-behaved than within-model likelihood function. Related idea that model selection schemes can be unstable (recall boostrapping ex.).

What about removing the role of 'theory' and comparing competing models on a more 'empirical' basis...