SYSC 551 Discrete Multivariate Modeling
Prof. Martin Zwick
Portland State University
Notes from 2/09 /2012
As taken by Juliana Arrighi

Information-Theoretic Reconstructability Analysis, part 2
(Putting it all together)

5. Choosing Models Statistically

Reference model is top (i.e. saturated)
We want to pick a model with small error, so we try not to go down the lattice too far, but we also want complexity to be small so we try to go down the lattice as far as possible. We want the biggest Δdf as long as L^2 is small

Null hypothesis: $m_j = m_o$ or $q(m_j) \approx p(m_o)$
$L^2(m_j) = 1.3863 \text{ n } I(m_o \rightarrow m_j) = 1.3863 \text{ n } T(m_j)$

In this case we do not want to reject the null hypothesis, so we want a high p-value.

If we make a Type I error (incorrectly reject null hyp.), we end up with a model that is more complex than necessary.

If we make a Type II error (incorrectly accept null hyp.), we accept a model that is too simpler than is statistically justified.

Type I error is preferable.

Reference model is bottom (e.g. independence)
We want biggest L^2 as long as Δdf is small
Null hypothesis: $m_j = m_{ind}$
We want to reject this null hypothesis

If we make a Type I error (incorrectly reject null hyp.), we are positing constraints that are not statistically justified.

If we make a Type II error (incorrectly accept null hyp.) we missed detecting a real constraint.

Generally Type II error is preferable here.

6. Choosing Models Based on Information Content

Models can be chosen by information content only. For example, if we wanted to search for models that had at least 90% information captured, we would not need to calculate L^2.

OCCAM gives normalized information (I)

$$I_{OCCAM}(m_j \rightarrow m_k) = \frac{I(m_j \rightarrow m_k)}{T(m_{ind})}$$

7. Choosing Models With AIC and BIC

$$\begin{aligned} AIC &= -2n\sum p\ln q + 2df \\ &= -2n\sum p\ln q + 2df + 2n\sum p\ln p - 2n\sum p\ln p \\ &= L^2 - 2n\sum p\ln p + 2df \end{aligned}$$

$$\begin{aligned} BIC &= -2n\sum p\ln q + \ln(n)df \\ &= -2n\sum p\ln q + \ln(n)df + 2n\sum p\ln p - 2n\sum p\ln p \\ &= L^2 - 2n\sum p\ln p + \ln(n)df \end{aligned}$$

n = sample size

When comparing between models for the same data, -Σplogp is a constant because it is based on the data, not the model, and it is not included in the AIC or BIC calculation

-Σplogq (L^2 -Σplogp) is a measure of error. 2df and ln(n)df are measures of complexity. We want both error and complexity to be low, so we want to minimize AIC.

BIC weights the complexity term higher (ln(n) is larger than 2 for any reasonable n).

OCCAM give ΔAIC and ΔBIC where $\Delta AIC = AIC_{Reference} - AIC_{Model}$. We want to maximize this.