

# Akaike information criterion

From Wikipedia, the free encyclopedia

**Akaike's information criterion**, developed by Hirotugu Akaike under the name of "an information criterion" (**AIC**) in 1971 and proposed in *Akaike (1974)*,<sup>[1]</sup> is a measure of the goodness of fit of an estimated statistical model. It is grounded in the concept of entropy, in effect offering a relative measure of the information lost when a given model is used to describe reality and can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking that of accuracy and complexity of the model.

The AIC is not a test of the model in the sense of hypothesis testing; rather it is a test between models - a tool for model selection. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. From the AIC value one may infer that e.g. the top three models are in a tie and the rest are far worse, but it would be arbitrary to assign a value above which a given model is 'rejected'.<sup>[2]</sup>

## Contents

- 1 Definition
- 2 Relevance to  $\chi^2$  fitting (maximum likelihood)
- 3 AICc and AICu
- 4 QAIC
- 5 See also
- 6 Notes
- 7 References
- 8 External links

## Definition

In the general case, the AIC is<sup>[citation needed]</sup>

$$AIC = 2k - 2 \ln(L)$$

where  $k$  is the number of parameters in the statistical model, and  $L$  is the maximized value of the likelihood function for the estimated model.

From a Bayesian perspective, the AIC corresponds to adopting a prior  $P(k) \propto \exp(-k)$  to penalise models with large numbers of parameters  $k$ .<sup>[citation needed]</sup> The posterior is then

$$P(k|data) \propto P(data|k)P(k) = \exp(-AIC/2).$$

Over the remainder of this entry, it will be assumed that the model errors are normally and independently distributed. Let  $n$  be the number of observations and  $RSS$  be

$$RSS = \sum_{i=1}^n \hat{\epsilon}_i^2,$$

the residual sum of squares. We further assume that the variance of the model errors is unknown but equal for them all. Maximizing the likelihood with respect to this variance, the AIC becomes

$$AIC = 2k + n[\ln(2\pi RSS/n) + 1].$$

This can be simplified by factoring out the term  $n * \ln(2\pi)$ . This is a constant term added to the AIC value of all the competing models. Therefore it can't affect the order in which we rank them and we can safely remove this term. When we also factor out the constant  $n$ , AIC simplifies to:

$$AIC = 2k + n[\ln(RSS)].$$

Increasing the number of free parameters to be estimated improves the goodness of fit, regardless of the number of free parameters in the data generating process. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. *The preferred model is the one with the lowest AIC value.* The AIC methodology attempts to find the model that **best explains the data with a minimum of free parameters**. By contrast, more traditional approaches to modeling start from a null hypothesis. The AIC penalizes free parameters less strongly than does the Schwarz criterion.

AIC judges a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. But it is important to realize that the AIC value assigned to a model is only meant to *rank* competing models and tell you which is the best among the given alternatives. The absolute values of the AIC for different models have no meaning; only relative differences can be ascribed meaning.

## Relevance to $\chi^2$ fitting (maximum likelihood)

Often, one wishes to select amongst competing models where the likelihood function assumes that the underlying errors are normally distributed. This assumption leads to  $\chi^2$  data fitting.

For any set of models where the data points are used, one can use a slightly altered AIC. For the purposes of this article, this will be called  $AIC_{\chi^2}$ . It differs from the AIC only through an additive constant, which is a function only of the data points and not of the model. As only differences in the AIC are relevant, this constant can be ignored.

For  $\chi^2$  fitting, the likelihood is given by

$$L = \prod_{i=1}^n \left( \frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left( - \sum_{i=1}^n \frac{(y_i - f(\mathbf{x}))^2}{2\sigma_i^2} \right)$$

$$\therefore \ln L = \ln \left( \prod_{i=1}^n \left( \frac{1}{2\pi\sigma_i^2} \right)^{1/2} \right) - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(\mathbf{x}))^2}{\sigma_i^2}$$

$\therefore \ln L = C - \chi^2/2$ , where C is a constant independent of the model used, and dependent only on the use of particular data points. i.e. it does not change if the data do not change.

The AIC is therefore given by  $AIC = 2k - 2 \ln(L) = 2k - 2(C - \chi^2/2) = 2k - 2C + \chi^2$ . As only differences in AICc are meaningful, the constant C can be omitted provided the same data points are used, giving

$$AIC_{\chi^2} = \chi^2 + 2k$$

This form is often convenient in that data fitting programs produce  $\chi^2$  as a statistic for the fit. For models with the same number of data points, the one with the lowest  $AIC_{\chi^2}$  should be preferred.

Similarly, if one has available the statistic  $R^2$  ("Variance Explained"), one may write

$$AIC_{R^2} = n \ln \frac{1 - R^2}{n} + 2k.$$

The Pearson correlation  $r = R$  is a special case of this. Here, independence of the observations is assumed.

## AICc and AICu

AICc is AIC with a second order correction for small sample sizes, to start with:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} = \chi^2 + \frac{2k}{1 - \frac{k-1}{n}}.$$

where k denotes the number of model parameters (one of them being the intercept). Since AICc converges to AIC as  $n$  gets large, AICc should be employed regardless of sample size (Burnham and Anderson, 2004).

McQuarrie and Tsai (1998: 22) define AICc as:

$$AICc = \ln \frac{RSS}{n} + \frac{n+k}{n-k-2},$$

and propose (p. 32) the closely related measure:

$$AICu = \ln \frac{RSS}{n-k} + \frac{n+k}{n-k-2}.$$

McQuarrie and Tsai ground their high opinion of AICc and AICu on extensive simulation work.

## QAIC

QAIC (the quasi-AIC) is defined as:

$$QAIC = 2k - \frac{1}{c} 2 \ln L$$

where  $c$  is a variance inflation factor. QAIC adjusts for over-dispersion or lack of fit. The small sample version of QAIC is

$$QAIC_c = QAIC + \frac{2k(k+1)}{n-k-1}.$$

## See also

- Bayesian information criterion
- deviance
- Deviance information criterion
- Hannan-Quinn information criterion
- Jensen-Shannon divergence
- Kullback-Leibler divergence
- Likelihood ratio test
- Occam's Razor

## Notes

## References

1. ^ Akaike, Hirotugu (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* **19** (6): 716–723. doi:10.1109/TAC.1974.1100705 (<http://dx.doi.org/10.1109%2FTAC.1974.1100705>) . MR0423716 (<http://www.ams.org/mathscinet-getitem?mr=0423716>) .
  2. ^ Burnham, Anderson, 1998, "Model Selection and Inference - A practical information-theoretic approach" ISBN 0-387-98504-2
- Brieman, L., 1992. *The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error*. Journal of the American Statistical Association, Vol 87 pp. 738–754
  - Burnham, K. P., and D. R. Anderson, 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer-Verlag. ISBN 0-387-95364-7.
  - -----, 2004. *Multimodel Inference: understanding AIC and BIC in Model Selection* (<http://www2.fmg.uva.nl/modelselection/presentations/AWMS2004-Burnham.pdf>) , Amsterdam Workshop on Model Selection.
  - Hurvich, C. M., and Tsai, C.-L., 1989. *Regression and time series model selection in small samples*. Biometrika, Vol 76. pp. 297–307
  - McQuarrie, A. D. R., and Tsai, C.-L., 1998. *Regression and Time Series Model Selection*. World Scientific. ISBN 981023242X

## External links

- Hirotogu Akaike comments on how he arrived at the AIC in This Week's Citation Classic (<http://www.garfield.library.upenn.edu/classics1981/A1981MS54100001.pdf>)
- Brieman's "a quiet scandal in the statistical community" (<http://www.kdnuggets.com/news/2009/n16/27i.html>)

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

Categories: Regression variable selection

---

- This page was last modified on 22 October 2010 at 22:07.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.  
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.