This Week's Citation Classic

This paper describes how the problem of statistical model selection can systematically be handled by using an information criterion (AIC) introduced by the author in 1971. The basic idea underlying the introduction of the criterion is explained and its practical utility is demonstrated by numerical examples. [The Science Citation Index® (SCI®) and the Social Sciences Citation Index® (SSCI®) indicate that this paper has been cited over 180 times since 1974.]

Hirotugu Akaike
Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku
Tokyo 106
Japan

October 7, 1981

"In 1968, I was developing a statistical identification procedure for a cement rotary kiln under normal noisy operating conditions by using a multi-variate autoregressive time series model. It quickly became clear that the main problem was the decision on the order, the number of past observations used to predict the behavior of the kiln. A solution was obtained by the introduction of the concept of final prediction error (FPE), the expected mean squared error of prediction by a model with the parameters determined by a statistical method.[1] The order selection was realized so as to minimize an estimate of FPE.

"In 1970, I received an invitation to the Second International Symposium on Information Theory, which was to be held in Tsahkadsor, Armenia, USSR. At that time, I was interested in extending FPE to the determination of the number of factors in a factor analysis model, a statistical model originally developed in psychology. However, it was not at all clear what the prediction error of this model was. The pressure of the impending deadline for the submission of the conference paper was increasing and this caused several weeks of sleepless nights.

"On the morning of March 16, 1971, while taking a seat in a commuter train, I suddenly realized that the parameters of the factor analysis model were estimated by maximizing the likelihood and that the mean value of the logarithmus of the likelihood was connected with the Kullback-Leibler information number. This was the quantity that was to replace the mean squared error of prediction. A new measure of the badness of a statistical model with parameters determined by the method of maximum likelihood was then defined by the formula[2] AIC = (-2) $\log_e$ (maximum likelihood) + 2 (number of parameters). AIC is an acronym for "an information criterion" and was first introduced in 1971. A model with a lower value of AIC is considered to be a better model.

"It is the general applicability and simplicity of model selection by AIC that prompted its use in such diversified areas as hydrology, geophysics, engineering, econometrics, psychometrics, and medicine. The procedure has some proof of its optimality[3] Nevertheless, due to its nonconventional style, AIC is not yet fully accepted by professional statisticians. It is mainly the increasing number of successful applications that caused the frequent citation of the paper."

1. Akaike H. Fitting autoregressive models for prediction. Ann. Inst. Statist. Math. 21:243-7, 1969.
2. Information theory and an extension of the maximum likelihood principle. (Petrov B N & Csaki F, eds.) Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 1973. p. 267-81.
3. A Bayesian analysis of the minimum AIC procedure. Ann. Inst. Statist. Math. 30A:9-14. 1978.