High-dimensional properties of AIC, BIC and $C_p$ for estimation of dimensionality in canonical correlation analysis

Yasunori Fujikoshi

(Received November 1, 2016; Revised January 29, 2017)

Abstract.
This paper is concerned with consistency properties of the dimensionality estimation criteria AIC, BIC and $C_p$ in CCA (Canonical Correlation Analysis) between $p$ variables and $q$ ($\leq p$) variables, based on a sample of size $N = n + 1$. The consistency properties of the criteria are studied under a high-dimensional asymptotic framework such that $p$ and $n$ tend to infinity satisfying $p/n \to c \in [0,1)$, and under two types of assumptions on the order of the population canonical correlations, where $q$ is fixed. We note that there are cases that the criteria based on AIC and $C_p$ are consistent, but the criterion based on BIC is not consistent. Through a Monte Carlo simulation experiment, our results are checked numerically, and the estimation criteria are compared.

AMS 2010 Mathematics Subject Classification. 62H12, 62H30.
Key words and phrases. AIC, BIC, Canonical correlation analysis, $C_p$, Consistency property, Dimensionality, High-dimensional asymptotic framework.

§1. Introduction
In this paper we are concerned with the dimensionality estimation method by use of the model selection criteria AIC (Akaike (1973)), BIC (Schwarz (1978)) and $C_p$ (Mallows (1973)) in CCA (canonical correlation analysis) with two random vectors of $p$ and $q$ ($q \leq p$) components, based on the sample size $N = n + 1$. The criteria based on AIC and $C_p$ were proposed by Fujikoshi and Veitch (1979), and the criterion based on BIC was studied by Gunderson and Muirhead (1997). It is known in a large-sample asymptotic framework that AIC and $C_p$ are not consistent, but BIC is consistent. For these results, see Fujikoshi (1985), Gunderson and Muirhead (1997).
However, recently it is known that there is a situation such that AIC and $C_p$ have a consistency property, but BIC is not consistent, when the number $p$ of response variables and the sample size $n$ are large under a high-dimensional framework such that $p/n \to c \in [0, 1)$. These results can be found in Fujikoshi, Sakurai and Yanagihara (2014) and Yanagihara, Wakaki and Fujikoshi (2015) for selection of variables in multivariate regression model. Further, such properties have been shown in Fujikoshi and Sakurai (2016) for model selection criteria of estimating the dimensionality in multivariate linear model.

In this paper we consider asymptotic properties of AIC, BIC and $C_p$ for estimation of dimensionality in CCA under a high-dimensional asymptotic framework such that

\begin{equation}
q \text{ is fixed, } p \to \infty, \ n \to \infty, \ p/n \to c \in [0, 1).
\end{equation}

It is shown that the AIC and $C_p$ for estimation of dimensionality have consistency properties under two types of assumptions on population canonical correlations, but BIC is not consistent under one type of assumptions on population canonical correlations. It may be noted that these properties are different from the ones in a large-sample case, since in general AIC and $C_p$ have a positive probability of selecting each of the overspecified models. Our results are checked numerically by conducting a Monte Carlo simulation experiment. Further, we compare with the selection rates of the three criteria.

The present paper is organized as follows. In Section 2, we summarize the criteria for estimating the dimensionality in CCA. High-dimensional properties of the criteria are given in Section 3. In Section 4 we check our theoretical results by conducting a Monte Carlo simulation experiment, and compare with the selection rates of the three criteria. In Section 5, we discuss our conclusions and future subjects. The proofs of our results are given in Appendix.

\section{The criteria for estimation of dimensionality}

Let

\begin{equation}
\begin{pmatrix}
x_1 \\
y_1 \\
\vdots \\
x_N
\end{pmatrix}, \ldots, \begin{pmatrix}
x_N \\
y_N
\end{pmatrix}
\end{equation}

be a sample of size $N = n + 1$ of $(x', y')'$ from $(p + q)$-dimensional normal distribution $N_{q+p}(\mu, \Sigma)$, with $x : p \times 1$ and $y : q \times 1$. Let $S$ be the sample covariance matrix formed from the sample. In this paper we assume that $q \leq p$ without loss of generality. Corresponding to a partition $(x', y')$, we partition $\mu, \Sigma$ and $S$ as

$$\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, \quad \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}, \quad \begin{pmatrix}
S_{11} & S_{12} \\
S_{21} & S_{22}
\end{pmatrix}.$$
Let \( \rho_1 \geq \cdots \geq \rho_q \geq 0 \) and \( r_1 > \cdots > r_q > 0 \) be the population and the sample canonical correlations between \( x \) and \( y \). Then \( \rho_1^2 \geq \cdots \geq \rho_q^2 \geq 0 \) and \( r_1^2 > \cdots > r_q^2 > 0 \) are the characteristic roots of \( \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \) and \( S_{11}^{-1} S_{12} S_{22}^{-1} S_{21} \), respectively.

We are interested in the number of nonzero population canonical correlations, which is called the dimensionality in canonical correlation analysis. Related to the estimation of the dimensionality we consider a dimensionality model:

\[
M_j : \quad \rho_j > \rho_{j+1} = \cdots = \rho_q = 0, \\
\iff \operatorname{rank}(\Sigma_{12}) = j. \quad (2.2)
\]

If \( M_j \) is true, we can explain the correlation structure between \( x \) and \( y \) by the first \( j \) canonical correlation variables, since the remaining canonical variables have no power of prediction.

Based on the likelihood of \( S \), it is known (Fujikoshi and Veitch (1979)) that AIC for \( M_j \) is given by

\[
\text{AIC}_j = -\sum_{i=j+1}^q n \log(1 - r_i^2) + n(p + q) + (p + q + 1) \log |S| + K \\
+ 2 \left\{ j(p + q - j) + \frac{1}{2} p(p + 1) + \frac{1}{2} q(q + 1) \right\},
\]

where \( K = 2 \log \left\{ \Gamma_{p+q}(\frac{1}{2}/(\frac{1}{2} n)^{(1/2)n(p+q)}) \right\} \). Instead of AIC, we may use

\[
A_j = \text{AIC}_j - \text{AIC}_q \\
= -\sum_{i=j+1}^q n \log(1 - r_i^2) - 2(p - j)(q - j), \quad j = 0, \ldots, q. \quad (2.4)
\]

Here \( A_q = 0 \). The BIC and \( C_{p,j} \) corresponding to \( A_j \) are given by

\[
B_j = -\sum_{i=j+1}^q n \log(1 - r_i^2) - (\log n)(p - j)(q - j), \quad j = 0, \ldots, q, \quad (2.5)
\]

\[
C_{p,j} = n \sum_{i=j+1}^q \frac{r_i^2}{1 - r_i^2} - 2(p - j)(q - j), \quad j = 0, \ldots, q. \quad (2.6)
\]

Here \( B_q = 0 \) and \( C_{p,q} = 0 \). Note that the \( A_j \), \( B_j \) and \( C_{p,j} \) based on the likelihood of \( \mathbf{X} = (x_1, \ldots, x_N)' \) and \( \mathbf{Y} = (y_1, \ldots, y_N)' \) can be expressed as the ones in (2.4), (2.5) and (2.6) replaced \( n \) by \( N \).

The estimation methods based on \( A_j \), \( B_j \) and \( C_{p,j} \) are expressed as

\[
\hat{j}_A = \arg\min_{j \in \mathcal{J}} A_j, \quad \hat{j}_B = \arg\min_{j \in \mathcal{J}} B_j, \quad \hat{j}_C = \arg\min_{j \in \mathcal{J}} C_{p,j},
\]

respectively.
§3. High-dimensional properties

We denote the model $M_j$ in (2.2) by $j$ simply. Then, the set of all the models is $F = \{0, 1, \ldots, q\}$. It is assumed that the true covariance matrix is

$$
\Sigma^* = \begin{pmatrix}
\Sigma_{11}^* & \Sigma_{12}^* \\
\Sigma_{21}^* & \Sigma_{22}^*
\end{pmatrix}
$$

with $\text{rank}(\Sigma_{12}^*) = j^*$. Then, the true dimension is $j^*$, where $0 \leq j^* \leq q$, and the model $M_{j^*}$ means the minimum dimensionality model including $M^*$. We separate $F$ into two sets, one is a set of overspecified models, i.e., $F_+ = \{j^*, j^* + 1, \ldots, q\}$ and the other is a set of underspecified models, i.e., $F_- = F^* \cap F = \{0, 1, \ldots, j^* - 1\}$. Further, we denote the set of models deleting the true model from $F_+$ by $F_+ \setminus \{j^*\}$, i.e., $F_+ \setminus \{j^*\} = \{j^* + 1, \ldots, q\}$.

When we treat the distributions of the canonical correlations themselves or their function such as $\hat{\rho}_A$, $\hat{\rho}_B$ and $\hat{\rho}_C$, without loss of generality we may assume that

$$
(3.1) \quad \Sigma = \begin{pmatrix}
I_p & \mathcal{R} \\
\mathcal{R} & I_q
\end{pmatrix},
\quad \mathcal{R} = (\mathcal{R}_1, 0), \quad \mathcal{R}_1 = \text{diag}(\rho_1, \ldots, \rho_q).$$

The number of possible nonzero canonical correlations is $q$. We will consider the transformed population and sample canonical correlations defined by

$$
(3.2) \quad \gamma_j = \frac{\rho_j}{(1 - \rho_j^2)^{1/2}}, \quad d_j = \frac{r_j}{(1 - r_j^2)^{1/2}}, \quad j = 1, \ldots, q.
$$

Depending on our results, the following assumptions are taken up:

**B1** (The true model and dimension $j^*$): The true model is that the samples in (2.1) are independently and identically distributed as $N_{q+p}(\mu, \Sigma)$ with $\mu = \mu^*$ and $\Sigma = \Sigma^*$. The true dimensionality is $j^*$, and the true canonical correlations are

$$
\rho_1^* \geq \cdots \geq \rho_{j^*} > \rho_{j^*+1} = \cdots = \rho_q^* = 0.
$$

**B2** (The asymptotic framework): $q$ is fixed,

$p \to \infty$, $n \to \infty$, $p/n \to c \in [0, 1]$.

**B3** (The canonical correlations-1): Let $\gamma_j^* = \rho_j^* / \{1 - (\rho_j^*)^2\}, j = 1, \ldots, q$. For any $i$ ($1 \leq i \leq j^*$),

$$
\rho_i^* = O(1), \quad \gamma_i^* = O(1) \quad \text{and} \quad \lim_{p/n \to c} \rho_i^* = \rho_0^*.
$$

**B4** (The canonical correlations-2): For any $i$ ($1 \leq i < j^*$),

$$
\gamma_i^* = \sqrt{p} \theta_i^* = O(\sqrt{p}) \quad \text{and} \quad \lim_{p/n \to c} \theta_i^* = \theta_{0i}^*.
$$
In B3 and B4, it is assumed that the multiplicities of the $\rho_i$’s do not depend on $p$ and $q$. In this paper, the true dimension $j_*$ is fixed, though our results shall be generalized for the case $\lim j_* = j_{*0}$. In the following, we give sufficient conditions for the three criteria to be consistent. Here, the consistency of, e.g., $\hat{j}_\Lambda$ means that the probability that $\hat{j}_\Lambda$ selects the true dimension tends asymptotically to 1, i.e.,

$$\lim_{p/n \to c} \Pr(\hat{j}_\Lambda = j_*) = 1.$$ 

Here, the notation $\lim_{p/n \to c}$ is used as an abbreviation for the asymptotic framework (1.1).

**Theorem 3.1.** Suppose that assumptions B1 and B2 are satisfied. Further, assume that $c \in [0, c_a)$, where $c_a (\approx 0.797)$ is the constant satisfying $\log(1 - c_a) + 2c_a = 0$.

1. $\hat{j}_\Lambda$ is consistent if B3 and the inequality $\frac{1}{\log(1 - (\rho_j^*)^2)} - 1 > c(1 - 2c)$ are satisfied.

2. $\hat{j}_\Lambda$ is consistent if assumption B4 is satisfied.

**Theorem 3.2.** Suppose that assumptions B1 and B2 are satisfied.

1. BIC is not consistent if assumptions B2 with $c > 0$ and B3 are satisfied.

2. BIC is consistent if assumptions B2 and B4 are satisfied.

**Theorem 3.3.** Suppose that assumptions B1 and B2 are satisfied. Further, assume that $c \in [0, 1/2)$.

1. $\hat{j}_C$ is consistent if B3 and the inequality $(\rho_{j,0}^*)^2 \left\{1 - (\rho_{j,0}^*)^2\right\}^{-1} > c(1 - 2c)$ are satisfied.

2. $\hat{j}_C$ is consistent if assumption B4 is satisfied.

From the proofs of Theorems 3.1 and 3.3 we can see that the AIC and $C_p$ criteria on the dimensionality in canonical correlation analysis satisfy the followings:

(i) if $c \in [0, c_a)$, $\lim_{p/n \to c} \Pr(\hat{j}_\Lambda \in \mathcal{F}_+ \setminus \{j_*\}) = 0$,

(ii) if $c \in [0, 0.5)$, $\lim_{p/n \to c} \Pr(\hat{j}_C \in \mathcal{F}_+ \setminus \{j_*\}) = 0$. 

(3.3) (3.4)
These results hold without the assumptions on the order of population canonical correlations.

Under a large-sample framework; $n \to \infty$ and $B3$ it is known that

\[
\lim_{n \to \infty} \Pr(\hat{j}_A = j) = \lim_{n \to \infty} \Pr(\hat{j}_C = j) = h(j|j_*),
\]

where for $j = 0, 1, \ldots, j* - 1$, $h(j|j_*) = 0$, and for $j = j_*, \ldots, q$, $h(j|j_*)'$s are positive and for their explicit expressions, see Fujikoshi (1985). Gunderson and Muirhead (1997) extended the result to the case of an elliptical distribution.

§ 4. Numerical study

In this section, we numerically examine the validity of some of our claims, and point some tendencies for the dimensionalities estimated by AIC, BIC and $C_p$.

In our simulation setting, $q = 5$ and the true dimensionality is $j_* = 3$. We consider two types of population canonical correlations. The first type is:

(a); $\rho^*_1 = 2\rho$, $\rho^*_2 = \frac{3}{2}\rho$, $\rho^*_3 = \rho$, $\rho^*_4 = \rho^*_5 = 0$,

where

$$\rho = \sqrt{\frac{(4p)/(21)}{p+1+(4p)/(21)}}.$$  

Then

$\rho^*_1 \to 0.8$, $\rho^*_2 \to 0.6$, $\rho^*_3 \to 0.4$.

The second type is defined in terms of $\gamma^*_i = \rho^*_i/\sqrt{1-(\rho^*_i)^2}$ as follows:

(b); $\gamma^*_1 = \frac{\bar{\rho}}{\sqrt{1-(\bar{\rho})^2}}$, $\gamma^*_2 = \frac{3}{4}\gamma^*_1$, $\gamma^*_3 = \frac{1}{2}\gamma^*_1$, $\gamma^*_4 = \gamma^*_5 = 0$,

where

$$\bar{\rho} = \sqrt{\frac{p}{p+1}} \cdot \sqrt{\frac{(4p)/(21)}{1+(4p)/(21)}}.$$  

In this case

$\gamma^*_1/\sqrt{p} \to 0.8$, $\gamma^*_2/\sqrt{p} \to 0.6$, $\gamma^*_3/\sqrt{p} \to 0.4$,

and

(b); $\rho^*_i = \frac{\gamma^*_i}{\sqrt{1+(\gamma^*_i)^2}}$, $i = 1, 2, 3$, $\rho^*_4 = \rho^*_5 = 0$.

The cases (a) and (b) correspond to the canonical correlations under assumptions B3 and B4. Our experiments were done for $n = 6p$ and $p = 5, 10, 20, 35, 50, 80, 100$. In this case, $p/n = 1/6 \to c = 1/6$. $\in [0, c_0]$ and
c ∈ [0, 1/2). Further, the inequalities in Theorem 3.1(1) and Theorem 3.3.1 have been satisfied for case (a). Our simulation results are given in Tables 1-6. We use the selection probabilities as the relative frequencies.

Table 1. Selection rates under AIC for case (a)

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>M_0</th>
<th>M_1</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>M_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.00</td>
<td>0.14</td>
<td>0.48</td>
<td>0.31</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.00</td>
<td>0.03</td>
<td>0.47</td>
<td>0.45</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.43</td>
<td>0.55</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.39</td>
<td>0.61</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
<td>0.64</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.31</td>
<td>0.69</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.70</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2. Selection rates under BIC for case (a)

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>M_0</th>
<th>M_1</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>M_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.18</td>
<td>0.55</td>
<td>0.24</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.11</td>
<td>0.74</td>
<td>0.15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.06</td>
<td>0.91</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.04</td>
<td>0.96</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.04</td>
<td>0.96</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.06</td>
<td>0.94</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.09</td>
<td>0.91</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3. Selection rates under C_p for case (a)

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>M_0</th>
<th>M_1</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>M_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.00</td>
<td>0.06</td>
<td>0.43</td>
<td>0.41</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.00</td>
<td>0.01</td>
<td>0.31</td>
<td>0.58</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
<td>0.73</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
<td>0.83</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
<td>0.88</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.94</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.96</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 4. Selection rates under AIC for case (b)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>$M_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.02</td>
<td>0.21</td>
<td>0.46</td>
<td>0.26</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
<td>0.75</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.95</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.99</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5. Selection rates under BIC for case (b)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>$M_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.43</td>
<td>0.42</td>
<td>0.13</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.05</td>
<td>0.45</td>
<td>0.45</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.00</td>
<td>0.01</td>
<td>0.69</td>
<td>0.31</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 6. Selection rates under $C_p$ for case (b)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$M_0$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>$M_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5</td>
<td>0.00</td>
<td>0.10</td>
<td>0.45</td>
<td>0.37</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.74</td>
<td>0.18</td>
<td>0.01</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.86</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>210</td>
<td>35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.94</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>300</td>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.98</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>480</td>
<td>80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>600</td>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

From Tables 1, 3, 4 and 6 we can see that AIC and $C_p$ are consistent for the estimation of dimensionality in the high-dimensional settings considered. On the speed of convergence to the true dimension, the case of canonical correlations-2 is faster than the one of canonical correlations-1. Further, in the case of canonical correlations-1 $C_p$ is faster than AIC. These criteria have a tendency of underestimating the dimensionality when $(n, p)$ is not so large.
From Tables 2 and 5 we can see that BIC is not consistent in the case of canonical correlations-1, but is consistent in the case of canonical correlations-2.

§5. Concluding remarks

In general, it is known that under a large sample asymptotic framework; \( n \to \infty \) and \( p \) and \( q \) are fixed, AIC and \( C_p \) have no consistency property, in the sense that the probabilities of selecting the true model do not approach to one, but BIC is consistent. However, in this paper, we demonstrated that the AIC and \( C_p \) for estimating the dimensionality in canonical correlation analysis have a consistency property, under a high-dimensional framework (1.1). For the consistency, it is required to satisfy some additional assumptions. For AIC, it needs that \( c \in [0, c_a) \), where \( c_a \approx 0.797 \). For \( C_p \), it needs that \( c \in [0, 1/2) \). Further, the consistency was considered under two types of assumptions on the population canonical correlations. On the other hand, in a high-dimensional case, we note that BIC is not always consistent.

In this paper we assume that our sample of size \( N = n + 1 \) comes from \((p + q)\)-dimensional normal distribution \( N_{p+q}(\mu, \Sigma) \). Further, we assume that for the random vectors of \( p \) and \( q \) \((\leq p)\), \( q \) is finite, and \( p \) tends to infinity, satisfying \( p/n \to c \in [0, 1) \). It is expected to remove the normality assumption, and to consider consistency properties of AIC, BIC and \( C_p \) under a general high-dimensional asymptotic framework such that \( p/n \to c \in [0, 1) \) and \( q/n \to d \in [0, 1) \).

§6. Appendix: The proofs of Theorems 3.1, 3.2 and 3.3

For the proofs of Theorems 3.1, 3.2 and 3.3, it is basic to establish the limiting behavior of the squares \( r_1^2 > \cdots > r_q^2 \) of the canonical correlations under a high-dimensional framework. The following results follows from the limiting distributions of \( r_1^2 > \cdots > r_q^2 \) given by Fujikoshi (2017) (For some special cases, see Fujikoshi and Sakurai (2008)).

Lemma 6.1. Let \( r_1^2 > \cdots > r_q^2 \) and \( \rho_1^2 \geq \cdots \geq \rho_q^2 \) be the squares of the sample and the population canonical correlations between \( x; p \times 1 \) and \( y; q \times 1 \) with \( p \geq q \), based on a sample of size \( N = n + 1 \) from \( N_{p+q}(\mu, \Sigma) \). Let \( d_j^2 = r_j^2/(1-r_j^2) \), \( \gamma_j^2 = \rho_j^2/(1-\rho_j^2) \), \( j = 1, \ldots, q \). We assume that the number of nonzero population canonical correlations is \( a \), and hence \( \rho_1 \geq \cdots \geq \rho_a > \rho_{a+1} = \cdots = \rho_q = 0 \), and the multiplicities of \( \rho_i \)'s do not depend on \( p \) and...
Suppose that for the limiting behaviors of $r_1^2 > \cdots > r_q^2$ and $d_1^2 > \cdots > d_q^2$ under a high-dimensional asymptotic framework

$$p \to \infty, \quad n \to \infty, \quad p/n \to c \in [0,1),$$

we have the following results:

1. Suppose that for any $i \ (1 \leq i \leq a)$, $\rho_i^2 = O(1)$ and

$$\lim_{p/n \to c} \rho_i^2 = \rho_{0i}^2 > 0; \quad \lim_{p/n \to c} \gamma_{i0}^2 = \rho_{i0}/(1 - (\rho_{i0})^2)^{1/2}, i = 1, \ldots, a. \quad \text{Then}$$

$$r_i^2 \to r_{0i}^2 + c(1 - \rho_{0i}^2), \quad d_i^2 \to c(1 - \gamma_{i0}^2); \quad i = 1, \ldots, a,$$

$$r_i^2 \to c, \quad d_i^2 \to c; \quad i = a + 1, \ldots, q.$$

2. Suppose that for any $i \ (1 \leq i \leq a)$, $\gamma_i^2 = \theta_i^2 = O(p)$ and

$$\lim_{p/n \to c} \theta_i = \theta_{i0} > 0. \quad \text{Then}$$

$$\frac{1}{p} d_i^2 \to c, \quad d_i^2 \to c; \quad i = a + 1, \ldots, q.$$

**The proofs of Theorems 3.1 and 3.2**

We consider a general criterion including $A_j$ and $B_j$ defined by

$$G_{\nu_n,j} = - \sum_{i=j+1}^{q} n \log(1 - r_i^2) - \nu_n(p - j)(q - j), \quad j = 0, \ldots, q,$$

with $\nu_n > 0$. Note that $G_{2,j} = A_j$ and $G_{\log,n,j} = B_j$. Let $d_j^2 = r_j^2/(1 - r_j^2)$. Using $1 - r_j^2 = (1 + d_j^2)^{-1}$, the difference between $G_j$ and $G_{j,s}$ is expressed in terms of $d_1^2, \ldots, d_q^2$ as follows. For $j > j_s$,

$$G_{\nu_{n,j}} - G_{\nu_{n,j,s}} = -n \log((1 + d_{j+1}^2) \cdots (1 + d_q^2)) + \nu_n(j - j_s)(p + q - j - j_s),$$

and for $j < j_s$,

$$G_{\nu_{n,j}} - G_{\nu_{n,j,s}} = n \log((1 + d_{j+1}^2) \cdots (1 + d_{j_s}^2)) + \nu_n(j - j_s)(p + q - j - j_s).$$

From Lemma 6.1 we have that for $j_s < i \leq q$, $d_i^2 \to c/(1 - c)$ in both cases B3 and B4. Therefore we have that for $j > j_s$,

$$\frac{1}{p} (G_{\nu_{n,j}} - G_{\nu_{n,j,s}}) - (j - j_s)\nu_n \to (j - j_s)c^{-1} \log(1 - c).$$
Noting that \( \lim_{c \to 0^+} c^{-1} \log(1 - c) = -1 \), we have that in both cases B3 and B4, for any \( j > j^* \),

\[
(6.2) \quad \frac{1}{p} (A_j - A_{j^*}) \overset{p}{\rightarrow} (j - j^*) \left\{ \frac{1}{c} \log(1 - c) + 2 \right\} > 0,
\]

if \( c \in [0, c_a) \), and for any \( j > j^* \),

\[
(6.3) \quad \frac{1}{p \log n} (B_j - B_{j^*}) \overset{p}{\rightarrow} (j - j^*) > 0.
\]

Next consider the case \( j < j^* \). Noting that under B3

\[
(6.4) \quad 1 + d_i^2 \overset{p}{\rightarrow} \frac{1}{1 - c} \left\{ 1 - \left( \rho_{i0}^* \right)^2 \right\}^{-1}, \quad \text{for } i < j^*,
\]

we have

\[
\frac{1}{p} \left( G_{v_n,j} - G_{v_n,j^*} \right) + (j^* - j) \nu_n \overset{p}{\rightarrow} - (j^* - j) c^{-1} \log(1 - c) - c^{-1} \sum_{i=j+1}^{j*} \log \left\{ 1 - \left( \rho_{i0}^* \right)^2 \right\}^2
\]

\[
\geq (j^* - j) \left\{ - \frac{1}{c} \log(1 - c) - \frac{1}{c} \log \left\{ 1 - (\rho_{j*}^*)^2 \right\} \right\}.
\]

Combining (6.2), (6.3) and the above results, we can get Theorem 3.1 (1) and Theorem 3.2 (1). Next we assume B4 and \( j < j^* \). For \( j < j^* \), we have

\[
\frac{1}{n} (G_{v_n,j} - G_{v_n,j^*}) \overset{p}{\rightarrow} (j^* - j) \nu_n (p + q - j^* - j) \equiv D_{v_n,j}
\]

Further, for \( i < j^* \), \( (1/p) d_i^2 \overset{p}{\rightarrow} (1 - c)^{-1}(\theta_i^*)^2 \), and hence

\[
\log(1 + d_i^2) - \log p \overset{p}{\rightarrow} \log(1 - c)^{-1}(\theta_i^*)^2.
\]

Therefore, we have

\[
\frac{1}{\log p} D_{2,j} \overset{p}{\rightarrow} j - j^*, \quad \text{and} \quad \frac{1}{\log p} D_{\log n,j} \overset{p}{\rightarrow} (j - j^*)(1 - c),
\]

which implies Theorems 3.1 and 3.2. Here, for the last result, we use

\[
\frac{p}{n} \left( \frac{\log n}{\log p} \right) = \frac{p}{n} \left( \frac{\log n}{\log n + \log(p/n)} \right) \to c.
\]
The proof of Theorem 3.3

Theorem 3.3 is proved by a similar line as in the proof of Theorem 3.1 as follows. Note that

\[ C_{p,j} - C_{p,j_*} = -n \sum_{i=j_*+1}^{j} d_i^2 + 2(j - j_*)(p + q - j - j_*), \quad \text{for } j > j_*, \]

\[ C_{p,j} - C_{p,j_*} = n \sum_{i=j+1}^{j_*} d_i^2 - 2(j_* - j)(p + q - j - j_*), \quad \text{for } j < j_*, \]

where \( d_i^2 = r_i^2/(1-r_i^2) \). Under both cases B3 and B4, we can see that by using \( d_i^2 \overset{P}{\to} c(1-c)^{-1} \) for \( i > j_* \),

\[ \frac{1}{p} (C_{p,j} - C_{p,j_*}) \overset{P}{\to} (j - j_*) \left( -\frac{1}{1-c} + 2 \right), \quad \text{for } j > j_. \]

Next consider the case \( j < j_* \). Using (6.4), under B3 we have

\[ \frac{1}{n} (C_{p,j} - C_{p,j_*}) \overset{P}{\to} \sum_{i=j_*+1}^{j_*} (1-c)^{-1} \left( c + (\gamma_{i_0})^2 \right) - 2c(j_* - j) \]

\[ \geq (j - j_*) \left[ (1-c)^{-1} \left( c + (\gamma_{i_0})^2 \right) - 2c \right]. \]

Combining this result with (6.5), we get Theorem 3.3(1). Under B4, \((1/p)d_i^2 \overset{P}{\to} (1-c)^{-1}(\theta_{i,0})^2\) for \( i < j_* \), and it is easy to see

\[ \frac{1}{np} (C_{p,j} - C_{p,j_*}) \overset{P}{\to} \frac{1}{1-c} \sum_{i=j_*+1}^{j_*} (\theta_{i,0})^2. \]

Combining this result with (6.5), we get Theorem 3.3(2).

Acknowledgments

The author would like to express his gratitude to the referee and the editor for their many valuable comments and suggestions. The author also would like to express his gratitude to Dr. Tetsuro Sakurai for his help of simulation study. The author’s research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), 16K00047, 2016-2018.
References


Yasunori Fujikoshi
Department of Mathematics, Graduate School of Science, Hiroshima University
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8626, Japan