The Information Criterion

Masume Ghahramani
Department of Statistics, Payam Noor University, Iran, mghdatagilan@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Available at: http://digitalcommons.wayne.edu/jmasm/vol13/iss2/25

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
The Information Criterion

M. Ghahramani
Payam Noor University
Tehran, Iran

The Akaike information criterion, AIC, is widely used for model selection. Using the AIC as the estimator of asymptotic un bias for the second term Kullbake-Leibler risk considers the divergence between the true model and offered models. However, it is an inconsistent estimator. A proposed approach the problem is the use of AIC, a consistently offered information criterion. Model selection of classic and linear models are considered by a Monte Carlo simulation.

Keywords: Consistency, AIC, information criterion, Kullbake-Leibler risk, model selection

Introduction

Statistical modeling is used for investigating a random phenomenon that isn’t completely predictable. One of the criteria frequently used in model selection is the Kullbake-Leibler (KL) information criterion (Kullback and Leibler, 1951). This information criterion was introduced as one risk in model selection. Akaike (1973) introduced information criterion, AIC, as an estimator of asymptotic un bias for the second term KL risk and to form a penalty likelihood function. Akaike stated modeling isn’t only finding a model which describes the behavior of the observed data, but its main aim is predicated as a possible good, and the future of the process is under investigation. Hall (1987) used the Kullbake-Leibler risk considered bias and variance in the approximate density function. Bozdogan (2000), with the error distinction in the model selection, considered two errors from bias and variance in the estimation of model selection. Choi and Kiffer (2006), and Cawley and Talbot (2010) have considered the over fitting in model selection, and they showed over fitting results from the bias when modeling phenomena have been considered. Over the years, corrections have been made on penalty term, and criteria such as AIC (Akaike, 1973), TIC (Takeuchi, 1976), and KIC (Cavanaugh, 1994) have been

M. Ghahramani is in the Department of Mathematical Statistics. Email him at estimatormgh@gmail.com.
introduced. In section 2, we state the Kullback-Liebler risk, and the necessity of definitions. In section 3, a consistent information criterion is proposed instead of the AIC. In section 4, we present the results of our simulation studies.

**Kullback-Leibler Risk**

Let $X = (X_1, X_2, ..., X_n)$ be a (i.i.d) random sample from true model and unknown, $h(.)$, and the family $F_{\hat{\theta}} = \{f(\cdot; \hat{\theta}_k) = f_{\hat{\theta}_k}; \hat{\theta}_k \in \Theta \subseteq \mathbb{R}^k\}$ from offered models has been considered for approximate true model.

**Definition 1**

The family $F_{\hat{\theta}}$ is well specified if there is a $\theta_0 \in \Theta$ such that $h(\cdot) = f(\cdot; \theta_0)$; otherwise it is misspecified.

**Definition 2**

The KL risk defines for generate model and unknown $h(.)$, and offered model $f_{\hat{\theta}}$ as

$$KL(h, f_{\hat{\theta}}) = E_{\hat{\theta}} \left[ \log \left( \frac{h(.)}{f(\cdot; \hat{\theta})} \right) \right] = E_{h} [\log h(.)] - E_{h} [\log f(\cdot; \hat{\theta})]$$

where the expectation is taken with respect to the unknown model $h(.)$. The first term in the right hand side of (1) is called irrelevant part, because it doesn’t depend on $\theta_k$, and the second term is called relevant part. Based on the properties of the KL risk, the smaller value showed the closeness of the offered model to the unknown and true model. Therefore the problem is reduced to obtain a good estimate of the expected log-likelihood. Since the expectation is with respect to the model with unknown parameters, one estimator is

$$E_{h} [\log f(\cdot; \hat{\theta}_n)] = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i; \hat{\theta}_n).$$

Thus, $\hat{\theta}_n$ is the maximum likelihood estimator of $\theta_k$ and $f(\cdot; \hat{\theta}_n)$ is the maximum likelihood function. The bias of maximum log-likelihood is as
PAIRWISE COMPARISON IN REPEATED MEASURES

Bias estimator = $E_h\{\log f(.; \hat{\theta}_n) - nE_h\{\log f(Z ; \hat{\theta}_n)\}$

where $Z$ is a random variable (i.i.d) with $X_i s$. The general form of the information criterion that has been shown by IC, as

$$IC = -2\sum_{i=1}^n \log f(X_i ; \hat{\theta}_n) + 2\{\text{bias estimator}\} = -2l_f(\hat{\theta}_n) + 2\{\text{bias estimator}\}.$$ 

Akaike, when offered family is well specified, size of bias is estimated with dimensional parameter $\hat{\theta}_n$, means $k$, and the Akaike information criterion is stated as

$$AIC = -2l_f(\hat{\theta}_n) + 2k.$$ 

With attention to form the AIC by increasing the number of parameters in the offered model the penalty term, $2k$ will be increased and the term $-2\sum_{i=1}^n \log f(X_i ; \hat{\theta}_n)$ will be decrease. Penalty term is constant to chance of size sample in the information criterion AIC, and by increasing the size sample, AIC cannot distinguish the true model with the probability one. Therefore this problem is the same concept of inconsistency for an information criterion. Following the inconsistency of information criterion AIC, based on the definition similar to the definition of AIC, a consistent of information criterion, which called A'IC has presented. Akaike information criterion, by Akaike for model selection is introduced, but this useful criterion is inconsistent (Akaike, 1973).

The information criterion is obtained as follows. The basis of the log-likelihood function is

$$b = E_h\{\log f(.; \hat{\theta}_n) - nE_h\log f(Z ; \hat{\theta}_n)\}$$

where in the second term of the right hand side the inner expectation is calculated with respect to $h(z)$ and the outer expectation is calculated with respect to $h(x)$. By evaluating the bias it is decomposed as follows:

$$b = E_h\{gf(.; \hat{\theta}_n) - \log f(.; \theta_0)\} + E_h\{\log f(.; \theta_0) - nE_h\{\log f(Z ; \theta_0)\}\}$$

$$+ nE_h\{E_h\{\log f(Z ; \theta_0) - E_h\{\log f(Z ; \hat{\theta}_n)\}\} = b_1 + b_2 + b_3.$$
The three expectations are calculated separately \( b_1, \ b_2, \) and \( b_3 \).

a) For calculation of \( b_1 \) by writing 
\[
l_f(\theta_0) = l_f(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^T \frac{\partial l_f(\theta)}{\partial \theta} \big|_{\theta = \hat{\theta}_n} + \frac{1}{2} (\theta_0 - \hat{\theta}_n)^T \frac{\partial^2 l_f(\theta)}{\partial \theta^2} \big|_{\theta = \hat{\theta}_n} (\theta_0 - \hat{\theta}_n) + o_p(1),
\]

\( O_p(1) \) is an expression of quantity that in the probability tends to zero.

\[
J(\theta_0) = -E_h \left[ \frac{\partial^2 l_f(\theta)}{\partial \theta \partial \theta^T} \right] |_{\theta = \hat{\theta}_n}
\]

Thus, the relation above can be approximated, as

\[
l_f(\hat{\theta}_n) - l_f(\theta_0) \approx \frac{n}{2} (\theta_0 - \hat{\theta}_n)^T J(\theta_0)(\theta_0 - \hat{\theta}_n) + o_p(1),
\]

This based on the \( b_1 \) can be written as

\[
b_1 = E_h \{ l_f(\hat{\theta}_n) - l_f(\theta_0) \} \approx E_h \left\{ \frac{n}{2} (\theta_0 - \hat{\theta}_n)^T J(\theta_0)(\theta_0 - \hat{\theta}_n) \right\} \quad (2)
\]

b) The \( b_2 \) doesn’t contain an estimator and it can easily be written as

\[
b_2 = E_h \{ gf(\cdot; \theta_0) - nE_h \{ \log f(Z; \theta_0) \} \} = 0
\]

(3)

c) For calculation of value the \( b_3 \), first, the phrase \( E_h \{ \log f(Z; \theta_0) \} \) be defined equally of \( \eta(\hat{\theta}_n) \). By using from Taylor expectation \( \eta(\hat{\theta}_n) \) around \( \theta_0 \),

\[
\eta(\hat{\theta}_n) = \eta(\theta_0) + (\hat{\theta}_n - \theta_0)^T \frac{\partial \eta(\theta)}{\partial \theta} \big|_{\theta = \theta_0} + \frac{1}{2} (\hat{\theta}_n - \theta_0)^T \frac{\partial^2 \eta(\theta)}{\partial \theta \partial \theta^T} \big|_{\theta = \theta_0} (\hat{\theta}_n - \theta_0) + o_p(1)
\]
PAIRWISE COMPARISON IN REPEATED MEASURES

with attention to the \( \frac{\partial \eta(\theta)}{\partial \eta} \bigg|_{\theta=\theta_0} = 0 \). Thus when \( n \) tends to infinity, the relation above can be approximated as

\[
\eta(\hat{\theta}_n) \approx \eta(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^T J(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(1).
\]

Thus the \( b_3 \) can be written as

\[
b_3 = nE_{\theta_0}\{E_{\theta_n}\{\log f(Z;\theta_n)\} - E_{\theta_0}\{\log f(Z;\hat{\theta}_n)\}\}
\]

\[
\approx \frac{n}{2} E_{\theta_0}\{(\theta_n - \hat{\theta}_n)^T J(\theta_0)(\hat{\theta}_n - \theta_0)\} \tag{4}
\]

If the family of \( F_{\theta_i} \) is well specified, with attention to quadratic forms in relations (2) and (4), that converge to centrally distributed chi-square with \( k \) degrees of freedom, then \( b_1 \) and \( b_3 \) can be written as

\[
b_1 = \frac{n}{2} k
\]

By combining of \( b_1 \) and \( b_3 \), in relation (5) and \( b_2 \), in relation (3), bias the \( b \) is

\[
b = b_1 + b_2 + b_3 = nk.
\]

By replacing the value of \( b \) in the general form of the information criterion, the offered information criterion called, \( A^{IC} \) is obtained as

\[
A^{IC} = -2\sum_{i=1}^{n} \log f(X_i;\hat{\theta}_n) + 2nk \tag{6}
\]

In the offered information criterion \( A^{IC} \), penalty term 2nk changes will change with sample size. So, if sample size will be very large, information criterion \( A^{IC} \), with the probability of one, find the true model data. In other words, information criterion \( A^{IC} \) is the only consistent information criterion that has been obtained based on the Kullback-Leibler risk. To show consistency of information criterion \( A^{IC} \), let the maximum likelihood function estimator for the offered model \((f(.;\theta) = f(\theta))\) and optimal model \((f(.;\theta_0) = f(\theta_0))\) with respectively
$l_f(\hat{\theta}_{k(n)})$ and $l_f(\hat{\theta}_{k_0(n)})$. With regard to relation (6) information criterion A'IC, for the model $f(\theta_k)$ and $f(\theta_{k_0})$, we have

$$A'IC(f(\theta_k)) = -2l_f(\hat{\theta}_{k(n)}) + 2nk,$$

$$A'IC(f(\theta_{k_0})) = -2l_f(\hat{\theta}_{k_0(n)}) + 2nk_0$$

If there is $k > k_0$, consistency for information criterion A'IC is given by

$$P(A'ICf(\theta_k) - A'ICf(\theta_{k_0}) > 0)$$

$$= P(-2l_f(\hat{\theta}_{k(n)}) + 2nk - (-2l_f(\hat{\theta}_{k_0(n)}) + 2nk_0) > 0)$$

$$= P(2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)}) < 2nk - 2nk_0)$$

$$= P(U_n < 2n(k - k_0)) = F(2n(k - k_0))\xrightarrow{p} F(\infty) = 1 \quad (7)$$

In relation (7), $U_n$ is $2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)})$ and the distribution function of chi-square has been shown by F. Therefore it tends in of the probability to one. Thus A'IC is a consistent information criterion. (For further study about the consistency of an information criterion, see Hu and Shao 2008).

**Simulation**

A simulation was conducted for usage and comparison of the offered information criterion, A'IC, with the information criterion AIC, by using Monte-Carlo simulation, for linear regression and classic models. This simulation of linear regression model is supposed that well specified family

$$F_{\theta_k} = \{ f(\cdot; \theta_k) = f_{\theta_k} \cdot \theta_k \in \Theta \subseteq \mathbb{R}^k \}$$

and misspecified family

$$G_{\beta_d} = \{ g(\cdot; \beta_d) = g_{\beta_d} \cdot \beta_d \in \mathbb{R}^d \}$$

are given for estimating the true model. Let $f: y_i = 0.3 + 0.5x_{i1} + x_{i2} + 0.7x_{i3} + \varepsilon_{i1}, i = 1, \ldots, n$ as the true model so that $\varepsilon_{i1}$, has been generated as random from distribution N(0,2). Models $f_1: y_i = \hat{\theta}_0 + \hat{\theta}_1x_{i1} + \hat{\theta}_2x_{i2} + \hat{\theta}_3x_{i3}, i = 1, \ldots, n$ and,
\[ f_2 : y_i = \hat{\theta}_0 + \hat{\theta}_1 x_{i1} + \hat{\theta}_2 x_{i2} + \hat{\theta}_3 x_{i3} + \hat{\theta}_4 x_{i4} \quad i = 1, \ldots, n \]  
offered models, which have been generated from \( F_{\theta} \). Also we have \( g : y_i = 0.3 + 0.5 z_{i1} + 3 z_{i2} + 1.1 z_{i3} + \varepsilon_{i2} i = 1, \ldots, n \)

Table 1. Comparison of AIC with A'IC by using from Monte-Carlo simulation for linear regression models \( f_1, f_2, g_1, \) and \( g_2 \).

<table>
<thead>
<tr>
<th>Size</th>
<th>Model</th>
<th>AIC</th>
<th>A'IC</th>
<th>( \Delta \text{AIC} )</th>
<th>( \Delta \text{A'IC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=50</td>
<td>( f_1 )</td>
<td>-2990</td>
<td>-2598</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>-2700</td>
<td>-2210</td>
<td>290</td>
<td>388</td>
</tr>
<tr>
<td></td>
<td>( g_1 )</td>
<td>200</td>
<td>592</td>
<td>3190</td>
<td>2006</td>
</tr>
<tr>
<td></td>
<td>( g_2 )</td>
<td>248</td>
<td>738</td>
<td>3238</td>
<td>1860</td>
</tr>
<tr>
<td>n=100</td>
<td>( f_1 )</td>
<td>-3500</td>
<td>-2708</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>-3200</td>
<td>-2210</td>
<td>300</td>
<td>498</td>
</tr>
<tr>
<td></td>
<td>( g_1 )</td>
<td>430</td>
<td>1222</td>
<td>3930</td>
<td>3930</td>
</tr>
<tr>
<td></td>
<td>( g_2 )</td>
<td>455</td>
<td>1445</td>
<td>3955</td>
<td>4153</td>
</tr>
<tr>
<td>n=200</td>
<td>( f_1 )</td>
<td>-5400</td>
<td>-3808</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>-4360</td>
<td>-2370</td>
<td>1040</td>
<td>1438</td>
</tr>
<tr>
<td></td>
<td>( g_1 )</td>
<td>210</td>
<td>1802</td>
<td>5610</td>
<td>5610</td>
</tr>
<tr>
<td></td>
<td>( g_2 )</td>
<td>240</td>
<td>2230</td>
<td>5640</td>
<td>6038</td>
</tr>
<tr>
<td>n=350</td>
<td>( f_1 )</td>
<td>-7230</td>
<td>-4438</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>-6400</td>
<td>-2910</td>
<td>30</td>
<td>1528</td>
</tr>
<tr>
<td></td>
<td>( g_1 )</td>
<td>325</td>
<td>3117</td>
<td>7555</td>
<td>7555</td>
</tr>
<tr>
<td></td>
<td>( g_2 )</td>
<td>360</td>
<td>3850</td>
<td>7590</td>
<td>8288</td>
</tr>
<tr>
<td>n=500</td>
<td>( f_1 )</td>
<td>-9730</td>
<td>-5738</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>( f_2 )</td>
<td>-9300</td>
<td>-4310</td>
<td>430</td>
<td>1428</td>
</tr>
<tr>
<td></td>
<td>( g_1 )</td>
<td>400</td>
<td>4392</td>
<td>10130</td>
<td>10130</td>
</tr>
<tr>
<td></td>
<td>( g_2 )</td>
<td>425</td>
<td>5415</td>
<td>10155</td>
<td>11153</td>
</tr>
</tbody>
</table>

450
Thus, \( \varepsilon_i \), has been generated as random from distribution \( N(0,1) \), and Models

\[ g_1 : y_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \hat{\beta}_3 z_{i3} \ i = 1, ..., n \]
and

\[ g_2 : y_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \hat{\beta}_3 z_{i3} + \hat{\beta}_4 z_{i4} \ i = 1, ..., n. \]

The models are generated from \( G_{\beta_j} \). This simulation is achieved by using from software R, and the number of repetitions are \( 10^3 \), and samples \( n = 50, 100, 200, 350, 600 \), have been considered. The results of simulation are presented in the Table 1.

In the third and fourth columns of Table 1, the value of AIC and A'IC are presented in order to various values of \( n \) and for offered models \( f_1, f_2, g_1, \) and \( g_2 \). Therefore the relation between values AIC for offering models is obvious as

\[ AIC(f_1) < AIC(f_2) < AIC(g_1) < AIC(g_2). \]

The family \( F_{\theta} \) is well specified, but the family \( G_{\beta_j} \) is misspecified. Thus, this relation is logical. With attention to the fourth column of Table 1 recent relation also is confirmed for A'IC. In other words

\[ A'IC(f_1) < A'IC(f_2) < A'IC(g_1) < A'IC(g_2). \]

With increasing \( n \), the value of AIC has been increased for the offered models, but the direction is confirmed unequally. The absolute magnitude difference of the value AIC and A'IC between the model of \( f_1 \) and other models is presented in the fifth and sixth columns of table. The absolute magnitude differences have been shown by the symbols of \( \Delta AIC \) and \( \Delta A'IC \). If there are symbols, as

\[ \Delta AIC_{f_1 \rightarrow f_2} = |AIC(f_1) - AIC(f_2)| \quad \text{and} \quad \Delta AIC_{g_1 \rightarrow g_2} = |AIC(g_1) - AIC(g_2)|, \quad j = 1, 2 \]
and

\[ \Delta A'IC_{f_1 \rightarrow f_2} = |A'IC(f_1) - A'IC(f_2)| \quad \text{and} \quad \Delta A'IC_{g_1 \rightarrow g_2} = |A'IC(g_1) - A'IC(g_2)|, \quad j = 1, 2. \]

For \( n = 50, 100, 150, 200, 350, 500 \), and models \( f_1, f_2, g_1, \) and \( g_2 \) will be confirmed the relation as

\[ \Delta AIC_{f_1 \rightarrow f_2} < \Delta AIC_{f_1 \rightarrow g_1} < \Delta AIC_{g_1 \rightarrow g_2} \quad \text{and} \quad \Delta A'IC_{f_1 \rightarrow f_2} < \Delta A'IC_{f_1 \rightarrow g_1} < \Delta A'IC_{g_1 \rightarrow g_2}. \]

With attention to these relations the direction of similarity the model selection for information criteria AIC and A'IC for various \( n \) have been shown with this the quality that the criterion A'IC is a consistent information criterion.
Table 2. Comparison of AIC with A'IC by using Monte-Carlo simulation, for the state that generates model data is Normal standard and offered models are from a Laplace family with different parameters.

<table>
<thead>
<tr>
<th>Size</th>
<th>Model</th>
<th>AIC</th>
<th>A'IC</th>
<th>Δ AIC</th>
<th>Δ A'IC</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=50</td>
<td>$f_1 = \text{lap}(0,1.3)$</td>
<td>-90</td>
<td>106</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$f_2 = \text{lap}(0,1)$</td>
<td>-70</td>
<td>126</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>$f_3 = \text{lap}(2,1)$</td>
<td>-56</td>
<td>140</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>$f_4 = \text{lap}(−2,0.9)$</td>
<td>-50</td>
<td>146</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>$f_1 = \text{lap}(0,1.3)$</td>
<td>-200</td>
<td>196</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$f_2 = \text{lap}(0,1)$</td>
<td>-160</td>
<td>236</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>$f_3 = \text{lap}(2,1)$</td>
<td>-143</td>
<td>253</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>$f_4 = \text{lap}(−2,0.9)$</td>
<td>-130</td>
<td>266</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>n=100</td>
<td>$f_1 = \text{lap}(0,1.3)$</td>
<td>-345</td>
<td>451</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$f_2 = \text{lap}(0,1)$</td>
<td>-295</td>
<td>501</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>$f_3 = \text{lap}(2,1)$</td>
<td>-255</td>
<td>541</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>$f_4 = \text{lap}(−2,0.9)$</td>
<td>-240</td>
<td>556</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>n=200</td>
<td>$f_1 = \text{lap}(0,1.3)$</td>
<td>-610</td>
<td>786</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$f_2 = \text{lap}(0,1)$</td>
<td>-525</td>
<td>871</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>$f_3 = \text{lap}(2,1)$</td>
<td>-487</td>
<td>909</td>
<td>123</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>$f_4 = \text{lap}(−2,0.9)$</td>
<td>-441</td>
<td>955</td>
<td>169</td>
<td>169</td>
</tr>
<tr>
<td>n=350</td>
<td>$f_1 = \text{lap}(0,1.3)$</td>
<td>-986</td>
<td>1010</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$f_2 = \text{lap}(0,1)$</td>
<td>-865</td>
<td>1131</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>$f_3 = \text{lap}(2,1)$</td>
<td>-777</td>
<td>1219</td>
<td>209</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>$f_4 = \text{lap}(−2,0.9)$</td>
<td>-670</td>
<td>11326</td>
<td>316</td>
<td>316</td>
</tr>
</tbody>
</table>

In the third and fourth column Table 2 values of AIC and A'IC for n=50, 100, 200, 350 and 500, have been respectively considered Laplace offered models $f_1$, $f_2$, $f_3$, and $f_4$. Therefore the relation between values AIC for offered models of Laplace family is obvious as $A'IC(f_1) < A'IC(f_2) < A'IC(f_3) < A'IC(f_4)$. 
With attention to the fourth column in the Table 2, the recent relations also confirmed for A’IC. In other words, \( A’IC(f_1) < A’IC(f_2) < A’IC(f_3) < A’IC(f_4) \).

In the fifth and sixth columns the absolute magnitude difference have been presented respectively for the value AIC and A’IC between the model of \( f_i \) and any which from other models to confirm with any \( n \), symbols of \( \Delta AIC \) and \( \Delta A’IC \) has been shown. With attention to these two columns for \( n \)’s different have \( \Delta AIC = \Delta A’IC \). If we have these symbols as

\[
\Delta AIC_{|f_i - f_j|} = |AIC(f_i) - AIC(f_j)| \quad \text{for } i \neq j \quad \text{and} \quad \Delta A’IC_{|f_i - f_j|} = |A’IC(f_i) - A’IC(f_j)| \quad \text{for } i \neq j
\]

for any \( n = 50, 100, 200, 350, 500 \), models \( f_1, f_2, f_3, \) and \( f_4 \), confirms the relation as

\[
\Delta AIC_{|f_1 - f_2|} < \Delta AIC_{|f_1 - f_3|} < \Delta AIC_{|f_1 - f_4|} \quad \text{and} \quad \Delta A’IC_{|f_1 - f_2|} < \Delta A’IC_{|f_1 - f_3|} < \Delta A’IC_{|f_1 - f_4|}.
\]

With attention to these relations, the direction of similarity model selection for information criteria AIC and A’IC for various \( n \) has been shown. But the information criterion A’IC is the consistent information criterion.

**Conclusion**

In this article investigating the inconsistent information criterion AIC, and by eliminating the inconsistency problem, a method for achieving an information criterion has been presented based on Kullback-Leibler risk and the consistent information criterion A’IC has been obtained. Therefore this information criterion is the only consistent information criterion and asymptotically unbiased. It is obtained based on Kullback-Leibler risk. Via simulation for linear regression and classic model, the quality of model selection was shown throughout the two information criterion, AIC and A’IC. According to the consistent information criterion of A’IC, it is possible for further discussion and to refine the other information criteria, which are based on Kullback-Leibler risk (as AICc and KICc) and add the consistency feature to the criteria.
References


