171:290 Model Selection
Lecture III: Corrected AIC and Modified AIC, AICc and MAIC

Joseph E. Cavanaugh

Department of Biostatistics
Department of Statistics and Actuarial Science
The University of Iowa

September 4, 2012

Introduction

- The Akaike information criterion, AIC, is derived as an estimator of the expected Kullback discrepancy between the true model and a fitted candidate model.
- The asymptotic justification of the criterion requires two strong assumptions:
  - (i) that the true model is contained in the candidate class under consideration,
  - (ii) that the vector of maximum likelihood estimators satisfies the conventional large-sample properties of MLE's.

Introduction

- Can these assumptions be relaxed?
- The corrected Akaike information criterion, AICc, is based on a development where the large-sample requirement in (ii) is relaxed.
- The Takeuchi (1976) information criterion, TIC, is based on a development where the true model assumption (i) is relaxed.
- The modified Akaike information criterion, MAIC, is based on a development where the both the true model assumption (i) and the large-sample requirement in (ii) are relaxed.

Introduction

- In Lecture III, we introduce AICc and MAIC.
- In Lecture IV, we introduce TIC.

Introduction

Outline:

- Review of AIC (Lecture II)
- Bias Investigation of AIC
- The Corrected Akaike Information Criterion, AICc
  - Introduction to AICc
  - Derivation of AICc
- Discussion
- Simulation Study
- The Modified Akaike Information Criterion, MAIC
- Discussion

Review of AIC

Key Constructs:

- **True** or **generating model**: $g(y)$.
- **Candidate** or **approximating model**: $f(y|\,\theta_k)$.
- **Candidate class:**

$$\mathcal{F}(k) = \{f(y|\,\theta_k) \mid \theta_k \in \Theta(k)\}\,.$$

- **Kullback discrepancy** between $g(y)$ and $f(y|\,\theta_k)$ with respect to $g(y)$:

$$d(\theta_k) = \mathrm{E}\{-2\ln f(y|\,\theta_k)\}.$$

Review of AIC

- **Fitted model:** $f(y|\hat{\theta}_k)$.
- **Kullback discrepancy** between $g(y)$ and $f(y|\hat{\theta}_k)$ with respect to $g(y)$:

$$d(\hat{\theta}_k) = \mathrm{E}\{-2\ln f(y|\theta_k)\}|_{\theta_k=\hat{\theta}_k}.$$

- **The Akaike information criterion:**

$$\mathrm{AIC} = -2\ln f(y|\hat{\theta}_k) + 2k.$$

Review of AIC

- The discrepancy

$$d(\hat{\theta}_k) = \mathrm{E}\{-2 \ln f(y|\, \theta_k)\}|_{\theta_k = \hat{\theta}_k}$$

  reflects the separation between the generating model $g(y)$ and a fitted model $f(y|\, \hat{\theta}_k)$.
- Under appropriate conditions, the expected value of AIC asymptotically approaches the expected value of $d(\hat{\theta}_k)$, say

$$\Delta(k) = \mathrm{E}\{d(\hat{\theta}_k)\}.$$

- Specifically, one can establish that

$$\mathrm{E}\{\mathrm{AIC}\} + o(1) = \Delta(k).$$

Review of AIC

- To justify the asymptotic unbiasedness of AIC, we impose the assumption that $g(y) \in \mathcal{F}(k)$.
- This assumption implies that the true model or density is a member of the parametric class $\mathcal{F}(k)$, and can therefore be written as $f(y|\,\theta_o)$, where $\theta_o \in \Theta(k)$.

Review of AIC

Consider writing $\Delta(k)$ as follows:

$$
\begin{aligned}
&\Delta(k) \\
&= \mathrm{E}\{d(\hat{\theta}_k)\} \\
&= \mathrm{E}\{-2 \ln f(y|\hat{\theta}_k)\} \\
&\quad + \Big[\mathrm{E}\{-2 \ln f(y|\theta_o)\} - \mathrm{E}\{-2 \ln f(y|\hat{\theta}_k)\}\Big] \qquad (1) \\
&\quad + \Big[\mathrm{E}\{d(\hat{\theta}_k)\} - \mathrm{E}\{-2 \ln f(y|\theta_o)\}\Big]. \qquad (2)
\end{aligned}
$$

Review of AIC

Lemma

$$\begin{aligned} E\{-2\ln f(y|\theta_o)\} - E\{-2\ln f(y|\hat{\theta}_k)\} &= k + o(1), \\ E\{d(\hat{\theta}_k)\} - E\{-2\ln f(y|\theta_o)\} &= k + o(1). \end{aligned}$$

- Conditions under which the lemma holds:
  - $g(y) \in \mathcal{F}(k)$,
  - the maximum likelihood vector $\hat{\theta}_k$ satisfies the conventional large-sample properties of MLE's.

Bias Investigation of AIC

- Under the assumption that $g(y) \in \mathcal{F}(k)$, AIC provides us with an approximately unbiased estimator of $\Delta(k)$ in settings where n is large and k is comparatively small.
- In settings where n is small and k is comparatively large (e.g., $k \approx n/2$), $2k$ is often much smaller than the bias adjustment, making AIC substantially negatively biased as an estimator of $\Delta(k)$.

Bias Investigation of AIC

- In the framework of normal linear regression, we will investigate the adequacy of the approximation of the bias adjustment by $2k$.
- When the candidate class $\mathcal{F}(k)$ consists of normal linear regression models, under the assumption that $g(y) \in \mathcal{F}(k)$, the bias adjustment terms (1) and (2) can be exactly evaluated for any values of n and k.
- The appropriate formulas will be later derived.

Bias Investigation of AIC

In the normal linear regression setting, the following tables list the exact values of the bias adjustment terms (1) and (2) for specific values of n and k.

k	n	(1)	(2)
3	320	3.01	3.06
5	320	5.04	5.15
9	320	9.13	9.45
17	320	17.46	18.56
3	160	3.03	3.13
5	160	5.08	5.31
9	160	9.26	9.94
17	160	17.97	20.34

Bias Investigation of AIC

k	n	(1)	(2)
3	80	3.06	3.26
5	80	5.16	5.65
9	80	9.55	11.03
17	80	19.11	24.76
3	40	3.11	3.55
5	40	5.34	6.43
9	40	10.19	13.81
17	40	22.12	39.70

Bias Investigation of AIC

k	n	(1)	(2)
3	30	3.15	3.77
5	30	5.46	7.04
9	30	10.69	16.31
17	30	25.06	59.94
3	20	3.24	4.26
5	20	5.74	8.55
9	20	11.93	24.07
17	20	37.60	302.40

The Corrected Akaike Information Criterion, AICc

- The corrected Akaike information criterion, AICc, was first suggested for normal linear regression by Sugiura (1978).
- AICc is defined by replacing the penalty term of AIC ($2k$) with the exact expression for the bias adjustment.
- Hurvich and Tsai (1989) demonstrated the small-sample superiority of AICc over AIC, and justified the use of AICc in the frameworks of nonlinear regression models and autoregressive models.

The Corrected Akaike Information Criterion, AICc

- In the last 20 years, AICc has been extended to a number of additional modeling frameworks, including the following:
  - autoregressive moving-average models (Hurvich, Shumway, and Tsai, 1990),
  - vector autoregressive models (Hurvich and Tsai, 1993),
  - multivariate linear regression models (Bedrick and Tsai, 1994),
  - generalized linear models with a dispersion parameter (Hurvich and Tsai, 1995),
  - models for longitudinal data analysis under the assumption of a known covariance structure (Azari, Li, and Tsai, 2006).

The Corrected Akaike Information Criterion, AICc

- In the framework of normal linear regression models (both univariate and multivariate), the penalty term of AICc provides an exact expression for the bias adjustment.
- In other frameworks for which AICc has been justified, the penalty term of AICc provides only an approximation to the bias adjustment (albeit an approximation that is generally more precise than $2k$).

The Corrected Akaike Information Criterion, AICc

- The advantage of AICc over AIC is that in small-sample applications, AICc estimates the expected discrepancy $\Delta(k)$ with less bias than AIC.
- The advantage of AIC over AICc is that AIC is more universally applicable, since the derivation of AIC is quite general whereas the derivation of AICc relies upon the form of the candidate model class $\mathcal{F}(k)$.

Derivation of AICc

- The derivation of AICc is based on the following lemma, to be formally established.
- We assume that the candidate class $\mathcal{F}(k)$ consists of normal linear regression models, and that $g(y) \in \mathcal{F}(k)$.
- As before, we write $g(y)$ as $f(y|\,\theta_o)$.
- We use p to denote the rank of the design matrix, meaning $k = (p+1)$.

Derivation of AICc

Lemma

$$E\{-2\ln f(y|\theta_o)\} - E\{-2\ln f(y|\hat{\theta}_k)\}$$
$$= n\ln(n/2) - n\psi\left(\frac{n-p}{2}\right),$$
$$E\{d(\hat{\theta}_k)\} - E\{-2\ln f(y|\theta_o)\}$$
$$= -n\ln(n/2) + n\psi\left(\frac{n-p}{2}\right) + \frac{2(p+1)n}{n-p-2}.$$

Derivation of AICc

- The statement of the preceding lemma involves the *digamma* or *psi* function $\psi(z)$.
- The digamma function $\psi(z)$ arises when evaluating the expectation of the log of a chi-squared random variable.
- $\psi(z)$ cannot be expressed in closed form, yet for $z > 1$, it can be approximated to any degree of accuracy by using an expansion:

$$\psi(z) = \ln(z) - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} - \frac{1}{252z^6} + \ldots.$$

Derivation of AICc

- Suppose that the generating model for the data is given by

$$y = X\beta_o + e, \qquad e \sim N_n(0, \sigma_o^2\ I),$$

and that the candidate model postulated for the data is of the form

$$y = X\beta + e, \qquad e \sim N_n(0, \sigma^2\ I).$$

- Here, y is an $n \times 1$ observation vector, e is an $n \times 1$ error vector, β_o and β are $p \times 1$ parameter vectors, and X is an $n \times p$ design matrix of full-column rank.

Derivation of AICc

- Let θ_o and θ_k respectively denote the $k = (p+1)$ dimensional vectors $(\beta_o', \sigma_o^2)'$ and $(\beta', \sigma^2)'$.
- Assume β_o is such that for some $0 < p_o \leq p$, the last $(p - p_o)$ components of β_o are zero.
- Thus, the true model is nested within the candidate model.
- Note that the nesting ensures that $f(y|\,\theta_o) \in \mathcal{F}(k)$.
- Let $\hat{\beta}$ denote the least-squares estimator of β, and let $\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/n$.
- Let $\hat{\theta}_k = (\hat{\beta}', \hat{\sigma}^2)'$ denote the MLE of θ_k.

Derivation of AICc

Preliminary Results:

- Let χ^2 be a random variable having a central chi-square distribution with d degrees of freedom.
- $$\mathsf{E}\left\{\frac{1}{\chi^2}\right\} = \frac{1}{d-2}.$$
- $$\mathsf{E}\left\{\ln \chi^2\right\} = \ln 2 + \psi\left(\frac{d}{2}\right).$$

Proof of AICc Lemma

Proof:

- The log-likelihood for the candidate model is given by

$$\ln f(y|\,\theta_k) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

- The following relations can be established:

$$\begin{aligned}
\mathrm{E}\{-2\ln f(y|\,\theta_o)\} &= n\ln\sigma_o^2 + n(1 + \ln 2\pi), \\
\mathrm{E}\{-2\ln f(y|\,\hat{\theta}_k)\} &= \mathrm{E}\{n\ln\hat{\sigma}^2\} + n(1 + \ln 2\pi), \\
d(\hat{\theta}_k) &= n\ln\hat{\sigma}^2 + \frac{n\sigma_o^2}{\hat{\sigma}^2} \\
&\quad + \frac{1}{\hat{\sigma}^2}(\hat{\beta} - \beta_o)'(X'X)(\hat{\beta} - \beta_o) + n\ln 2\pi.
\end{aligned}$$

Proof of AICc Lemma

To evaluate the expected value of $d(\hat{\theta}_k)$, note the following:

- $(n\hat{\sigma}^2/\sigma_o^2)$ has a chi-square distribution with $(n-p)$ degrees of freedom,
- the quadratic form $\{(\hat{\beta}-\beta_o)'\{(1/\sigma_o^2)(X'X)\}(\hat{\beta}-\beta_o)\}$ has a chi-square distribution with p degrees of freedom,
- $\hat{\sigma}^2$ and $\hat{\beta}$ are independent.

$$
\begin{aligned}
&\mathrm{E}\{d(\hat{\theta}_k)\} \\
&= \mathrm{E}\{n \ln \hat{\sigma}^2\} + \mathrm{E}\left\{\frac{n\sigma_o^2}{\hat{\sigma}^2}\right\} \\
&\quad + \mathrm{E}\left\{\frac{1}{\hat{\sigma}^2}\right\} \mathrm{E}\left\{(\hat{\beta} - \beta_o)'(X'X)(\hat{\beta} - \beta_o)\right\} + n \ln 2\pi \\
&= \mathrm{E}\{n \ln \hat{\sigma}^2\} + n \ln 2\pi + n^2 \mathrm{E}\left\{\frac{\sigma_o^2}{n\hat{\sigma}^2}\right\} \\
&\quad + n\mathrm{E}\left\{\frac{\sigma_o^2}{n\hat{\sigma}^2}\right\} \mathrm{E}\left\{(\hat{\beta} - \beta_o)'\{(1/\sigma_o^2)(X'X)\}(\hat{\beta} - \beta_o)\right\} \\
&= \left[\mathrm{E}\{n \ln \hat{\sigma}^2\} + n(1 + \ln 2\pi)\right] - n + n^2 \{1/(n - p - 2)\} \\
&\quad + n\{1/(n - p - 2)\}(p) \\
&= \left[\mathrm{E}\{n \ln \hat{\sigma}^2\} + n(1 + \ln 2\pi)\right] + \frac{2n(p+1)}{n - p - 2}.
\end{aligned}
$$

We now simplify the first and second terms of the bias adjustment.

$$
\begin{aligned}
&\mathrm{E}\{-2\ln f(y|\,\theta_o)\} - \mathrm{E}\{-2\ln f(y|\,\hat{\theta}_k)\} \\
&= \left\{n\ln\sigma_o^2 + n(1+\ln 2\pi)\right\} - \left\{\mathrm{E}\{n\ln\hat{\sigma}^2\} + n(1+\ln 2\pi)\right\} \\
&= \mathrm{E}\left\{n\ln\left(\frac{\sigma_o^2}{\hat{\sigma}^2}\right)\right\} \\
&= \mathrm{E}\left\{n\ln\left\{n\left(\frac{\sigma_o^2}{n\hat{\sigma}^2}\right)\right\}\right\} \\
&= n\ln n - n\mathrm{E}\left\{\ln\left(\frac{n\hat{\sigma}^2}{\sigma_o^2}\right)\right\} \\
&= n\ln n - n\left\{\ln 2 + \psi\left(\frac{n-p}{2}\right)\right\} \\
&= n\ln(n/2) - n\psi\left(\frac{n-p}{2}\right).
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}\{d(\hat{\theta}_k)\} - \mathrm{E}\{-2\ln f(y|\,\theta_o)\} \\
&= \left\{ \left[\mathrm{E}\{n\ln\hat{\sigma}^2\} + n(1+\ln 2\pi)\right] + \frac{2n(p+1)}{n-p-2} \right\} \\
&\quad - \left\{ n\ln\sigma_o^2 + n(1+\ln 2\pi) \right\} \\
&= \mathrm{E}\left\{ n\ln\left(\frac{\hat{\sigma}^2}{\sigma_o^2}\right)\right\} + \frac{2n(p+1)}{n-p-2} \\
&= \mathrm{E}\left\{ n\ln\left\{\frac{1}{n}\left(\frac{n\hat{\sigma}^2}{\sigma_o^2}\right)\right\}\right\} + \frac{2n(p+1)}{n-p-2} \\
&= -n\ln n + n\mathrm{E}\left\{\ln\left(\frac{n\hat{\sigma}^2}{\sigma_o^2}\right)\right\} + \frac{2n(p+1)}{n-p-2} \\
&= -n\ln n + n\left\{\ln 2 + \psi\left(\frac{n-p}{2}\right)\right\} + \frac{2n(p+1)}{n-p-2} \\
&= -n\ln(n/2) + n\psi\left(\frac{n-p}{2}\right) + \frac{2n(p+1)}{n-p-2}. \quad \square
\end{aligned}
$$

Derivation of AICc

- AICc is now obtained by adding the bias adjustment terms presented in the preceding lemma to the baseline estimator of $\Delta(k)$, $-2\ln f(y|\hat{\theta}_k)$.
- We have:

$$\begin{aligned}
\text{AICc} &= -2\ln f(y|\hat{\theta}_k) \\
&\quad + \left\{ n\ln(n/2) - n\psi\left(\frac{n-p}{2}\right)\right\} \\
&\quad + \left\{ -n\ln(n/2) + n\psi\left(\frac{n-p}{2}\right) + \frac{2(p+1)n}{n-p-2}\right\} \\
&= -2\ln f(y|\hat{\theta}_k) + \frac{2(p+1)n}{n-p-2}.
\end{aligned}$$

Discussion

- Since $k = (p+1)$, note that the penalty term of AICc can be written as

$$\frac{2(p+1)n}{n-p-2} = 2(p+1)\left(\frac{n}{n-p-2}\right) = 2k\left(\frac{n}{n-k-1}\right).$$

- When n is large and k is comparatively small, the penalty term of AICc is approximately $2k$.
- However, in settings where n is small and k is comparatively large, the fraction $\{n/(n-k-1)\}$ may be appreciably greater than one, making the penalty term of AICc considerably greater than the penalty term of AIC.
- In such settings, AICc often dramatically outperforms AIC as a model selection criterion.

Discussion

- From a practical standpoint, one should keep in mind that AICc has only been justified for certain modeling frameworks.
- In SAS, AICc is available for some of the frameworks in which the criterion has been justified (e.g., vector autoregressive modeling via PROC VARMAX), yet not for others (e.g., linear regression modeling via PROC REG).
- In PROC MIXED and PROC GLIMMIX, AICc is available. However, AICc has never been justified for mixed modeling or generalized linear mixed modeling under an unknown covariance structure.

Simulation Study

Simulation Setting:

- In each of four simulation sets, one thousand samples of size n are generated from a true regression model which has an n by $p_o = 4$ design matrix, a parameter vector of the form $\beta_o = (1, 1, 1, 1)'$, and a variance of $\sigma_o^2 = 4$.
- For every sample, candidate models with nested design matrices of ranks $p = 2, 3, \ldots, P = 11$ are fit to the data.
  - The first column of every design matrix is a vector of ones.
  - The design matrix of rank $p_o = 4$ is correctly specified.
- The covariates are generated as *iid* replicates from a $N(0, 4)$ distribution.

Simulation Study

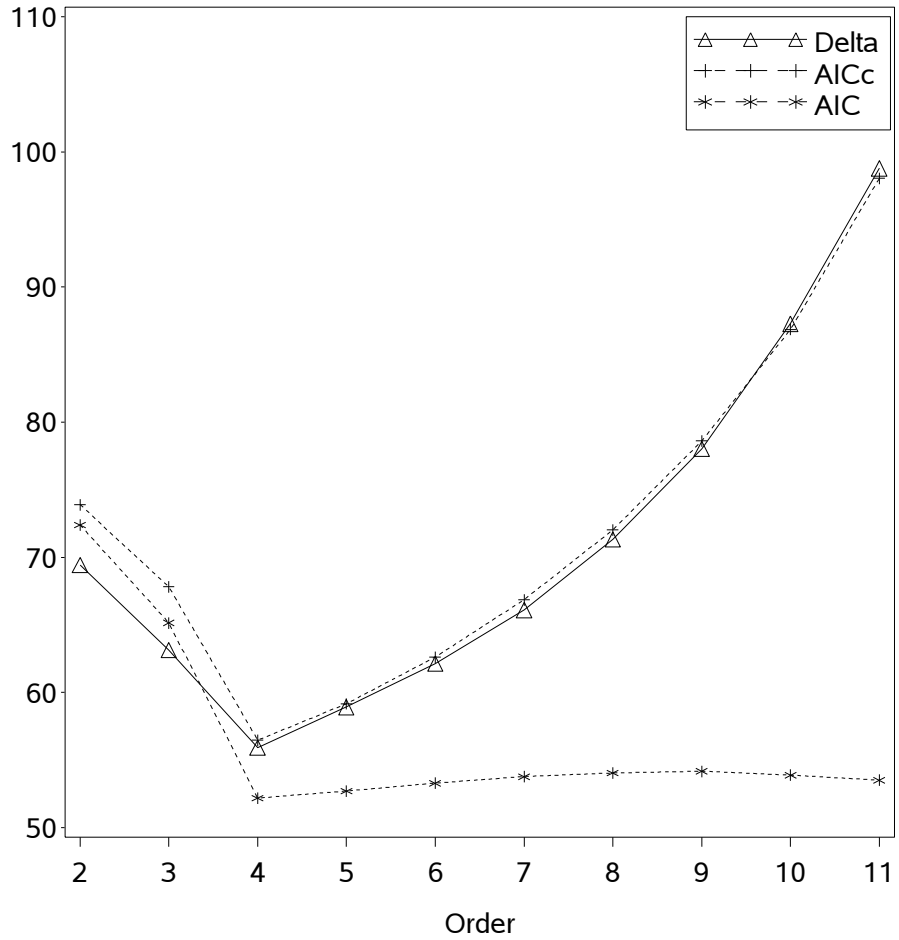
- We examine the effectiveness of AIC and AICc at selecting p.
- We refer to p as the *order* of the model.
- The sample sizes for the four sets are $n = 20$, $n = 50$, $n = 200$, and $n = 500$.

Simulation Study

Set I: Order selections for AIC and AICc, $n = 20$.

p	AIC	AICc
2	1	6
3	1	18
4	437	863
5	97	82
6	52	21
7	50	8
8	51	2
9	75	0
10	88	0
11	148	0

Figure: Plot of $\Delta(k)$ and the average values of AIC, AICc versus the order p for Set I. The figure illustrates the extreme negative bias of AIC.

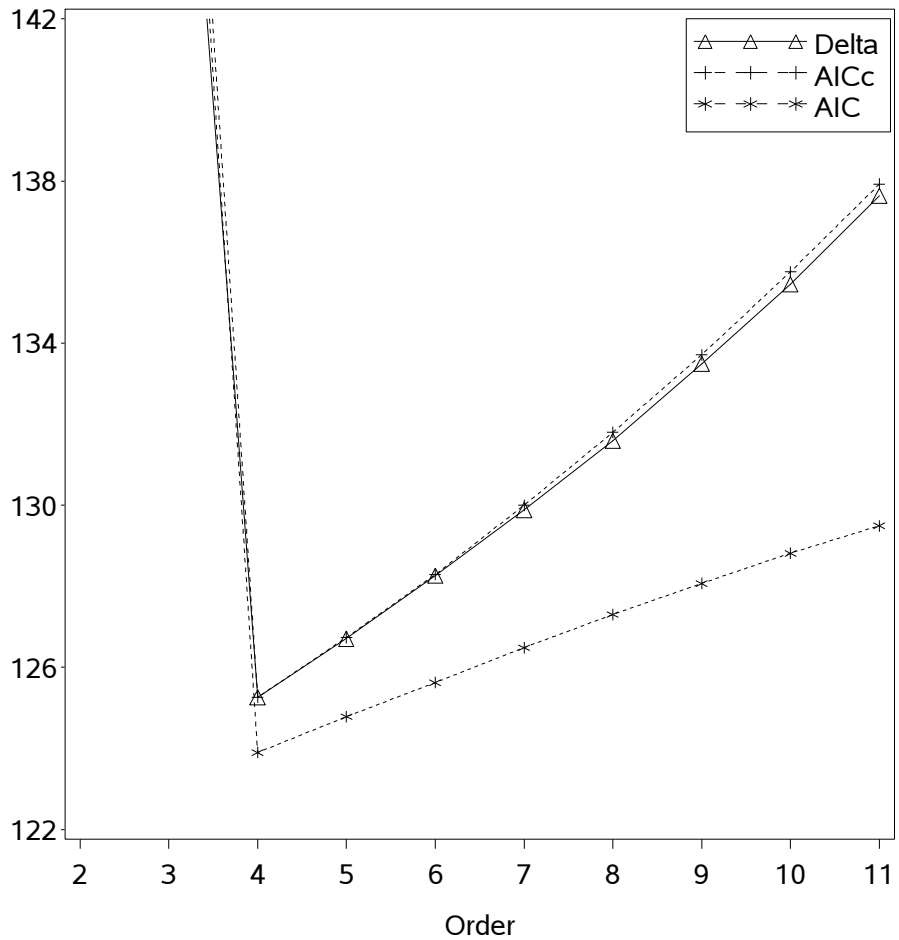


Simulation Study

Set II: Order selections for AIC and AICc, $n = 50$.

p	AIC	AICc
2	0	0
3	0	0
4	670	799
5	112	106
6	63	45
7	44	26
8	28	11
9	35	6
10	25	4
11	23	3

Figure: Plot of $\Delta(k)$ and the average values of AIC, AICc versus the order p for Set II. The figure illustrates the negative bias of AIC.

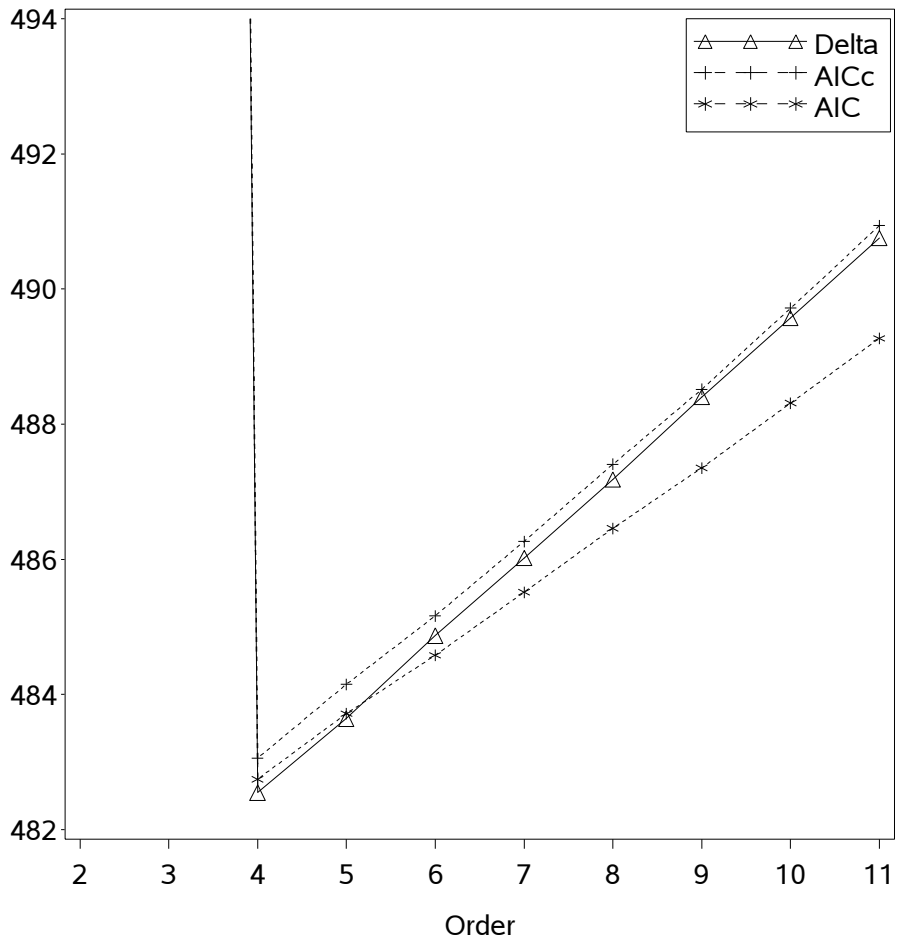


Simulation Study

Set III: Order selections for AIC and AICc, $n = 200$.

p	AIC	AICc
2	0	0
3	0	0
4	683	719
5	111	110
6	63	67
7	47	39
8	36	28
9	29	17
10	22	14
11	9	6

Figure: Plot of $\Delta(k)$ and the average values of AIC, AICc versus the order p for Set III.

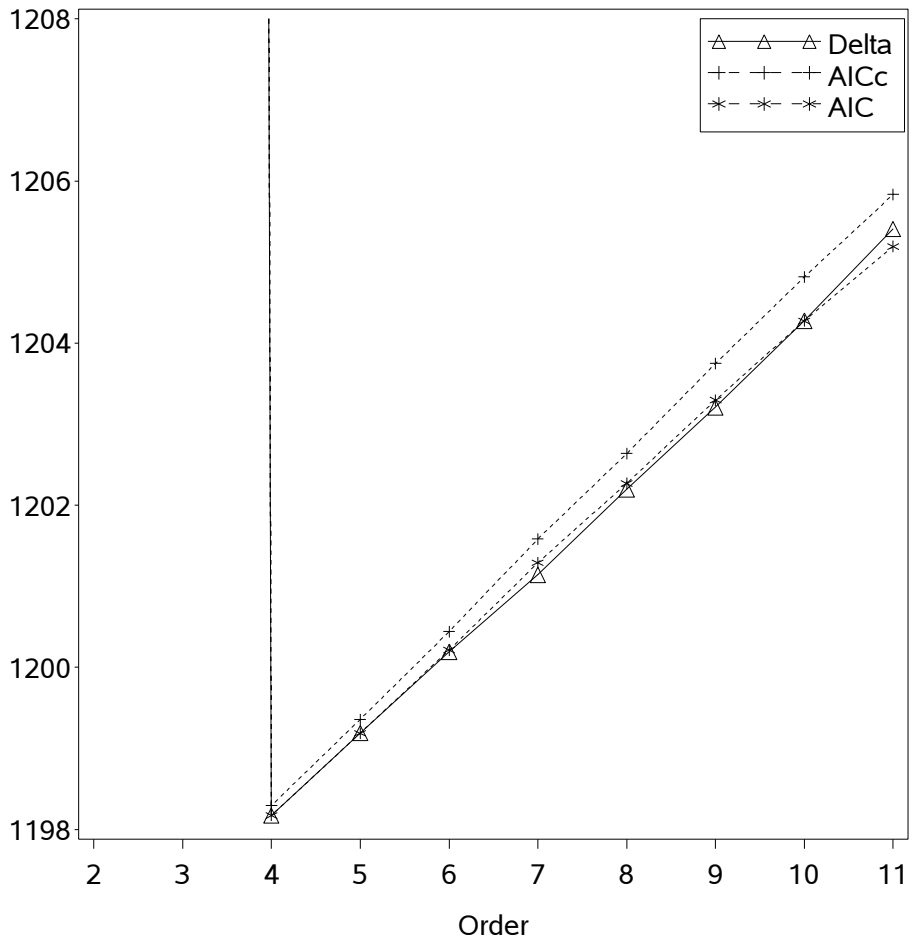


Simulation Study

Set IV: Order selections for AIC and AICc, $n = 500$.

p	AIC	AICc
2	0	0
3	0	0
4	743	755
5	103	107
6	58	55
7	31	28
8	24	21
9	14	13
10	9	8
11	18	13

Figure: Plot of $\Delta(k)$ and the average values of AIC, AICc versus the order p for Set IV.



The Modified Akaike Information Criterion, MAIC

- The asymptotic unbiasedness of AIC and exact unbiasedness of AICc (in the normal linear regression framework) require the assumption that $g(y) \in \mathcal{F}(k)$.
- This assumption implies that the candidate model of interest, $f(y|\,\theta_k)$, is either correctly specified or overspecified.
- The modified Akaike information criterion, MAIC, is one of several AIC variants based on a development that relaxes this assumption.
- MAIC was introduced for the framework of normal multivariate linear regression models by Fujikoshi and Satoh (1997).
- We will introduce MAIC in the framework of normal univariate linear regression models.

The Modified Akaike Information Criterion, MAIC

- Let $\mathcal{F} = \{\mathcal{F}(k_1), \mathcal{F}(k_2), \ldots, \mathcal{F}(k_L)\}$ represent the candidate family.
- Assume that the largest candidate class in $\mathcal{F}$ is $\mathcal{F}(K)$ (i.e., $K = \max\{k_1, k_2, \ldots, k_L\}$).
- MAIC is based on the assumption that $g(y) \in \mathcal{F}(K)$.
- Under this assumption, the largest candidate model, $f(y|\,\theta_K)$, is either correctly specified or overspecified.
- The candidate model of interest, $f(y|\,\theta_k)$, may be correctly specified, underspecified, or overspecified.

The Modified Akaike Information Criterion, MAIC

- In the regression framework, the model $f(y|\,\theta_K)$ would typically represent all covariates under consideration.
- Let P denote the rank of the design matrix for this model.
- Let σ_*^2 denote the error variance for this model.
- Let $\hat{\sigma}_*^2$ denote the maximum likelihood estimator of σ_*^2.
- Define MAIC as

$$\begin{aligned}\text{MAIC} &= -2\ln f(y|\,\hat{\theta}_k) + \frac{2n(p+1)}{(n-p-2)} \\ &\quad + \left[2p\left\{\frac{(n-p)\hat{\sigma}_*^2}{(n-P)\hat{\sigma}^2} - 1\right\} - 2\left\{\frac{(n-p)\hat{\sigma}_*^2}{(n-P)\hat{\sigma}^2} - 1\right\}^2\right].\end{aligned}$$

Discussion

- MAIC can be written as

$$\begin{aligned}\text{MAIC} &= \text{AICc} \\ &\quad + \left[2p\left\{\frac{(n-p)\hat{\sigma}_*^2}{(n-P)\hat{\sigma}^2}-1\right\}-2\left\{\frac{(n-p)\hat{\sigma}_*^2}{(n-P)\hat{\sigma}^2}-1\right\}^2\right].\end{aligned}$$

- The additional data-dependent penalization added to AICc is designed to be approximately zero for correctly specified and overfitted models.
- For underfitted models, this penalization is designed to reduce the bias of AICc.

Discussion

Bias Properties of AIC, AICc, MAIC
(Normal Linear Regression Models)

Fitted Model	AIC	AICc	MAIC
Underfitted	$O(1)$	$O(1)$	$O(1/n)$
Correctly Specified or Overfitted	$O(1/n)$	0	$O(1/n^2)$

Discussion

- Does the additional stochastic penalization added to AICc to produce MAIC yield a practical improvement?
- In simulation results featured in Fujikoshi and Satoh (1997), MAIC marginally outperforms AICc in terms of overfitted selections, yet performs the same as AICc in terms of underfitted selections.
- In the next lecture, we will examine the performance of AIC, AICc, and MAIC in a simulation study.
- We will also introduce a general criterion that provides an asymptotically unbiased estimator for $\Delta(k)$ without requiring $g(y) \in \mathcal{F}(k)$, the Takeuchi (1976) information criterion.

References

- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.
- Cavanaugh, J. E. (1997). Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, **33**, 201–208.
- Fujikoshi, Y., and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika* **84**, 707–716.