The Akaike Information Criterion

The Akaike Information Criterion is one of a range of ways of choosing between different types of models that seek an appropriate trade-off between goodness of fit and model complexity. The more complicated a model is the better generally will be its apparent goodness of fit, if the parameters are selected to optimise goodness of fit, but this does not necessarily make it a 'better' model overall for identifying how new data might behave.

A simple example of this is that if we have n datapoints, i.e. (x_i, y_i) for $i = 1, \dots, n$, relating to some unknown function then we can exactly fit all of these points with a polynomial of order $n-1$, i.e. $y(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$ where the a_i are fixed, but a smoother polynomial with lower order or some other function with few parameters may actually be a better guide as to what the value of y might be for the $n+1$'th datapoint even though it is unlikely to fit the first n points as well as the exact fit polynomial of order $n-1$.

As explained in e.g. Billah, Hyndman and Koehler (2003), a common way of handling this trade-off in the context of statistics is to choose the model (out of say N model types, each of which is characterised by a vector of the q unknown free parameters where q varies between the different model types) that provides the highest 'information criterion' of the form:

$$IC = \log L(\hat{\theta}) - f(n, q)$$

where $L(\hat{\theta})$ is the maximised log-likelihood function, θ is the vector of the q unknown free parameters within the relevant model, $f(n, q)$ is a penalty function that penalises more complex models and we are using a data series of length n for fitting purposes.

A range of information criteria have been proposed for this purpose including:

Criterion	**Penalty function**
AIC (Akaike's Information Criterion)	q
BIC (Bayesian Information Criterion)	$q \log(n)/2$
HQ (Hannan & Quinn's Criterion)	$q \log(\log(n))$
MCp (Mallow's Criterion)	$n \log(1 + 2q/r)/2$
GCV (Generalized Cross Validation Criterion)	$-n \log(1 - q/n)$
FPE (Finite Prediction Error Criterion)	$(n \log(n + q) - n \log(n - q))$

where (for MCp) $r = n - q^*$ and q^* is the number of free parameters in the smallest model that nests all models under consideration. Billah, Hyndman and Koehler's innovation is seek to estimate an 'ideal' $f(n, q)$ for the purpose in hand, thus deriving an 'empirical' information criterion rather than necessarily adopting a fixed penalty functional form.