Information Criterions

- We define two types of information criterion: the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC).
- In AIC and BIC, we choose the model that has the minimum value.
- *(Remark, a difference of less than two is considered negligible)*.

Information Criterions

Formulae

$$AIC = 2\log(L) + 2p,$$

$$BIC = 2\log(L) + p\log n$$

where

- L is the likelihood of the data with a certain model,
- n is the number of observations and
- p is the number of parameters in the model.

Information Criterions

However, which criterion should we choose?

- AIC is well-known for overestimating the correct number of segments
- BIC has a slight tendency to underestimate this number.
- However analyses usually renders concordant results.

AIC

- The Akaike information criterion is a measure of the relative **goodness of fit** of a statistical model.
- When using the AIC for selecting the parametric model class, choose the model for which the AIC value is lowest.

Akaike Information Criterion (AIC)

- Akaike's information criterion is a measure of the goodness of fit of an estimated statistical model.
- The AIC is a **model selection** tool, i.e. a method of comparing two or more candidate regression models.
- The AIC methodology attempts to find the model that best explains the data with a minimum of parameters. (i.e. in keeping with the law of parsimony)

Akaike Information Criterion

Akaike Information Criterion (AIC)

- The AIC is calculated using the "likelihood function" and the number of parameters
- Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

Akaike Information Criterion (AIC)

$$\mathrm{AIC} = 2p - 2\ln(L)$$

- p is the number of free model parameters.
- L is the value of the Likelihood function for the model in question.
- For AIC to be optimal, n must be large compared to p.

Schwarz's Bayesian Information Criterion

An alternative to the AIC is the Schwarz BIC, which additionally takes into account the sample size n.

$$\mathrm{BIC} = p \ln n - 2 \ln(L)$$

- These are relative measures of goodness-of-fit and are used to compare different solutions with different numbers of segments.
- ("Relative" means that these criteria are not scaled on a range of, for example, 0 to 1 but can generally take any value.)
- **Important**: Compared to an alternative solution with a different number of segments, smaller values in AIC or BIC indicate an increased fit.