*Journal of the Korean*
*Statistical Society*
*Vol. 20, No. 2, 1991*

# Behrens-Fisher Problem from a Model Selection Point of View[†]

**Jong-Woo Jeon[*] and Kee-Won Lee[**]**

## ABSTRACT

Behrens-Fisher problem is viewed from a model selection approach. Normal distribution is regarded as an approximating model. A criterion, called TIC, is derived and is compared with selection criteria such as AIC and a bootstrap estimator. Stochastic approximation is used since no closed form expression is available for the bootstrap estimator.

Key Words : Model Selection, Approximating Model, Operating Model, AIC, TIC, Bootstrap, Stochastic Approximation

## 1. Introduction

Let $X_1, \cdots, X_m$ be an independent and identically distributed random variables with distribution function $F_1$ such that the mean of $F_1$, $\mu_1$ and the variance of $F_1$, $\sigma_1^2$ exist. Also let $Y_1, \cdots, Y_n$ be independent and identically distributed random variables with distribution function $F_2$ such that the mean of $F_2$, $\mu_2$ and the variance of $F_2$, $\sigma_2^2$ exist. It is known that $\sigma_1^2 \neq \sigma_2^2$. We want to know whether $\mu_1 = \mu_2$ or not. When F's are specifed as normally distributed, this problem reduces to Behrens-Fisher problem.

This kind of problem occurs when, for instance, a treatment that increases mean response may increase the variance of the response, but we may still want to compare the mean responses. If we use the model selection terminology, F's are regarded as operating models and the normal distributions are regarded as approximating models. Linhart and Zucchini(1986) gives an excellent introduction to the model selection idea.

In this paper a model selection approach is used to find out which submodel among the approximating models gives the best fit to the observed values. The problem is to choose the best fitting submodel using some widely used selection criteria among the followings ;

$$\text{Model}(3) : X \sim N(\mu, \sigma_1^2),\ Y \sim N(\mu, \sigma_2^2)\ (\mu_1 = \mu_2 = \mu),$$

† Research supported by the Basic Science Research Institute Program, Ministry of Education 1990−1991.
* Department of Computer Science and Statistics, Seoul National University, Seoul.
** Department of Statistics, Hallym University, Chunchon, Korea.

$$\text{Model}(4): X \sim N(\mu_1, \sigma_1^2),\ Y \sim N(\mu_2, \sigma_2^2)\ (\mu_1 \neq \mu_2),$$

where the number in the parentheses indicates the number of parameters of the model. Let $N=m+n$ denote the total number of observations obtained.

Among the selection criteria, AIC, introuduced by Akaike(1973), is given by

$$(1.1)\quad -2\ \max(\text{log-likelihood}) + 2p,$$

where p denotes the number of parameters to be estimated from the observations. We are to choose the submodel which gives the smallest AIC.

Assume further that the fourth moments of X's and Y's exist. This condition is necessary to get the trace version of the selection criterion. An asymptotic criterion called TIC, which is derived and fully explained in Linhart and Zucchini(1986) is given by

$$(1.2)\quad -2\ \max(\text{log-likelihood}) + 2\ \mathrm{tr}(\Omega_N^{-1}\Sigma_N),$$

where $\Omega_N$, and $\Sigma_N$ will be explained in section 2.2.

A bootstrap version of the criterion will be explained in section 2.3. Since no closed form expression is available in this case, a stochastic approximation version will be given as well. If we denote the parameters of the approximating model by $\theta$, and their maximum likelihood estimators by $\hat{\theta}$ respectively, then a key idea in using the bootstrap method is to use $\hat{\theta}^*$, $\hat{F}_1$ and $\hat{F}_2$ in place of $\hat{\theta}$, $F_1$ and $F_2$, respectively, where $\hat{\theta}^*$ is a bootstrap version of $\hat{\theta}$ computed from the bootstrap sample and $\hat{F}_1$ and $\hat{F}_2$ are the empirical cumulative functions based on $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ respectively.

## 2. The Selection Criteria

In this section each criterion will be derived explicitly for each submodel. For the bootstrap estimators, stochastic approximation algorithms are fully explained since no closed form expression is available. In addition to the above two submodels we may consider the case with the same variance at the same time. But in this paper our attention will be restricted to the unequal variance case only.

Note that when $\mu_1 \neq \mu_2$ the observed $-2\times$log-likelihood under the specified submodel in the approximating family is given by,

$$(2.1)\quad N \log 2\pi + m \log \sigma_1^2 + n \log \sigma_2^2 + \Sigma(x_i-\mu_1)^2/\sigma_1^2 + \Sigma(y_j-\mu_2)^2/\sigma_2^2.$$

Similarly, when $\mu_1 = \mu_2$, $-2\times$log-likelihood is given by,

$$(2.2)\quad N \log 2\pi + m \log \sigma_1^2 + n \log \sigma_2^2 + \Sigma(x_i-\mu)^2/\sigma_1^2 + \Sigma(y_j-\mu)^2/\sigma_2^2.$$

Throughout, maximum likelihood estimator refers to the one that minimizes this expression.

### 2.1 AIC

Chapter 7 of Sakamoto, et. al.(1986) gives explicit formulae. Let $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ be the maximum likelihood estimators of $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$ respectively. That is,

$$(2.2.1)\quad \hat{\mu}_1=\Sigma X_i/m,\ \hat{\sigma}_1^2=\Sigma(X_i-\hat{\mu}_1)^2/m,$$
$$\hat{\mu}_2=\Sigma Y_j/n,\ \hat{\sigma}_2^2=\Sigma(Y_j-\hat{\mu}_2)^2/n.$$

Then the following table gives a short summary of AIC.

**Table 2.1. Table of AIC**

| Model(p) | $-2$ max(log-likelihood) $+2p$ |
|---|---|
| $p=3$ | $N \log 2\pi+m \log \tilde{\sigma}_1^2+n \log \tilde{\sigma}_2^2+N+2\times 3$ |
| $p=4$ | $N \log 2\pi+m \log \hat{\sigma}_1^2+n \log \hat{\sigma}_2^2+N+2\times 4$ |

Here $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ are the maximum likelihood estimators of $\sigma_1^2$ and $\sigma_2^2$ when $\mu_1=\mu_2$, respectively. If we denote the maximum likelihood estimator of $\mu=\mu_1=\mu_2$ by $\tilde{\mu}$, then $\tilde{\mu}$, $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ can be obtained by solving the following simultaneous equations:

$$(2.2.2)\quad \mu^3-A\mu^2+B\mu-C=0,$$
$$\sigma_1^2=(\mu-\hat{\mu})^2+\hat{\sigma}_1^2,$$
$$\sigma_2^2=(\mu-\hat{\mu})^2+\hat{\sigma}_2^2,$$

where A, B and C are defined by

$$A=(1+m/N)\ \hat{\mu}_2+(1+n/N)\ \hat{\mu}_1,$$
$$B=2\hat{\mu}_1\ \hat{\mu}_2+m\Sigma y_j^2\,/\,nN+n\Sigma x_i^2\,/\,mN,$$
$$C=\hat{\mu}_1\ m\ \Sigma\ y_j^2\,/\,nN+\hat{\mu}_2\, n\,\Sigma x_i^2\,/\,mN,$$

respectively.

### 2.2 TIC

In order to derive the asymptotic criterion of the form (1.2), we just have to put the trace term in place of p in the above table. To get the trace term which, in some sense, is equivalent to the number of parameters, the results from one sample case can be used. Lee(1991) presents the details regarding the trace version of the selection criterion.

When $\mu_1=\mu_2$, it can be argued that

$$(2.2.3)\quad tr(\Omega_N^{-1}\Sigma_N)=\tilde{\mu}_{1.4}\,/\,2\tilde{\sigma}_1^4+\tilde{\mu}_{2.4}\,/\,2\tilde{\sigma}_2^4,$$

where $\tilde{\mu}_{1.4}=\Sigma(X_i-\tilde{\mu})^4\,/\,m$ and $\tilde{\mu}_{2.4}=\Sigma(Y_j-\tilde{\mu})^4\,/\,n$, respectively.
Therefore the criterion is given by

$$(2.2.4)\quad N \log 2\pi+m \log \tilde{\sigma}_1^2+n \log \tilde{\sigma}_2^2+N+\tilde{\mu}_{1.4}/\tilde{\sigma}_1^4+\tilde{\mu}_{2.4}/\tilde{\sigma}_2^4.$$

Note that the trace term is approximately the number of parameters if the approximating family is really close to the operating family.

When $\mu_1 \neq \mu_2$, it can be checked that

$$(2.2.5)\quad tr(\Omega_N^{-1}\Sigma_N)=1+\hat{\mu}_{1.4}/2\hat{\sigma}_1^4+\hat{\mu}_{2.4}/2\hat{\sigma}_2^4,$$

where $\hat{\mu}_{1.4}=\Sigma(X_i-\hat{\mu}_1)^4/m$ and $\hat{\mu}_{2.4}=\Sigma(Y_j-\hat{\mu}_2)^4/n$, respectively.
Therefore the criterion is given by

$$(2.2.6)\quad N\log 2\pi+m\log\hat{\sigma}_1^2+n\log\hat{\sigma}_2^2+N+2+\hat{\mu}_{1.4}/\hat{\sigma}_1^4+\hat{\mu}_{2.4}/\hat{\sigma}_2^4.$$

Note again that the trace term is approximately the number of parameters if the operating family is really close to the approximating family.

### 2.3 Bootstrap Version of the Criterion

A key idea in using the bootstrap method is to put the consistent estimators in place of unknown parameters. It is important to remember that $F_1$ and $F_2$, the distributions of future observations $Z_1$ and $Z_2$ respectively, are independent of X's and Y's respectively in the expression (2.2.1). This implies that $F_1$ and $F_2$ are independent of $\hat{\theta}=\hat{\theta}(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$. This idea will be explicit while we construct the bootstrap version of the selection criterion for each submodel.

When $\mu_1=\mu_2$, a bootstrap version of the criterion is given by

$$(2.3.1)\quad E_{\hat{\theta}^*}\{N\log 2\pi+m\log\tilde{\sigma}_1^{*2}+n\log\tilde{\sigma}_2^{*2}$$
$$+\Sigma(x_i-\tilde{\mu}^*)^2/\tilde{\sigma}_1^{*2}+\Sigma(y_j-\tilde{\mu}^*)^2/\tilde{\sigma}_2^{*2}\},$$

where the starred estimators are calculated from the bootstrap sample, say, $(X_1^*, \cdots, X_m^*, Y_1^*, \cdots, Y_n^*)$. These bootstrap samples are drawn in a parametric way from the fitted model of the approximating family. That is, draw $X_1^*, \cdots, X_m^*$ from $N(\tilde{\mu}, \tilde{\sigma}_1^2)$ and $Y_1^*, \cdots, Y_n^*$ from $N(\tilde{\mu}, \tilde{\sigma}_2^2)$, respectively.

It is easy to see that no closed form expression for (2.3.1) is available. One of the nice features of doing bootstrap is that a stochastic approximation is always available. By repeating the basic resampling procedure for a large number of times, say, B times, we may obtain B independent copies of the terms inside the curly bracket of (2.3.1). The average of these will approximates (2.3.1). Furthermore, the empirical cumulative distribution function of these will approximate the sampling distribution of the criterion. We may express the stochastic approximation to (2.3.1) as follows ;

$$(2.3.2)\quad 1/B\sum_{k=1}^{B}\{N\log 2\pi+m\log\tilde{\sigma}_{1.k}^{*2}+n\log\tilde{\sigma}_{2.k}^{*2}$$
$$+\Sigma(x_i-\tilde{\mu}_k^*)^2/\tilde{\sigma}_{1.k}^{*2}+\Sigma(y_j-\tilde{\mu}_k^*)^2/\tilde{\sigma}_{2.k}^{*2}\}.$$

Similarly, when $\mu_1 \neq \mu_2$, a bootstrap version of the criterion is given by

$$(2.3.3)\quad E_{\hat{\theta}^*}\{N\log 2\pi+m\log\hat{\sigma}_1^{*2}+n\log\hat{\sigma}_2^{*2}$$
$$+\Sigma(x_i-\hat{\mu}_1^*)^2/\hat{\sigma}_1^{*2}+\Sigma(y_j-\hat{\mu}_2^*)^2/\hat{\sigma}_2^{*2}\},$$

where the starred estimators are calculated from the bootstrap sample, say, $(X_1^*, \cdots, X_m^*, Y_1^*, \cdots, Y_n^*)$. These bootstrap samples are drawn in a parametric way from the fitted model of the approximating family. That is, draw $X_1^*, \cdots, X_m^*$ from $N(\hat{\mu}_1, \hat{\sigma}_1^2)$ and $Y_1^* \cdots, Y_n^*$ from $N(\hat{\mu}_2, \hat{\sigma}_2^2)$, respectively.

Again, no closed form expression for (2.3.3) is available. By repeating the basic resampling procedure for a large number of times, say, B times, we may obtain B of the terms inside the curly bracket of (2.3.3). The average of these will approximate (2.3.3). Furthermore, the empirical cumulative distribution function of these will approximate the sampling distribution of the criterion. We may express the stochastic approximation to (2.3.3) as follows ;

$$(2.3.4)\quad 1/B \sum_{k=1}^{B} \{N \log 2\pi + m \log \hat{\sigma}_{1.k}^{*2} + n \log \hat{\sigma}_{2.k}^{*2} + \Sigma(x_i - \hat{\mu}_{1.k}^*)^2/\hat{\sigma}_{1.k}^{*2} + \Sigma(y_j - \hat{\mu}_{2.k}^*)^2/\hat{\sigma}_{2.k}^{*2}\}.$$

Note. We may draw the bootstrap sample in a nonparametric way, that is, by drawing with replacement from the original sample. But as Lee(1991) shows, it often generates identical bootstrap sample in small sample cases to give unsatisfactory results. Therefore our attention will be restricted to parametric bootstrap.

## 3. Numerical Study

In this section a Monte Carlo study is provided to compare the selection criteria previously discussed. Normal distributions are selected as operating families. Sample sizes are set at (5, 5), (5, 10), (10, 10), (20, 10), (20, 20), and $\mu$'s are set at (0, 0) and (0, 2). To set up an unequal variance situation, $\sigma_1/\sigma_2$ is set at 2 and 4.

When $\mu_1=\mu_2$, Lehmann(1983) states that the combined estimator, say, $\bar{\mu}=(m\ \hat{\mu}_1+n\ \hat{\mu}_2)/N$ is also an asymptotically efficient estimator. Therefore, in addition to the maximum likelihood estimator, $\tilde{\mu}$, which is obtained from solving the simultaneous equation, $\bar{\mu}$, will also be used for comparison.

To compare the performance, the number of times each criterion selects the correct submodel will be computed. The average values and standard errors are also recorded to assess the overall sampling variation.

This simulation study is run on Micro Vax II, and the total CPU time used is about 20 hours. The number of bootstrap samples are set at 100, and the total number of replications are set at 1000. IMSL library function, *ggnml* is used to generate the normal random numbers. To solve the cubic equation in (2.2.2) the exact solution formula from a mathematical handbook is used, since Newton-Raphson type iterative method sometimes fails to converge.

Table 1 and 2 give the results when the actual means are the same with different standard deviations as the sample sizes vary. Table 3, 4 and 5 give the results when the actual means are different in addition.

Table 1 shows the results when $\mu_1=\mu_2$, $\sigma_1=1$, $\sigma_2=2$. Hence it is expected that the average values of the criteria for Model(3) will be smaller than those for Model(4). It can be observed that $\tilde{\mu}$ gives more satisfactory results than $\bar{\mu}$ for most of the cases. Exceptions are the cases of bootstrap version with sample sizes (5, 5) and (5, 10). Regarding the avearge values $\tilde{\mu}$

## Tables

Note. Standard errors are around 0.2~0.3. PCS denotes the percentage of correct selection out of 1000 replications.

**Table 1. Average value of the criterion and the percentage of correct selection from 1000 replications when $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 0, 1, 2)$.**

| (m, n) | criterion | p=3, $\tilde{\mu}$ | p=3, $\bar{\mu}$ | p=4 | PCS($\tilde{\mu}$) | PCS($\bar{\mu}$) |
|---|---|---|---|---|---|---|
| (5, 5) | AIC | 38.87 | 38.77 | 38.59 | 78.8% | 63.6% |
| | TIC | 36.06 | 37.01 | 36.57 | 70.1% | 57.3% |
| | Bootstrap | 43.88 | 40.51 | 45.80 | 87.4% | 92.0% |
| (5, 10) | AIC | 58.98 | 59.63 | 59.70 | 79.4% | 65.4% |
| | TIC | 57.59 | 58.20 | 58.16 | 74.5% | 60.6% |
| | Bootstrap | 61.51 | 59.72 | 62.59 | 80.9% | 84.1% |
| (10, 10) | AIC | 73.58 | 74.33 | 74.41 | 80.9% | 68.9% |
| | TIC | 72.56 | 73.35 | 73.34 | 79.2% | 66.9% |
| | Bootstrap | 72.83 | 73.43 | 73.54 | 78.8% | 66.2% |
| (20, 10) | AIC | 101.86 | 102.37 | 102.73 | 81.9% | 71.9% |
| | TIC | 101.08 | 101.59 | 101.88 | 78.7% | 71.0% |
| | Bootstrap | 100.12 | 100.89 | 100.83 | 78.8% | 64.6% |
| (20, 20) | AIC | 143.71 | 145.07 | 144.71 | 83.6% | 74.0% |
| | TIC | 143.09 | 144.55 | 144.06 | 83.0% | 72.8% |
| | Bootstrap | 141.61 | 143.30 | 141.94 | 74.9% | 53.9% |

Note : This table is based on 100 bootstrap samples and 1000 total number of replications when the true models are N(0, 1) and N(0, $2^2$), respectively.

**Table 2. Average value of the criterion and the percentage of correct selection from 1000 replications when $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 0, 1, 4)$.**

| (m, n) | criterion | p=3, $\tilde{\mu}$ | p=3, $\bar{\mu}$ | p=4 | PCS($\tilde{\mu}$) | PCS($\bar{\mu}$) |
|---|---|---|---|---|---|---|
| (5, 5) | AIC | 44.82 | 47.54 | 45.45 | 77.6% | 40.3% |
| | TIC | 43.01 | 45.71 | 43.41 | 69.4% | 36.1% |
| | Bootstrap | 49.80 | 49.61 | 52.53 | 90.7% | 75.1% |
| (5, 10) | AIC | 73.00 | 75.00 | 73.76 | 80.0% | 46.7% |
| | TIC | 71.55 | 73.50 | 72.23 | 76.9% | 41.6% |
| | Bootstrap | 75.70 | 75.74 | 76.54 | 80.2% | 64.6% |
| (10, 10) | AIC | 87.45 | 89.75 | 88.23 | 79.7% | 48.4% |
| | TIC | 86.35 | 88.62 | 87.09 | 77.7% | 46.9% |
| | Bootstrap | 86.53 | 90.04 | 87.25 | 78.8% | 26.2% |
| (20, 10) | AIC | 115.99 | 118.51 | 116.85 | 80.6% | 48.9% |
| | TIC | 115.24 | 117.67 | 116.04 | 79.3% | 48.2% |
| | Bootstrap | 114.12 | 118.62 | 114.98 | 80.3% | 15.6% |
| (20, 20) | AIC | 171.60 | 174.25 | 172.38 | 78.4% | 49.8% |
| | TIC | 171.01 | 173.60 | 171.77 | 77.5% | 48.8% |
| | Bootstrap | 169.44 | 174.36 | 169.58 | 70.6% | 3.8% |

Note : This table is based on 100 bootstrap samples and 1000 total number of replications when the true models are N(0, 1) and N(0, $4^2$), respectively.

**Table 3. Average value of the criterion and the percentage of correct selection from 1000 replications when $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 2, 1, 2)$.**

| (m, n) | criterion | p=3, $\bar{\mu}$ | p=3, $\bar{\mu}$ | p=4 | PCS($\bar{\mu}$) | PCS($\bar{\mu}$) |
|---|---|---|---|---|---|---|
| (5, 5) | AIC | 40.84 | 43.60 | 38.41 | 72.8% | 82.8% |
| | TIC | 38.99 | 41.77 | 36.41 | 76.1% | 86.1% |
| | Bootstrap | 46.46 | 45.19 | 45.58 | 53.5% | 44.5% |
| (5, 10) | AIC | 66.86 | 66.16 | 59.86 | 90.6% | 93.2% |
| | TIC | 65.18 | 64.47 | 58.33 | 90.6% | 93.0% |
| | Bootstrap | 68.61 | 66.03 | 63.04 | 83.4% | 71.5% |
| (10, 10) | AIC | 80.52 | 83.55 | 74.82 | 93.1% | 95.3% |
| | TIC | 79.40 | 82.28 | 73.79 | 92.4% | 95.0% |
| | Bootstrap | 79.70 | 82.35 | 73.85 | 93.5% | 94.9% |
| (20, 10) | AIC | 108.55 | 114.35 | 102.33 | 94.7% | 97.7% |
| | TIC | 107.51 | 113.37 | 115.53 | 93.8% | 97.4% |
| | Bootstrap | 106.77 | 112.86 | 100.48 | 94.6% | 98.3% |
| (20, 20) | AIC | 157.20 | 163.19 | 145.41 | 99.5% | 99.9% |
| | TIC | 156.42 | 162.21 | 144.86 | 99.6% | 99.8% |
| | Bootstrap | 155.05 | 161.12 | 142.61 | 99.9% | 100.0% |

Note: This table is based on 100 bootstrap samples and 1000 total number of replications when the true models are N(0, 1) and N(2, $2^2$), respectively.

**Table 4. Average value of the criterion and the percentage of correct selection from 1000 replications when $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 2, 1, 4)$.**

| (m, n) | criterion | p=3, $\bar{\mu}$ | p=3, $\bar{\mu}$ | p=4 | PCS($\bar{\mu}$) | PCS($\bar{\mu}$) |
|---|---|---|---|---|---|---|
| (5, 5) | AIC | 45.91 | 49.69 | 45.40 | 42.4% | 72.0% |
| | TIC | 44.05 | 47.76 | 43.40 | 48.7% | 74.6% |
| | Bootstrap | 51.03 | 51.74 | 52.75 | 23.9% | 39.2% |
| (5, 10) | AIC | 75.24 | 79.05 | 73.65 | 55.6% | 82.0% |
| | TIC | 73.75 | 77.35 | 72.09 | 57.8% | 93.1% |
| | Bootstrap | 77.90 | 79.39 | 76.54 | 54.7% | 65.8% |
| (10, 10) | AIC | 89.73 | 96.13 | 88.28 | 59.1% | 83.3% |
| | TIC | 88.67 | 94.77 | 87.19 | 60.8% | 83.0% |
| | Bootstrap | 88.78 | 95.90 | 87.29 | 60.0% | 92.0% |
| (20, 10) | AIC | 117.74 | 125.93 | 116.37 | 57.0% | 82.1% |
| | TIC | 116.93 | 125.06 | 115.53 | 59.2% | 82.1% |
| | Bootstrap | 115.84 | 125.64 | 114.43 | 57.8% | 94.1% |
| (20, 20) | AIC | 176.34 | 187.98 | 172.96 | 76.5% | 93.9% |
| | TIC | 175.72 | 187.06 | 172.40 | 76.9% | 93.8% |
| | Bootstrap | 174.17 | 187.05 | 170.18 | 84.1% | 99.8% |

Note: This table is based on 100 bootstrap samples and 1000 total number of replications when the true models are N(0, 1) and N(2, $4^2$), respectively.

**Table 5. Average value of the criterion and the percentage of correct selection from 1000 replications when the parameter values are given by $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 1, 1, 2)$ and $(\mu_1, \mu_2, \sigma_1, \sigma_2)=(0, 1, 1, 4)$**

| $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ | (m, n) | | p=3, | p=4 | PCS |
|---|---|---|---|---|---|
| (0, 1, 1, 2) | (5, 10) | AIC | 61.06 | 59.72 | 49.2% |
| | | TIC | 59.58 | 58.12 | 53.4% |
| | | Bootstrap | 63.27 | 62.75 | 42.9% |
| | (20, 10) | AIC | 103.86 | 102.36 | 57.5% |
| | | TIC | 103.10 | 101.60 | 58.7% |
| | | Bootstrap | 102.12 | 100.48 | 59.2% |
| (0, 1, 1, 4) | (5, 10) | AIC | 73.54 | 73.76 | 30.0% |
| | | TIC | 72.07 | 72.23 | 33.5% |
| | | Bootstrap | 76.47 | 77.03 | 29.2% |
| | (20, 10) | AIC | 116.79 | 116.98 | 30.7% |
| | | TIC | 116.07 | 116.25 | 31.9% |
| | | Bootstrap | 114.95 | 115.12 | 31.4% |

gives more consistent results than $\tilde{\mu}$. The only exception is observed for the case of AIC with sample sizes (5, 5). With $\bar{\mu}$, bootstrap gives somewhat consistent results but its behaviour gets worse as the sample sizes increase. Among the three criteria it is not easy to tell which one gives the best performance. But the performance of bootstrap with $\bar{\mu}$ in small samples is notable.

Table 2 shows the results when $\mu_1=\mu_2$, $\sigma_1=1$, $\sigma_2=4$. With larger variance of Y values, it is observed that $\bar{\mu}$ gives worse results than the previous case. The case of the bootstrap criterion with sample sizes (20, 20) gives surprisingly poor result, only 3.8% of correct selection ! Again, it is noted that the performance of bootstrap with $\bar{\mu}$ gets worse as the sample sizes increase. Regarding the average values of the criteria, similar comments to the previous one can be made.

Table 3 and 4 shows the results when $\mu_1=0$, $\mu_2=2$. $\sigma_1$ is still 1 and $\sigma_2$ is set at 2 and 4 respectively. It is expected that the average values of the critera for Model(4) will be smaller than those for Model(3). It is observed that the overall perfomance gets better as the sample sizes increase. Now $\bar{\mu}$ outperforms $\tilde{\mu}$ in many a case.

What if $\mu_2$ is closer to $\mu_1$ than before ? Table 5 gives some results when $\mu_1$ is still 0 but $\mu_2$ is now 1 instead of 2. The overall performance gets worse than before. As $\sigma_2$ gets larger the perfomance gets even worse.

## References

1. Akaike, H. (1973). Information theroy and an extension of the maximum likelihood principle. In Petrov and Czaki, eds., Proceedings *of the 2nd International Symposium of Information Theory*, 267-281.
2. Lee, K-W. (1991). On the model selection criteria of normal approximating models., Unpublished manuscript.
3. Lehmann, E.L. (1983). *Theory of Point Estimation*, Wiley, New York.
4. Linhart, H. and Zucchini, W. (1986). *Model Selection*, Wiley, New York.
5. Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. Reidel, Dordrecht.