EFFECTIVE SIGNIFICANCE LEVEL, WHEN AIC IS EMPLOYED IN PRE-TEST ESTIMATION

Rep. Fac. Sci., Kagoshima Univ., (Math., Phys. & Chem.)
No. 15, p. 11–18, 1982

EFFECTIVE SIGNIFICANCE LEVEL, WHEN AIC IS EMPLOYED IN PRE-TEST ESTIMATION

By

Kôichi INADA* and Fumiyoshi KASAGI**

(Received August 31, 1982)

Abstract

In the present paper we evaluate the effective significance level when AIC is employed in pre-test estimation. Numerical values of the significance level are presented in Table 1–3.

1. Introduction

The estimation after preliminary test of significance has been studied by various authors. The earlier works include, among others, Bancroft [3], Asano [2] and Kitagawa [6]. The central issue in this type of problem is how to determine the significance level of the preliminary test. We cite the works of Sawa and Hiromatsu [7]. Hirano [5] applied AIC (Akaike's information criterion [1]) to determine when to pool and when not to pool. AIC is equal to $-2\log_e \mathrm{L}(\hat{\mu})+2\mathrm{k}$, where $L(\hat{\mu})$ is the maximum likelihood and k is the number of unknown parameters.

Assume we have two models H_0 and H_1, which we ought to determine before the esimator is numerically calculated, and we ought to select the model which has smaller value of AIC, and upon selecting the model we compute the maximum likelihood estimate assuming this model. When this principle is employed, it is clear we will arrive the maximum likelihood estimate under the model chosen, out of two, by the preliminary likelihood ratio test. The consideration on distribution does not determine the critical value but the likelihood function and the difference in the numbers of parameters determine it, and thus the effective significance level of the preliminary test is determined automatically.

The purpose of this paper is to evaluate the effective significance level. Assuming that the correlation is known, we consider in §2 estimation of mean vector and in §3 that of one component of the means, upon introducing certain types of pair of model. In §4, estimation of correlation coefficient is considered.

* Department of Mathematics, Faculty of Science, Kagoshima University, Kagoshima, Japan.
** National Cardiovascular Center Research Institute, Osaka, Japan.

2. Estimation of bivariate normal mean

Let $\binom{X_1}{Y_1}$, $\binom{X_2}{Y_2}$, $\ldots$, $\binom{X_n}{Y_n}$ be a random samples from a bivariate normal distribution $N(\mu, \Lambda)$ with the mean vector $\mu=\binom{\mu_1}{\mu_2}$ and the covariance matrix $\Lambda=\sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$.

2.1 One sided case

From the nature of the data it is known that $\mu_1=\mu_2=0$ or ($\mu_1\geqq 0$ and $\mu_2\geqq 0$) where the inequality is strict for at least one.

Here we have two alternative models about means;

Model H_0: $\mu_1=\mu_2=0$,

Model H_1: $\mu_1\geqq 0$ and $\mu_2\geqq 0$ where the inequality is strict for at least one.

Then we would determine to use the procedure minimizing AIC as the test criterion of the preference between Model H_0 and Model H_1. We denote AIC under H_0 and under H_1 as AIC (H_0) and AIC (H_1) respectively.

Case 1. When σ^2 and ρ are known parameters, we make use of the following relation.

$$AIC(H_0)-AIC(H_1)<0$$

$$\text{(1)} \qquad \Leftrightarrow \begin{cases} n\chi^2<4,\ \bar{X}\geqq 0,\ \bar{Y}\geqq 0 \\ \text{or } \dfrac{n}{1-\rho^2}(\bar{Y}-\rho\bar{X})^2<4,\ \bar{X}<0,\ \bar{Y}-\rho\bar{X}\geqq 0 \\ \text{or } \dfrac{n}{1-\rho^2}(\bar{X}-\rho\bar{Y})^2<4,\ \bar{Y}<0,\ \bar{X}-\rho\bar{Y}\geqq 0 \\ \text{or } \bar{X}-\rho\bar{Y}<0,\ \bar{Y}-\rho\bar{X}<0 \end{cases}$$

where $\chi^2=\dfrac{1}{1-\rho^2}(\bar{X}^2-2\rho\bar{X}\,\bar{Y}+\bar{Y}^2)$ and $(\bar{X},\bar{Y})$ is the sample mean vector. That is to say, if (1) is satisfied we would accept Model H_0, and the estimator of μ is the zero vector. On the other hand if (1) is not satisfied, the estimator $\hat{\mu}$ of μ is as follows;

$$\text{(2)} \qquad \hat{\mu}=\begin{cases} \begin{pmatrix}\bar{X}\\ \bar{Y}\end{pmatrix} & \text{if } \bar{X}\geqq 0 \text{ and } \bar{Y}\geqq 0, \\ \begin{pmatrix}0\\ \bar{Y}-\rho\bar{X}\end{pmatrix} & \text{if } \bar{X}<0 \text{ and } \bar{Y}-\rho\bar{X}\geqq 0, \\ \begin{pmatrix}\bar{X}-\rho\bar{Y}\\ 0\end{pmatrix} & \text{if } \bar{X}-\rho\bar{Y}\geqq 0 \text{ and } \bar{Y}<0, \\ \begin{pmatrix}0\\ 0\end{pmatrix} & \text{if } \bar{X}-\rho\bar{Y}<0 \text{ and } \bar{Y}-\rho\bar{X}<0. \end{cases}$$

When H_0 is true, the significance level α of the test criterion stated above is exactly

$$\text{(3)} \qquad \alpha=1-Pr\{AIC(H_0)-AIC(H_1)<0\,|\,H_0\}$$

$$=\frac{e^{-2}(\pi-\cos^{-1}\rho)}{2\pi}+\frac{1}{\sqrt{2\pi}}\int_2^\infty e^{-x^2/2}\,dx.$$

In Table 1 values of α are tabulated for varicus values of ρ.

Case 2. When σ^2 is an unknown parameter and ρ is a known parameter, after some calculations, we arrive at the simple result as follows;

$$\mathrm{AIC}(H_0)-\mathrm{AIC}(H_1)<0$$

$$(4)\qquad \Leftrightarrow \begin{cases} \dfrac{n(\bar{X},\bar{Y})\,\Sigma^{-1}(\bar{X},\bar{Y})'}{\sum_{i=1}^{n}(X_i,Y_i)\,\Sigma^{-1}(X_i,Y_i)'}<1-e^{-2/n},\ \bar{X}\geqq 0,\ \bar{Y}\geqq 0 \\ \text{or } \dfrac{n(0,\bar{Y}-\rho\bar{X})\,\Sigma^{-1}(0,\bar{Y}-\rho\bar{X})'}{\sum_{i=1}^{n}(X_i,Y_i)\,\Sigma^{-1}(X_i,Y_i)'}<1-e^{-2/n},\ \bar{X}<0,\ \bar{Y}-\rho\bar{X}\geqq 0 \\ \text{or } \dfrac{n(\bar{X}-\rho\bar{Y},0)\,\Sigma^{-1}(\bar{X}-\rho\bar{Y},0)'}{\sum_{i=1}^{n}(X_i,Y_i)\,\Sigma^{-1}(X_i,Y_i)'}<1-e^{-2/n};\ \bar{X}-\rho\bar{Y}\geqq 0,\ \bar{Y}<0 \\ \text{or } \bar{X}-\rho\bar{Y}<0,\ \bar{Y}-\rho\bar{X}<0 \end{cases}$$

where $\Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$.

As the estimation procedure of μ is similar to Case 1, we can also obtain the significance level α as follows;

$$(5)\qquad \alpha=1-\frac{1}{2\pi}\cos^{-1}\rho-\frac{1}{2}I_{1-e^{-2/n}}\left(\frac{1}{2},\frac{2n-1}{2}\right)-\frac{1}{2\pi}\cos^{-1}(-\rho)I_{1-e^{-2/n}}(1,n-1)$$

where
$$I_a(n,m)=\frac{1}{B(n,m)}\int_0^a x^{n-1}(1-x)^{m-1}\,dx$$

and
$$B(n,m)=\int_0^1 x^{n-1}(1-x)^{m-1}\,dx\,.$$

2.2 Two sided case

In this section we would discuss the two sided case.

Model H_0: $\mu_1=\mu_2=0$,

Model H_1': $(\mu_1\geqq 0$ and $\mu_2\geqq 0)$ or $(\mu_1\leqq 0$ and $\mu_2\leqq 0)$

where at least one of the inequalities are strict in both cases.

Case 1. When σ^2 and ρ are known parameters, we have the following relation.

$$AIC(H_0)-AIC(H_1')<0$$

$$(6)\qquad \Leftrightarrow \begin{cases} n\chi^2<4,\ \bar{X}\geqq 0,\ \bar{Y}\geqq 0 \\ \text{or } \dfrac{\sqrt{n}}{\sqrt{1-\rho^2}}(\bar{X}-\rho\bar{Y})>-2,\ \bar{X}<0,\ \bar{Y}>0,\ |\bar{X}|\geqq|\bar{Y}| \\ \text{or } \dfrac{\sqrt{n}}{\sqrt{1-\rho^2}}(\bar{Y}-\rho\bar{X})<2,\ \bar{X}<0,\ \bar{Y}>0,\ |\bar{X}|<|\bar{Y}| \\ \text{or } n\chi^2<4,\ \bar{X}\leqq 0,\ \bar{Y}\leqq 0 \end{cases}$$

$$\begin{cases} \text{or} \quad \dfrac{\sqrt{n}}{\sqrt{1-\rho^2}}(\bar{Y}-\rho\bar{X})>-2,\ \bar{X}>0,\ \bar{Y}<0,\ |\bar{X}|\leqq|\bar{Y}| \\ \text{or} \quad \dfrac{\sqrt{n}}{\sqrt{1-\rho^2}}(\bar{X}-\rho\bar{Y})<2,\ \bar{X}<0,\ \bar{Y}<0,\ |\bar{X}|>|\bar{Y}|\,. \end{cases}$$

Therefore our estimator $\hat{\mu}$ of μ is as follows;

$$(7) \qquad \hat{\mu}=\begin{cases} \text{if (2) is satisfied,} \\ \begin{pmatrix}0\\0\end{pmatrix}, \\ \text{if (2) is not satisfied,} \\ \begin{pmatrix}\bar{X}\\\bar{Y}\end{pmatrix} \text{ when } \bar{X} \text{ and } \bar{Y} \text{ are of the same sign,} \\ \begin{pmatrix}\bar{X}-\rho\bar{Y}\\0\end{pmatrix} \text{ when } \bar{X} \text{ and } \bar{Y} \text{ are of the different sign and } |\bar{X}|\geqq|\bar{Y}|, \\ \begin{pmatrix}0\\\bar{Y}-\rho\bar{X}\end{pmatrix} \text{ when } \bar{X} \text{ and } \bar{Y} \text{ are of the different sign and } |\bar{X}|<|\bar{Y}|. \end{cases}$$

The significance level α is given by

$$(8) \qquad \alpha=2\Big[\frac{e^{-2}(\pi-\cos^{-1}\rho)}{2\pi}+\frac{1}{\sqrt{2\pi}}\int_2^\infty \exp\Big(-\frac{x^2}{2}\Big)dx$$
$$-\frac{1}{2\pi\sqrt{1-\rho^2}}\int_2^\infty\int_2^\infty \exp\Big\{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\Big\}\,dxdy\Big]\,.$$

Values of α are also tabulated in Table 1.

Case 2. When σ^2 is an unknown parameter and ρ is a known parameter, we have the following relation.

$$AIC(H_0)-AIC(H_1')<0$$

$$(9) \qquad \Leftrightarrow \begin{cases} \dfrac{n(\bar{X},\bar{Y})\,\Sigma^{-1}(\bar{Y},\bar{Y})'}{\sum_{i=1}^n (X_i,Y_i)\,\Sigma^{-1}(X_i,Y_i)'}<1-e^{-2/n} \text{ and } \{(\bar{X}\geqq0,\bar{Y}\geqq0) \text{ or } (\bar{X}\leqq0,\bar{Y}\leqq0)\} \\ \text{or } \dfrac{n(0,\bar{Y}-\rho\bar{X})\,\Sigma^{-1}(0,\bar{Y}-\rho\bar{X})'}{\sum_{i=1}^n (X_i,Y_i)\,\Sigma^{-1}(X_i,Y_i)'}<1-e^{-2/n} \text{ and } \{(\bar{X}+\bar{Y}\geqq0,\bar{Y}\leqq0) \\ \qquad\qquad \text{or } (\bar{X}+\bar{Y}\leqq0,\bar{Y}\geqq0)\} \\ \text{or } \dfrac{n(\bar{Y}-\rho X,0)\,\Sigma^{-1}(0,\bar{Y}-\rho\bar{X})'}{\sum_{i=1}^n (X_i,Y_i)\,\Sigma^{-1}(X_tY_i)'}<1-e^{-2/n} \text{ and } \{(\bar{X}+\bar{Y}\leqq0,\bar{X}\geqq0) \\ \qquad\qquad \text{or } (\bar{X}+\bar{Y}\geqq0,\bar{X}\leqq0)\}. \end{cases}$$

As the estimation procedure of μ is similar to Case 1, we can also get the significance level α as follows;

$$(10) \qquad \alpha=1-\frac{1}{\pi}\Big\{\cos^{-1}\rho I_{1-e^{-3/n}}\Big(\frac{1}{2},\frac{2n-1}{2}\Big)+\cos^{-1}(-\rho)\,I_{1-e^{-2/n}}(1,n-1)\Big\}\,.$$

TABLE 1

ρ	α(one-sided)	α(two-sided)	ρ	α(one-sided)	α(two-sided)
−0.9	.0325	.0650	0.0	.0566	.1122
−0.8	.0366	.0732	0.1	.0588	.1158
−0.7	.0399	.0798	0.2	.0609	.1190
−0.6	.0427	.0854	0.3	.0631	.1222
−0.5	.0453	.0906	0.4	.0654	.1250
−0.4	.0477	.0954	0.5	.0679	.1276
−0.3	.0500	.0998	0.6	.0704	.1298
−0.2	.0522	.1042	0.7	.0733	.1318
−0.1	.0544	.1082	0.8	.0767	.1336
			0.9	.0807	.1346

3. Estimation of one component of the bivariate normal mean

In this section we would consider the estimation of μ_2. If we can known the value of μ_1, it is natural that we should use this knowledge to estimate μ_2.

3.1 One sided case

Our aim is to estimate μ_2 where two alternative models are given as follows;

Model H_0: $\mu_1=0$,

Model H_1: $\mu_1>0$.

Then the estimator $\hat{\mu}_2$ of μ_2 is as follows;

$$\hat{\mu}_2=\begin{cases}\bar{Y}-\rho\bar{X} & \text{if } H_0 \text{ is accepted,}\\ \bar{Y} & \text{if } H_1 \text{ is accepted.}\end{cases} \tag{11}$$

Case 1. When σ^2 and ρ are known parameters, whether Model H_0 is accepted or not is judged through AIC similary as § 2.

Then we have the following relation.

$$\begin{aligned} & AIC(H_0)-AIC(H_1)<0 \\ \Leftrightarrow\ & \frac{\sqrt{n}\,\bar{X}}{\sigma}<\sqrt{2} \end{aligned} \tag{12}$$

Therefore the significance level α of this test criterion is exactly

$$\begin{aligned} \alpha &= 1-Pr\left(\frac{\sqrt{n}\,X}{\sigma}<\sqrt{2}\right) \\ &= 0.07864\cdots . \end{aligned} \tag{13}$$

Case 2. When σ^2 is an unknown parameter and ρ is a known parameter, we have the following relation.

$$\begin{aligned} & AIC(H_0)-AIC(H_1)<0 \\ \Leftrightarrow\ & t_{2(n-1)}=\frac{\sqrt{n}\,\bar{X}}{\frac{1}{\sqrt{1-\rho^2}}\sqrt{\sum_{i=1}^{n}\{(X_i-\bar{X})^2-2\rho(X_i-\bar{X})(Y_i-\bar{Y})+(Y_i-\bar{Y})^2\}}} \\ & <\sqrt{2(n-1)}\sqrt{e^{1/n}-1}, \end{aligned} \tag{14}$$

where t_m is distributed according to the t-distribution with m degrees of freedom. The estimator $\hat{\mu}_2$ of μ_2 is the same as (11) and the significance level α is exactly

(15)
$$\alpha = 1 - Pr\{t_{2(n-1)} < \sqrt{2(n-1)}\sqrt{e^{1/n}-1}\}$$
$$= \frac{1}{2}\left\{1 - I_{1-e^{-1/n}}\left(\frac{1}{2}, n-1\right)\right\}.$$

Values of α are tabulated in Table 2 for various values of n. It should be noted that under H_0

$$\alpha = \lim_{n\to\infty} Pr\{AIC(H_0) - AIC(H_1) \geqq 0 \mid H_0\}$$
$$= 0.07864\cdots.$$

TABLE 2

n	α	$\sqrt{2(n-1)}\sqrt{e^{1/n}-1}$	n	α	$\sqrt{2(n-1)}\sqrt{e^{1/n}-1}$
5	.1100	1.3309	30	.0831	1.4021
10	.0929	1.3759	35	.0825	1.4039
15	.0878	1.3894	40	.0820	1.4052
20	.0854	1.3958	50	.0813	1.4070
25	.0840	1.3996	100	.0800	1.4106

3.2 Two sided case

Our aim is to estimate μ_2 where two alternative models are given as follows;

Model H_0: $\mu_1=0$,

Model H_1': $\mu_1 \neq 0$,

When σ^2 is an unknown parameter and ρ is a known parameter, we have the following relation.

(16)
$$AIC(H_0) - AIC(H_1') < 0$$
$$\Leftrightarrow |t_{2(n-1)}| < \sqrt{2(n-1)}\sqrt{e^{1/n}-1}.$$

Therefore the significance level α is exactly

(17)
$$\alpha = 1 - Pr\{|t_{2(n-1)}| < \sqrt{2(n-1)}\sqrt{e^{1/n}-1}\}$$
$$= 1 - I_{1-e^{-1/n}}\left(\frac{1}{2}, n-1\right).$$

4. Estimation of correlation coefficient

Let the random samples $\begin{pmatrix}X_1\\Y_1\end{pmatrix}, \ldots, \begin{pmatrix}X_n\\Y_n\end{pmatrix}$ be taken from a bivariate normal distribution $N\left[\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right]$, where μ_1, μ_2, σ_1 and σ_2 are unknown prameters. Our aim is to estimate the correlation coefficient ρ.

Model H_0: $\rho=0$,

Model H_1: $\rho \neq 0$.

Then the preliminary test estimator $\hat{\rho}$ of ρ is

$$(18) \qquad \hat{\rho} = \begin{cases} 0 \text{ if } H_0 \text{ is accepted,} \\ r \text{ if } H_1 \text{ is accepted,} \end{cases}$$

where

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}} .$$

By simple computations, we have

$$(19) \qquad AIC(H_0) - AIC(H_1) < 0$$

$$\Leftrightarrow |r| < \sqrt{1 - e^{-2/n}} .$$

When H_0 is true, the significance level α is exactly

$$(20) \qquad \alpha = 1 - Pr(|t_{n-2}| < \sqrt{n-2} \sqrt{e^{2/n} - 1})$$

$$= 1 - I_{1-e^{2/n}}\left(\frac{1}{2}, \frac{n-2}{2}\right).$$

Next when two models are given as follows;

Model H_0': $\rho = \rho_0$

Model H_1': $\rho \neq \rho_0$ where ρ_0 $(\neq 0)$ is the known constant,

the preliminary test estimator $\hat{\rho}$ of ρ is

$$(21) \qquad \hat{\rho} = \begin{cases} \rho_0 \text{ if } H_0' \text{ is accepted,} \\ r \text{ if } H_1' \text{ is accepted.} \end{cases}$$

By tedious computations we get the following relation.

$$(22) \qquad AIC(H_0') - AIC(H_1') < 0$$

$$\Leftrightarrow \rho_L < r < \rho_U$$

where

$$\rho_L = \frac{1}{e^{2/n} + (1 - e^{2/n})\rho_0^2} \{\rho_0 - (1 - \rho_0^2) e^{1/n} \sqrt{e^{2/n} - 1}\},$$

$$\rho_U = \frac{1}{e^{2/n} + (1 - e^{2/n})\rho_0^2} \{\rho_0 + (1 - \rho_0^2) e^{1/n} \sqrt{e^{2/n} - 1}\}.$$

When H_0 is true, the significance level α is

$$\alpha = 1 - Pr(\rho_L < r < \rho_U).$$

We can get values of α by referring to the table by David [4]. These are shown in Table 3, including the case of $\rho_0 = 0$. It should be noted that under H_0 in both cases

$$\begin{aligned}(23)\qquad \alpha &= \lim_{n\to\infty} Pr\{AIC(H_0)-AIC(H_1)\geqq 0\,|\,H_0\} \\ &= \lim_{n\to\infty} Pr\{AIC(H_0')-AIC(H_1')\geqq 0\,|\,H_0'\} \\ &= 0.1572\cdots.\end{aligned}$$

TABLE 3

ρ_0 \ n	4	6	8	10	15	20	25	50	100
0.0	.373	.277	.240	.220	.196	.186	.180	.168	.163
0.2	.372	.276	.239	.220	.196	.186	.180	.168	.163
0.4	.370	.275	.238	.219	.196	.185	.179	.168	.163
0.6	.366	.273	.237	.218	.195	.185	.179	.168	.163
0.8	.359	.268	.234	.216	.194	.184	.178	.167	.162
0.9	.354	.266	.232	.214	.191	.182	.177	.167	.162

The authors are deeply indebted to Professor A. Kudô of Kyushu University for his helpful advices and encouragements.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, 267–281.

[2] Asano, C. (1960). Estimation after preliminary tests of significance and their applications to biometrical researches, Bull. Math. Statist., 9, 1–14.

[3] Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance, Ann. Math. Statist., **15**, 190–204.

[4] David, F.N. (1938). Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples, Cambridge University Press, London.

[5] Hirano, K. (1978). On level of significance of the preliminary test in pooling means, Ann. Inst. Statis. Math., **30**, A, 1–8.

[6] Kitagawa, T. (1963). Estimation after preliminary tests of significance, Univ. Calif. Pub. Statist., **3**, 147–186.

[7] Sawa, T. and Hiromatsu, T. (1973). Minimax regret significance points for a preliminary test in regression analysis, Econometrica, **41**, 1093–1101.