Open Journal of Statistics, **2012, 2, 309-312**
http://dx.doi.org/10.4236/ojs.2012.23038 Published Online July 2012 (http://www.SciRP.org/journal/ojs)

A Revision of *AIC* for Normal Error Models

Kunio Takezawa
Agroinformatics Division, Agricultural Research Center, National Agriculture and Food Research Organization,
Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
Email: nonpara@gmail.com

Received March 21, 2012; revised April 22, 2012; accepted May 4, 2012

ABSTRACT

Conventional Akaike's Information Criterion (*AIC*) for normal error models uses the maximum-likelihood estimator of error variance. Other estimators of error variance, however, can be employed for defining *AIC* for normal error models. The maximization of the log-likelihood using an adjustable error variance in light of future data yields a revised version of *AIC* for normal error models. It also gives a new estimator of error variance, which will be called the "third variance". If the model is described as a constant plus normal error, which is equivalent to fitting a normal distribution to one-dimensional data, the approximated value of the third variance is obtained by replacing $(n-1)$ (n is the number of data) of the unbiased estimator of error variance with $(n-4)$. The existence of the third variance is confirmed by a simple numerical simulation.

Keywords: *AIC*; AIC_c; Normal Error Models; Third Variance

1. Introduction

Akaike's Information Criterion (*AIC*) for multiple linear models with normal i.i.d. errors is defined as (e.g., [1,2])

$$\begin{aligned} AIC &= n\log(2\pi) + n\log(RSS/n) + n + 2q + 4 \\ &= -2l\left(\boldsymbol{X}, \boldsymbol{y} \middle| \{\hat{a}_j\}, \hat{\sigma}^2\right) + 2q + 4, \end{aligned} \tag{1}$$

where n is the number of data and q is the number of predictors of the multiple linear model. Hence, the number of regression coefficients in this model is $(q+1)$ when the error variance is regarded as a regression coefficient. $\boldsymbol{X}$ is a design matrix composed of the predictor values in the data. $\boldsymbol{y}$ is the vector composed of values of the target variable in the data. *RSS* stands for the residual sum of squares:

$$RSS = \sum_{i=1}^{n}\left(\hat{a}_0 + \sum_{j=1}^{q}\hat{a}_j x_{ij} - y_i\right)^2, \tag{2}$$

where $\{\hat{a}_0, \hat{a}_1, \cdots, \hat{a}_q\}$ are the estimators of regression coefficients of a multiple linear model. $x_{ij}(1 \le i \le n,\ 1 \le j \le q)$ is an element of $\boldsymbol{X}$. $y_i\ (1 \le i \le n)$ is an element of $\boldsymbol{y}$. $l\left(\boldsymbol{X}, \boldsymbol{y} \middle| \{\hat{a}_j\}, \hat{\sigma}^2\right)$ is the log-likeli- hood of the regression model in light of the data at hand. It is defined as

$$l\left(\boldsymbol{X}, \boldsymbol{y} \middle| \{\hat{a}_j\}, \hat{\sigma}^2\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\hat{\sigma}^2\right) - \frac{n}{2}. \tag{3}$$

The multiple linear model for obtaining Equations (1) and (3) contains $\{\hat{a}_0, \hat{a}_1, \cdots, \hat{a}_q\}$ given by the least squares method (also called the maximum likelihood method for normal errors), and the error variance ($\hat{\sigma}^2$) given by the maximum likelihood method. $\hat{\sigma}^2$ is derived using

$$\hat{\sigma}^2 = RSS/n. \tag{4}$$

$\hat{\sigma}^2$ defined above is used as the error variance in *AIC* because *AIC* is a statistic based on the maximum-likelihood estimator. However, the unbiased error variance shown below rather than the maximum-likelihood estimator of error variance is utilized in most statistical calculations.

$$\hat{\sigma}_{ub}{}^2 = RSS/(n-q-1). \tag{5}$$

The maximum-likelihood estimator of error variance may not be the only choice for the error variance for *AIC*. Hence, in this paper, we discusses the adjustment of error variance to calculate *AIC* for normal error models after recalling the derivation of conventional *AIC* for normal error models. Then, this consideration leads to a new estimator of error variance, which will be called the "third variance". Finally, the existence of the third variance is shown by a simple numerical simulation.

2. Derivation of *AIC* for Normal Error Models

Conventional *AIC* for normal error models is easily derived when the multiple linear model with normal error assumed by an analyst contains the real equation producing the data as a special case. *AIC* based on these assumption is an approximation of

$$\begin{aligned} &-2E\left[l\left(\boldsymbol{X}, \boldsymbol{y}^* \middle| \{\hat{a}_j\}, \hat{\sigma}^2 \right) \right] \\ &= E\left[n\log(2\pi) + n\log(\mathrm{RSS}/n) + n\mathrm{RSS}^*/RSS \right], \end{aligned} \tag{6}$$

where $\boldsymbol{y}^*$ is a vector comprising the values of the target variable in future data. The design matrix of future data is identical to that of the data at hand ($\boldsymbol{X}$). RSS^* is the residual sum of squares when future data are employed:

$$RSS^* = \sum_{i=1}^{n} \left(y^*{}_i - \hat{a}_0 - \sum_{j=1}^{q} \hat{a}_j x_{ij} \right)^2, \tag{7}$$

where $y_i \left(i \le i \le n \right)$ is an element of $\boldsymbol{y}^*$.

The expectation of *RSS* is given by

$$\begin{aligned} E[RSS] &= E\left[(\boldsymbol{y} - \hat{\boldsymbol{y}})^t (\boldsymbol{y} - \hat{\boldsymbol{y}}) \right] \\ &= E\left[\boldsymbol{y}^t (\boldsymbol{I} - \boldsymbol{H})^t (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y} \right] \\ &= E\left[\boldsymbol{y}^t (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y} \right] \\ &= E\left[\left(\tilde{\boldsymbol{y}}^t + \boldsymbol{\varepsilon}^t \right) (\boldsymbol{I} - \boldsymbol{H}) (\tilde{y} + \boldsymbol{\varepsilon}) \right] \\ &= E\left[\boldsymbol{\varepsilon}^t \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^t \boldsymbol{H} \boldsymbol{\varepsilon} \right] \end{aligned} \tag{8}$$

where $\boldsymbol{H}$ is the symmetric matrix ($\boldsymbol{H}^t = \boldsymbol{H}$) and idempotent ($\boldsymbol{H}^2 = \boldsymbol{H}$). Furthermore, it is assumed that if $\tilde{\boldsymbol{y}}$ (the values of the target variable with no errors) is employed, $\boldsymbol{H}\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}$ holds because it is assumed that the regression equation adopted here contains the real equation producing the data as a special case.

Since $\boldsymbol{\varepsilon}$ is a normal error (the mean is 0 and the variance is σ^2), the following equation is obtained:

$$E\left[\boldsymbol{\varepsilon}^t \boldsymbol{\varepsilon} \right] = E\left[\sum_{i=1}^{n} \varepsilon_i \varepsilon_i \right] = n\sigma^2. \tag{9}$$

The following equation is also derived:

$$\begin{aligned} E\left[\boldsymbol{\varepsilon}^t \boldsymbol{H} \boldsymbol{\varepsilon} \right] &= E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \left[\boldsymbol{H} \right]_{ij} \varepsilon_j \right] \\ &= \mathrm{trace}\left(\boldsymbol{H} \right) \sigma^2 = \left(q+1 \right) \sigma^2, \end{aligned} \tag{10}$$

where $\mathrm{trace}\left(\boldsymbol{H} \right)$ is the trace of $\boldsymbol{H}$. Hence, Equations (8)-(10) give

$$E[RSS] = (n - q - 1)\sigma^2. \tag{11}$$

Therefore, RSS/σ^2 obeys the χ^2 distribution with ($n - q - 1$) degrees of freedom. A similar calculation yields

$$\begin{aligned} E\left[RSS^* \right] &= E\left[\left(\boldsymbol{y}^* - \hat{\boldsymbol{y}} \right)^t \left(\boldsymbol{y}^* - \hat{\boldsymbol{y}} \right) \right] \\ &= E\left[\left(\tilde{\boldsymbol{y}} + \boldsymbol{\varepsilon}^* - \boldsymbol{H}\left(\tilde{\boldsymbol{y}} + \boldsymbol{\varepsilon} \right) \right)^t \left(\tilde{\boldsymbol{y}} + \boldsymbol{\varepsilon}^* - \boldsymbol{H}\left(\tilde{\boldsymbol{y}} + \boldsymbol{\varepsilon} \right) \right) \right] \\ &= E\left[\left(\boldsymbol{\varepsilon}^* \right)^t \boldsymbol{\varepsilon}^* + \boldsymbol{\varepsilon}^t \boldsymbol{H} \boldsymbol{\varepsilon} \right] = \left(n + q + 1 \right) \sigma^2. \end{aligned} \tag{12}$$

Hence, RSS^*/σ^2 obeys the χ^2 distribution with ($n + q + 1$) degrees of freedom.

Considering Equations (11) and (12), the $E[]$ content in the third term on the right-hand side of Equation (6) is transformed into

$$\frac{nRSS^*}{RSS} \sim \frac{n\chi^2_{n+q+1}}{\chi^2_{n-q-1}} = n\frac{n+q+1}{n-q-1} F_{n+q+1, n-1-1}. \tag{13}$$

where χ^2_{n-q-1} is a random variable that obeys the χ^2 distribution with $\left(n - q - 1 \right)$ degrees of freedom. χ^2_{n+q+1} is a random variable that obeys the χ^2 distribution with $\left(n + q + 1 \right)$ degrees of freedom, and $F_{n+q+1, n-q-1}$ is an F distribution. The first degrees of freedom is $\left(n + q + 1 \right)$ and the second degrees of freedom is $\left(n - q - 1 \right)$. Hence, the expectation of the random variable given by Equation (13) is

$$\begin{aligned} n\frac{n+q+1}{n-q-1} E\left[F_{n+q+1, n-q-1} \right] &= n\frac{n+q+1}{n-q-1} \cdot \frac{n-q-1}{n-q-1-2} \\ &= n\frac{n+q+1}{n-q-3}. \end{aligned} \tag{14}$$

By substituting this equation into Equation (6) and using Equation (3), the following equation is obtained:

$$\begin{aligned} &-2E\left[l\left(\boldsymbol{X}, \boldsymbol{y}^* \middle| \{\hat{a}_j\}, \hat{\sigma}^2 \right) \right] \\ &\approx n\log(2\pi) + n\log\left(\frac{RSS}{n} \right) + n\frac{n+q+1}{n-q-3} \\ &= -2l\left(\boldsymbol{X}, \boldsymbol{y} \middle| \{\hat{a}_j\}, \hat{\sigma}^2 \right) - n + n\frac{n+q+1}{n-q-3}. \end{aligned} \tag{15}$$

This is AIC_c for normal error models ([1,3,4]).

When n is large, the approximation below holds:

$$\begin{aligned} n\frac{n+q+1}{n-q-3} = n\frac{1 + (q+1)/n}{1 - (q+3)/n} &\approx n\left(1 + \frac{q+1}{n} + \frac{q+3}{n} \right) \\ &\approx n\left(1 + \frac{2q+4}{n} \right). \end{aligned} \tag{16}$$

By substituting this equation into Equation (6) and using Equation (3), the following equation is obtained:

$$\begin{aligned} &-2E\left[l\left(\boldsymbol{X}, \boldsymbol{y}^* \middle| \{\hat{a}_j\}, \hat{\sigma}^2 \right) \right] \\ &\approx n\log(2\pi) + n\log(\mathrm{RSS}/n) + n + 2q + 4 \\ &= -2l\left(\boldsymbol{X}, \boldsymbol{y} \middle| \{\hat{a}_j\}, \hat{\sigma}^2 \right) + 2q + 4. \end{aligned} \tag{17}$$

This is conventional *AIC* for normal error models.

3. Adjustment of Error Variance of *AIC* for Normal Error Models

The estimator of error variance is assumed to be adjustable. That is, error variance (σ^2_{AIC}) is defined as

$$\sigma_{AIC}^2 = RSS/(n-\alpha), \tag{18}$$

where α is a constant for adjusting error variance. The use of σ_{AIC}^2 in AIC_c (Equation (15)) yields AIC_c^a (AIC-adjustable):

$$AIC_c^a = n\log(2\pi) + n\log\left(\frac{RSS}{n-\alpha}\right) + (n-\alpha)\frac{n+q+1}{n-q-3}. \tag{19}$$

Then, $\hat{\alpha}$ which minimizes AIC_c^a is

$$\hat{\alpha} = n\left(1-\frac{n-q-3}{n+q+1}\right) \approx 2q+4. \tag{20}$$

Hence, the following $\hat{\sigma}_{AIC}^2$ is different from the unbiased estimator of error variance:

$$\hat{\sigma}_{AIC}^2 = RSS/(n-\hat{\alpha}). \tag{21}$$

$\hat{\sigma}_{AIC}^2$ will be called the "third variance" because the discovery of this variance follows those of the maximum-likelihood estimator of error variance and the unbiased estimator of error variance. In particular, when $q=0$ which indicates the fitting of a normal distribution to one-dimensional data. Although $\alpha=0$ or $\alpha=1$ is adopted conventionally, $\alpha=4$ is preferable in terms of log-likelihood in light of future data.

The substitution of Equations (20) and (21) to Equation (15) leads to

$$\begin{aligned} AIC_c^u &= n\log(2\pi) \\ &\quad + n\log\left(\frac{RSS}{n-n\left(1-(n-q-3)/(n+q+1)\right)}\right) \\ &\quad + \left(n-n\left(1-\frac{n-q-3}{n+q+1}\right)\right)\cdot\frac{n+q+1}{n-q-3} \\ &= n\log(2\pi) + n\log\left(\frac{RSS(n+q+1)}{n-q-3}\right) + n, \end{aligned} \tag{22}$$

where AIC_c^u denotes the "ultimate AIC". Simulation studies show that the model selection characteristics of AIC_c^u falls somewhere between AIC and AIC_c.

4. Numerical Simulation

The simulation data consists of $\{y_i\}(1\le i\le 100)$ (realizations of $N(-13.0, 4^2)$) and $\{y_i^*\}(1\le i\le 100)$ (realizations of $N(-13.0, 4^2)$). $\hat{a}_0$, $\hat{\sigma}^2$, RSS, and RSS^* are expressed as follows:

$$\hat{a}_0 = \frac{1}{n}\sum_{i=1}^n y_i,\ \hat{\sigma}^2 = RSS/(n-\alpha), \tag{23}$$

$$RSS = \sum_{i=1}^n (y_i-\hat{a}_0)^2,\ RSS^* = \sum_{i=1}^n (y_i^*-\hat{a}_0)^2, \tag{24}$$

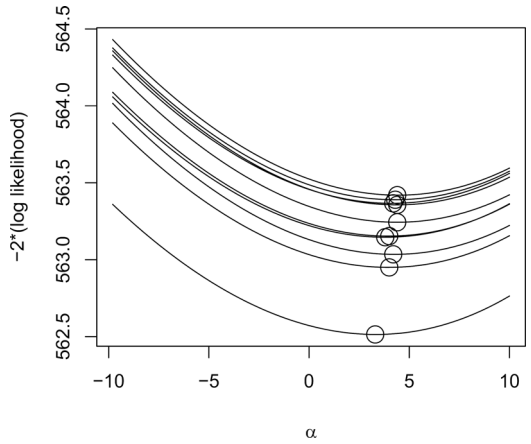

Figure 1. Relationship between α and average $-2l\left(\{y_i^*\}\middle|\hat{a}_0,\hat{\sigma}^2\right)$. A circle indicates the minimum point of each line. Ten lines reflect 10 repeats of the simulations.

where $n=100$. By altering the seed of random values, 5000 sets of $\{y_i\}$ and $\{y_i^*\}$ are obtained. Then, 5000 values of $-2l\left(\{y_i^*\}\middle|\hat{a}_0,\hat{\sigma}^2\right)$ are obtained and averaged. This procedure is carried out using one of the values $\{-9.8,-9.6,-9.4,\cdots,10\}$ as α.

Figure 1 shows the result of this simulation. Ten lines show that the simulation is repeated 10 times by changing the seed of random values. Each minimum point is located around the $\alpha=4$ point; these ten points apparently deviate from the $\alpha=1$ and $\alpha=0$ points. This shows that $\alpha=4$ gives a better log-likelihood in light of future data and that the third variance should be considered.

5. Conclusion

The error variance for AIC is adjustable. The optimization of the errror variance yields AIC_c^u in which the third variance is adopted as the error variance. The third variance is different from both the unbiased estimator of error variance and the maximum-likelihood estimator of error variance. The features and usage of the third variance remains to be elucidated.

REFERENCES

[1] K. P. Burnham and D. R. Anderson, "Model Selection and Multi-Model Inference A Practical Information-Theoretic Approach," Springer, Berlin, 2010.

[2] S. Konishi and G. Kitagawa, "Information Criteria and Statistical Modeling," Springer, Berlin, 2007.

[3] C. M. Hurvich and C.-L. Tsai, "Regression and Time Series Model Selection in Small Samples," *Biometrika*, Vol. 76, No. 2, 1989, pp. 297-307. doi:10.1093/biomet/76.2.297

[4] N. Sugiura, "Further Analysis of the Data by Akaike's Information Criterion and Finite Corrections," *Communications in Statistics-Theory and Methods*, Vol. 7, No. 1, 1978, pp. 13-26. doi:10.1080/03610927808827599

