Evaluation of Information Criteria

Matt Oehler

March 8, 2018

Abstract

When performing an analysis statisticians often select a particular model based on how the model is scored in regards to one or various information criteria such as the Akaike Information Criterion or Bayesian Information Criterion. In this study we will assess the ability of these criteria to select the true model. This will be assessed by generating data under different covariance structures, and using different information criteria to select one of several plausible models. The accuracy and variance of the performance will be assessed through a Monte Carlo simulation study.

1 Introduction

When performing data analyses, statisticians often use one or several information criteria to determine which model to use. However, these different criteria are not necessarily guaranteed to select the true model. In this study we assessed the efficacy of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in choosing the true underlying model of simulated data. We performed this assessment by means of a Monte Carlo simulation study, allowing us to derive results and obtain a measure of uncertainty Since there are essentially infinite structures under which data can be simulated, we have limited the scope of this study to assess only a few scenarios. The details of which will be expounded upon in the following sections.

2 Model

We chose to limit the scope of this study by only assessing AIC and BIC performance with data sets that were generated under two main covariance structures: first-order autoregressive (AR(1)) and compound symmetric (CS). We chose to use these structures since they are commonly used when fitting mixed models, and because we suspected that these structures would be similar enough to potentially cause the AIC and/or BIC to choose the wrong model. This created a rigorous testing environment in which we could perform our study.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Figure 1: Autoregressive covariance structure

$$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}$$

Figure 2: Compound Symmetric covariance structure

The format of the AR(1) and CS covariance structures is displayed in figures 1 and 2 respectively. At first glance, the structures may not appear to be very similar, but it turns out that the CS structure can be rewritten in terms of an overall variance σ^2 and a correlation coefficient ρ to better correspond with the layout of the AR(1) covariance structure in figure 1. This can be seen more clearly in the matrix shown below.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

In this context, ρ (if positive) can be thought of as the proportion of the overall variance that can be attributed to the random effects.

3 Methods

The data used for this study were generated under a specific format. We generated data for two treatment groups and each group had an equal amount of subjects (we tested various numbers of subjects). The subjects were measured at 4 equally spaced time periods. We selected 10 as the true intercept value for both groups, 1 for the true slope value of the first group, and 1.5 for the true slope value of the second group. Lastly, we selected an overall variance of 2.5 and a correlation coefficient (we tested various correlations as well) . We selected the values of 1 and 1.5 for the slope terms, which are relatively close to each other, to maintain a more rigorous testing environment for AIC and BIC. For the sake of consistency, we chose to keep the true slope/intercept values constant throughout the study. The subject number/correlation combinations that we tested are shown table 1.

Correlation	Subject Number
0.10	10
0.40	10
0.80	10
0.10	40
0.40	40
0.80	40
0.10	80
0.40	80
0.80	80

Table 1: Correlation and subject number combinations that were tested

By performing a Monte Carlo simulation study, we iteratively (for 1000 iterations) generated data under each of the covariance structures and then fit both AR(1) and CS models to all of the generated data sets. Then we picked the best model according to AIC and BIC, and kept track of whether the selected models were consistent with the model under which the data were generated. We then generated 95% Monte Carlo confidence intervals to assess the uncertainty of the estimated values.

When fitting the models for this study, we had treatment group and measurement time as fixed effects, and we included subject as a random effect. This model format was held constant throughout the entire study. The following section contains the results from the simulation study.

4 Results

First we will look at the performance of AIC and BIC in choosing the correct model when data were generated under the CS covariance structure. The results for AIC are compiled into table 2.

Correlation	Subject Number	2.5%	Estimate	97.5%
0.10	10	0.47	0.50	0.53
0.40	10	0.48	0.51	0.54
0.80	10	0.47	0.50	0.53
0.10	40	0.49	0.52	0.55
0.40	40	0.48	0.51	0.54
0.80	40	0.49	0.52	0.55
0.10	80	0.46	0.49	0.52
0.40	80	0.46	0.49	0.52
0.80	80	0.47	0.50	0.53

Table 2: Results of AIC for data generated under CS covariance structure

In the estimate column of tables 2 - 5 the estimate indicates the proportion of times that the information criterion incorrectly picked the true model. It is interesting to note that for models generated under CS data, AIC only picks the true model about 50% of the time regardless of the subject number/correlation combination.

Next we will look at the performance of BIC choosing the correct model when data were generated under the CS covariance structure. The results are compiled into table 3.

Correlation	Subject Number	2.5%	Estimate	97.5%
0.10	10	0.47	0.50	0.53
0.40	10	0.48	0.51	0.54
0.80	10	0.47	0.50	0.53
0.10	40	0.49	0.52	0.55
0.40	40	0.48	0.51	0.54
0.80	40	0.49	0.52	0.55
0.10	80	0.46	0.49	0.52
0.40	80	0.46	0.49	0.52
0.80	80	0.47	0.50	0.53

Table 3: Results of AIC for data generated under CS covariance structure

After looking at these results, it is interesting to note that the proportion of times that BIC chose the incorrect model is identical to that of AIC when data are generated under the CS covariance structure. This probably indicates that the generated data are a bigger factor in determining the selected model more so than the specific information criterion. Perhaps it would have been better to evaluate different slope/intercept terms throughout the study to see if that would have any effect.

Lastly, we will look at the performance of AIC and BIC for data which were generated under the AR(1) covariance structure. Tables 4 and 5 show the results for AIC and BIC respectively.

Correlation	Subject Number	2.5%	Estimate	97.5%
0.10	10	0.47	0.50	0.53
0.40	10	0.26	0.29	0.31
0.80	10	0.12	0.14	0.16
0.10	40	0.36	0.39	0.42
0.40	40	0.08	0.10	0.11
0.80	40	0.00	0.01	0.01
0.10	80	0.29	0.32	0.35
0.40	80	0.01	0.02	0.03
0.80	80	0.00	0.00	0.00

Table 4: Results of AIC for data generated under AR(1) covariance structure

Correlation	Subject Number	2.5%	Estimate	97.5%
0.10	10	0.47	0.50	0.53
0.40	10	0.26	0.29	0.31
0.80	10	0.12	0.14	0.16
0.10	40	0.36	0.39	0.42
0.40	40	0.08	0.10	0.11
0.80	40	0.00	0.01	0.01
0.10	80	0.29	0.32	0.35
0.40	80	0.01	0.02	0.03
0.80	80	0.00	0.00	0.00

Table 5: Results of BIC for data generated under AR(1) covariance structure

For data generated under the AR(1) structure, AIC and BIC seemed to be more effective, especially for data sets where there was a larger sample size and/or a larger correlation. We can see this because as sample size and correlation increased, the estimate for proportion of incorrectly choosing the model decreased. Additionally, just as with data generated with a CS covariance structure, AIC and BIC performed identically, indicating that the nuance lies more so in the generated data generated than in the information criterion (at least for AIC and BIC in these scenarios).

5 Conclusions

The conclusions we reached in this study are limited due to the small scope of scenarios that we tested. This study could be expanded upon and improved in many ways. Based on our results we conclude that AIC and BIC perform identically in the scenarios that we tested. This may indicate that (at least for the scenarios we looked at) the performance of the information criteria is more dependent on the generated data rather than the nuances of the specific criterion. It also appears that AIC and BIC are more prone to correctly choose a true AR(1) model than they are to choose a true CS model. This result was intriguing and leads us to want to explore this result in more depth.