

COMPARING PREDICTIVE VALUES OF STATISTICAL TEST VS. AIC AND BIC IN ONE SAMPLE PROBLEM

Tajiri, Ryo
Graduate School of Medicine, Kurume University

Iwamoto, Kasumi
Graduate School of Medicine, Kurume University

Fujikawa, Kei
Graduate School of Medicine, Kurume University

Yanagawa, Takashi
Biostatistics Center, Kurume University : Professor

<https://doi.org/10.5109/1909524>

出版情報 : Bulletin of informatics and cybernetics. 47, pp.59-67, 2015-12. Research Association
of Statistical Sciences

バージョン :

権利関係 :

COMPARING PREDICTIVE VALUES OF STATISTICAL TEST VS. AIC AND BIC IN ONE SAMPLE PROBLEM

By

Ryo TAJIRI*, Kasumi IWAMOTO†, Kei FUJIKAWA‡
and
Takashi YANAGAWA§

Abstract

Often, AIC and BIC are applied to statistical decision problems where methods of testing statistical hypothesis have been unrivaled. We define decision rules based on statistical test, on AIC and on BIC and compare their positive predictive values, and negative predictive values as well, in this paper in the framework of one-sample problem, assuming underlying normal distribution.

Key Words and Phrases: Akaike information criterion, Bayesian information criterion, positive predictive value, negative predictive value.

1. Introduction

The AIC (Akaike Information Criterion) was proposed by H. Akaike in 1973 as a tool for selecting an optimal model among candidates, implemented in many statistical soft wares and widely applied for model selection in scientific problems. The AIC is based on the Kullback-Leibler information. On the other hand, the BIC, the Bayesian information criterion, was introduced by Schwarz (1978) as a competitor to the AIC.

Often, AIC and BIC are applied to statistical decision problems where methods of testing statistical hypothesis have played a major role. For examples, Nagano et.al. (2012) conducted an exploratory clinical trial for comparing two types of nutrients. The sample size of each group in their trial was only ten and it was speculated that no significant result was produced with such small sample sizes unless increased the level of statistical tests from 5% to some value. However, it was not easy to decide what value of the level they should employ. They circumvented the problem by using the AIC, instead of statistical test, that takes the balance between the Type I error and Type II error probabilities. Also, Yanagawa et.al. (1997) advocated the use of AIC, instead of statistical hypothesis tests, in problems of determining whether chemical compound was toxic substance or not in a rodent bioassay where number of animals was limited. Emphasizing the importance of controlling the Type II error probability, rather than

* Graduate School of Medicine, Kurume University, 67 Asahi-Machi, Kurume-Shi 830-0011, Japan

† Graduate School of Medicine, Kurume University, 67 Asahi-Machi, Kurume-Shi 830-0011, Japan

‡ Graduate School of Medicine, Kurume University, 67 Asahi-Machi, Kurume-Shi 830-0011, Japan

§ Biostatistics Center, Kurume University, 67 Asahi-Machi, Kurume-Shi 830-0011, Japan

Type I error probability, they suggested the use of the AIC since it provided smaller Type II error than statistical tests.

As far as we know no formal comparison has been made between statistical tests and AIC. We also take into account the BIC in the comparison. We define in this paper the positive predictive value (PPV) and negative predictive value (NPV), and compare the PPV's, and also NPV's, of decision rules that are based on statistical test, the AIC and the BIC in the simplest framework, namely, in one sample statistical testing problems, assuming normality.

In Section 2, we formulate the framework of the present paper, by defining PPV and NPV; and also defining those decision rules based on statistical test, AIC and BIC. In Section 3, we give Type I error and Type II error probabilities of these decision rules. In Section 4, the PPV of these decision rules are compared. Finally in Section 5 NPV's of those decision rules are compared.

2. Formulation of the problem

Let X_1, X_2, \dots, X_n be independently and identically distributed random variables according to a normal distribution with mean μ and known variance σ^2 . We assume without loss of generality $\sigma^2 = 1$. We consider in this paper the decision rules for deciding whether $H_0 : \mu = 0$ or $H_1 : \mu \neq 0$. We introduce a prior probability $P(\mu = 0)$ for $\mu = 0$. Suppose that T^+ (T^-) expresses the decision for selecting H_1 (H_0) based on a decision rule. Then the positive predictive value (PPV) and negative predictive value (NPV) are defined as follows.

$$PPV = P(\mu \neq 0 | T^+), \quad NPV = P(\mu = 0 | T^-).$$

Now, we define the decision rules based on statistical test, AIC and BIC.

Decision rule based on statistical test

$$\begin{cases} \sqrt{n}|\bar{x}| \geq z_{\alpha/2} \Rightarrow \text{accept } H_1, \\ \sqrt{n}|\bar{x}| < z_{\alpha/2} \Rightarrow \text{accept } H_0. \end{cases}$$

where, $z_{\alpha/2}$ is the upper $\alpha/2 \times 100\%$ point of the standard normal distribution. Although in a routine testing statistical hypothesis, the decision of not accept H_1 is made if $\sqrt{n}|\bar{x}| < z_{\alpha/2}$, H_0 is accepted in this case in this decision.

Decision rule based on AIC

Denote by AIC_0 and AIC_1 the AIC computed under H_0 and H_1 , respectively. Then the decision rule based on AIC is given by

$$\begin{cases} AIC_0 \geq AIC_1 \Rightarrow \text{accept } H_1, \\ AIC_0 < AIC_1 \Rightarrow \text{accept } H_0, \end{cases}$$

It follows from the definition of AIC the decision rule is equivalently represented by

$$\begin{cases} \sqrt{n}|\bar{x}| \geq \sqrt{2} \Rightarrow \text{accept } H_1, \\ \sqrt{n}|\bar{x}| < \sqrt{2} \Rightarrow \text{accept } H_0. \end{cases}$$

Decision rule based on BIC

Similarly, the decision rule based on BIC is represented by

$$\begin{cases} \sqrt{n}|\bar{x}| \geq \sqrt{\log n} \Rightarrow \text{accept } H_1, \\ \sqrt{n}|\bar{x}| < \sqrt{\log n} \Rightarrow \text{accept } H_0. \end{cases}$$

3. The probabilities of the first and second kinds of errors.

All decision rules given above may be unified by

$$\begin{cases} \sqrt{n}|\bar{x}| \geq c \Rightarrow \text{accept } H_1, \\ \sqrt{n}|\bar{x}| < c \Rightarrow \text{accept } H_0. \end{cases}$$

We call this the unified decision rule.

The Type I error probability and Type II error probability at $\mu = \mu_1 (\neq 0)$ of the unified decision rule are given by

$$\alpha_c = P(\sqrt{n}|\bar{X}| \geq c | \mu = 0), \quad \beta_c = P(\sqrt{n}|\bar{X}| < c | \mu = \mu_1).$$

We immediately have

$$\alpha_c = 2(1 - \Phi(c)), \quad \beta_c = \Phi(c + \sqrt{n}\mu_1) - \Phi(-c + \sqrt{n}\mu_1), \quad (1)$$

where Φ is the cumulative distribution function of a standard normal distribution.

Substituting $c = z_{\alpha/2}$, $\sqrt{2}$, or $\sqrt{\log n}$, we have the following lemma.

Lemma 1. Type I error probability and II error probability at $\mu = \mu_1 (\neq 0)$ of the three decision rules are given as follows, where $A_n = \sqrt{n}\mu_1$.

(1) The decision rule that based on statistical test:

$$\alpha_T = \alpha, \quad \beta_T = \Phi(z_{\alpha/2} + A_n) - \Phi(-z_{\alpha/2} + A_n).$$

(2) The decision rule that based on AIC:

$$\alpha_A = 2\Phi(-\sqrt{2}), \quad \beta_A = \Phi(\sqrt{2} + A_n) - \Phi(-\sqrt{2} + A_n).$$

(3) The decision rule that based on BIC:

$$\alpha_B = 2\Phi(-\sqrt{\log n}), \quad \beta_B = \Phi(\sqrt{\log n} + A_n) - \Phi(-\sqrt{\log n} + A_n).$$

4. Comparison of positive predictive values.

Consider the positive predictive value defined in Section 2 at $\mu = \mu_1 (\neq 0)$. Applying the Bayes theorem, the positive predictive value may be represented by

$$\begin{aligned} PPV &= \frac{P(T^+ | \mu = \mu_1) P(\mu = \mu_1)}{P(T^+ | \mu = \mu_1) P(\mu = \mu_1) + P(T^+ | \mu = 0) P(\mu = 0)} \\ &= \frac{RP(T^+ | \mu = \mu_1)}{RP(T^+ | \mu = \mu_1) + P(T^+ | \mu = 0)}, \end{aligned}$$

where $R = P(\mu = \mu_1)/P(\mu = 0)$. Note that for the unified decision rule we have

$$P(T^+|\mu = \mu_1) = 1 - \beta_c, \quad P(T^+|\mu = 0) = \alpha_c.$$

Therefore, PPV of the unified decision rule may be represented by

$$PPV_c = \frac{R(1 - \beta_c)}{R(1 - \beta_c) + \alpha_c}.$$

From (1) we have

$$1 - \beta_c = 1 - (\Phi(c + A_n) - \Phi(-c + A_n)) = \Phi(-c - A_n) + \Phi(-c + A_n)$$

Thus

$$PPV_c = \frac{R\{\Phi(-c - A_n) + \Phi(-c + A_n)\}}{R\{\Phi(-c - A_n) + \Phi(-c + A_n)\} + 2\Phi(-c)},$$

where $A_n = \sqrt{n}\mu_1$.

Substituting $c = z_{\alpha/2}$, $\sqrt{2}$, or $\sqrt{\log n}$, we have the following lemma.

Lemma 2. The PPV of each decision rule is represented as follows.

(1) The decision rule that based on statistical test:

$$PPV_T = \frac{R\{\Phi(-z_{\alpha/2} - A_n) + \Phi(-z_{\alpha/2} + A_n)\}}{R\{\Phi(-z_{\alpha/2} - A_n) + \Phi(-z_{\alpha/2} + A_n)\} + 2\Phi(-z_{\alpha/2})}.$$

(2) The decision rule that based on AIC:

$$PPV_A = \frac{R\{\Phi(-\sqrt{2} - A_n) + \Phi(-\sqrt{2} + A_n)\}}{R\{\Phi(-\sqrt{2} - A_n) + \Phi(-\sqrt{2} + A_n)\} + 2\Phi(-\sqrt{2})}$$

(3) The decision rule that based on BIC:

$$PPV_B = \frac{R\{\Phi(-\sqrt{\log n} - A_n) + \Phi(-\sqrt{\log n} + A_n)\}}{R\{\Phi(-\sqrt{\log n} - A_n) + \Phi(-\sqrt{\log n} + A_n)\} + 2\Phi(-\sqrt{\log n})}.$$

Lemma 3. For any constant A ,

$$h(y) = \frac{\Phi(-y - A) + \Phi(-y + A)}{\Phi(-y)}$$

is a strongly increasing function of $y > 0$.

Proof. We have

$$\frac{dh(y)}{dy} = \frac{u(y)}{\Phi(-y)^2},$$

where

$$u(y) = -\phi(y - A)\Phi(-y) + \phi(y)\Phi(-y + A) - \phi(y + A)\Phi(-y) + \phi(y)\Phi(-y - A),$$

and ϕ is the density function of a standard normal distribution. Since

$$\begin{aligned}\Phi(-y + A) &= \int_{-\infty}^{-y+A} \phi(x) dx \\ &= \int_{-\infty}^{-y} \phi(x + A) dx = \int_{-\infty}^{-y} \exp\left(\frac{A^2 + 2Ax}{2}\right) \phi(x) dx,\end{aligned}$$

and similarly

$$\Phi(-y - A) = \int_{-\infty}^{-y} \exp\left(\frac{A^2 - 2Ax}{2}\right) \phi(x) dx,$$

$u(y)$ may be represented by

$$u(y) = \int_{-\infty}^{-y} \exp\left(\frac{A^2}{2}\right) \phi(x) \phi(y) \left(e^{Ax} + e^{-Ax} - (e^{Ay} + e^{-Ay}) \right) dx.$$

Put

$$v(x) = e^{Ax} + e^{-Ax},$$

then $v(x) > v(y)$ for any x such that $-\infty < x < -y$ when $y > 0$. Therefore, the integrand is positive for $-\infty < x < -y$, and thus $u(y) > 0$ for any $y > 0$, and then $dh(y)/dh > 0$. Thus $h(y)$ is a strongly increasing function of $y > 0$. This completes the proof of the lemma.

Lemma 4. For any constant $R > 0$ and $A_n = \sqrt{n}\mu_1$, it follows that

$$g(y) = \frac{R \left(\Phi(-y - A_n) + \Phi(-y + A_n) \right)}{R \left(\Phi(-y - A_n) + \Phi(-y + A_n) \right) + 2\Phi(-y)}$$

is a strongly increasing function of $y > 0$.

Proof. By using $h(y)$ defined in Lemma 3 we may represent $g(y)$ at $A = A_n$ as

$$g(y) = \frac{Rh(y)}{2 + Rh(y)} = 1 - \frac{2}{2 + Rh(y)}.$$

Since $h(y)$ is a strongly increasing function of $y > 0$ for any constant A_n from Lemma 3, it follows that $g(y)$ is a strongly increasing function of y .

Theorem 1. Consider decision rules based on statistical test, AIC and BIC for $H_0: \mu = 0$ vs. $H_1: \mu \neq 0$ and their PPV's at $\mu = \mu_1$. Then it follows the following relationships for any prior odds R and μ_1 .

- (1) $PPV_A > PPV_T \iff \alpha > 0.157$,
- (2) When $n > 1$, $PPV_B > PPV_T \iff n > \exp(Z_{\alpha/2}^2)$,
- (3) When $n > 1$, $PPV_B > PPV_A \iff n \geq 8$,

Proof. We first prove (3). Employing function $g(y)$ defined in Lemma 4, we may represent

$$PPV_A - PPV_B = g(\sqrt{2}) - g(\sqrt{\log n}).$$

Since $g(y)$ is a strongly increasing function of $y > 0$ from Lemma 4, it follows that

$$PPV_A > PPV_B \iff \sqrt{2} > \sqrt{\log n}.$$

Thus

$$\begin{aligned} n < 8 &\iff PPV_A < PPV_B, \\ n \geq 8 &\iff PPV_A > PPV_B. \end{aligned}$$

This completes the proof of (3). Since

$$\begin{aligned} PPV_A - PPV_T &= g(\sqrt{2}) - g(Z_{\alpha/2}), \\ PPV_T - PPV_B &= g(Z_{\alpha/2}) - g(\sqrt{\log n}), \end{aligned}$$

we may prove (1) and (2) similarly as the proof of (3).

5. Comparison of negative predictive values.

Applying the Bayes theorem, the negative predictive value defined in Section 2 at $\mu = \mu_1 (\neq 0)$ is represented by

$$\begin{aligned} NPV &= \frac{P(T^-|\mu=0)P(\mu=0)}{P(T^-|\mu=0)P(\mu=0) + P(T^-|\mu=\mu_1)P(\mu=\mu_1)} \\ &= \frac{P(T^-|\mu=0)}{P(T^-|\mu=0) + RP(T^-|\mu=\mu_1)}, \end{aligned}$$

where $R = P(\mu = \mu_1)/P(\mu = 0)$. For the unified decision rule defined in Section 2, since

$$P(T^-|\mu=0) = 1 - \alpha_c, \quad P(T^-|\mu=\mu_1) = \beta_c.$$

we have

$$NPV_c = \frac{1 - \alpha_c}{(1 - \alpha_c) + R\beta_c}.$$

From (1) we have

$$1 - \alpha_c = 1 - 2(1 - \Phi(c)) = \Phi(c) - \Phi(-c).$$

Thus

$$NPV_c = \frac{\Phi(c) - \Phi(-c)}{(\Phi(c) - \Phi(-c)) + R\{\Phi(c + A_n) - \Phi(-c + A_n)\}},$$

where $A_n = \sqrt{n}\mu_1$.

Substituting $c = z_{\alpha/2}, \sqrt{2}, \sqrt{\log n}$, we have the following lemma.

Lemma 5. The NPV of each decision rule at $\mu = \mu_1$ is represented as follows.

(1) The decision rule that based on statistical test:

$$NPV_T = \frac{\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})}{\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) + R\{\Phi(z_{\alpha/2} + A_n) - \Phi(-z_{\alpha/2} + A_n)\}}.$$

(2) The decision rule that based on AIC:

$$NPV_A = \frac{\Phi(\sqrt{2}) - \Phi(-\sqrt{2})}{\Phi(\sqrt{2}) - \Phi(-\sqrt{2}) + R\{\Phi(-\sqrt{2} + A_n) - \Phi(-\sqrt{2} + A_n)\}}.$$

(3) The decision rule that based on BIC:

$$NPV_B = \left(\Phi(\sqrt{\log n}) - \Phi(-\sqrt{\log n}) \right) \left(\Phi(-\sqrt{\log n}) - \Phi(-\sqrt{\log n}) \right) \\ + R \{ \Phi(-\sqrt{\log n} + A_n) - \Phi(-\sqrt{\log n} + A_n) \}^{-1}.$$

We need following lemmas to compare these NPV's.

Lemma 6. For any constant A put

$$s(y) = \int_{-y}^y \left(e^{-Ay} + e^{Ay} - 2e^{-At} \right) \phi(t) dt,$$

then $s(y) > 0$ for any $y > 0$.

Proof. Since $s(y)$ is rewritten by

$$s(y) = 2 \int_0^y \left(e^{-Ay} + e^{Ay} - (e^{-At} + e^{At}) \right) \phi(t) dt,$$

and

$$e^{-Ay} + e^{Ay} - (e^{-At} + e^{At}) > 0$$

for any t such that $0 < t < y$, we have $s(y) > 0$ for any constant A .

Lemma 7. For any constant A put

$$k(y) = \left(\phi(y + A) + \phi(y - A) \right) \left(\Phi(y) - \Phi(-y) \right) \\ - 2 \left(\Phi(y + A) - \Phi(-y + A) \right) \phi(y),$$

then $k(y) > 0$ for any $y > 0$.

Proof. We have

$$\phi(y + A) = \exp(-Ay - A^2/2) \phi(y), \\ \Phi(y + A) - \Phi(-y + A) = \exp(-A^2/2) \int_{-y}^y \phi(x) e^{-Ax} dx.$$

Substituting these formula to $k(y)$, we may represent $k(y)$, by using $s(y)$ defined in Lemma 6, as $k(y) = \exp(A^2/2) \phi(y) s(y)$. Thus from Lemma 6 we have $k(y) > 0$ for any $y > 0$.

Lemma 8 For $A_n = \sqrt{n} \mu_1$ put

$$\ell(y) = \frac{\Phi(y + A_n) - \Phi(-y + A_n)}{\Phi(y) - \Phi(-y)}.$$

Then $\ell(y)$ is a strictly increasing function of $y > 0$.

Proof. Since

$$\begin{aligned} \frac{d\ell(y)}{dy} &= \left(\{\phi(y + A_n) + \phi(y - A_n)\} \{\Phi(y) - \Phi(-y)\} \right. \\ &\quad \left. - 2\{\Phi(y + A_n) - \Phi(-y + A_n)\} \phi(y) \right) \left(\Phi(y) - \Phi(-y) \right)^{-2} \\ &= \frac{k(y)}{\left(\Phi(y) - \Phi(-y) \right)^2} \end{aligned}$$

at $A = A_n$, we have $d\ell(y)/dy > 0$ from Lemma 7. Thus $\ell(y)$ is a strongly increasing function of $y > 0$.

Theorem 2. Consider decision rules based on statistical test, AIC and BIC for $H_0: \mu = 0$ vs. $H_1: \mu \neq 0$ and their NPV's at $\mu = \mu_1$. Then we have the following relationships for any R and μ_1 .

- (1) $NPV_A > NPV_T \iff \alpha < 0.157$,
- (2) When $n > 1$, $NPV_B > NPV_T \iff n < \exp(Z_{\alpha/2}^2)$,
- (3) When $n > 1$, $NPV_B > NPV_A \iff n < 8$,

Proof. Put

$$f(x) = \frac{1}{1 + R\ell(x)},$$

by using function $\ell(x)$ defined in Lemma 7. Then we may represent NPV's by

$$NPV_T = f(Z_{\alpha/2}), \quad NPV_A = f(\sqrt{n}), \quad NPV_B = f(\sqrt{\log n})$$

Since $f(x)$ is a strictly decreasing function of $x > 0$ from Lemma 8, we have

$$\begin{aligned} NPV_A > NPV_T &\iff f(\sqrt{2}) > f(Z_{\alpha/2}) \iff \sqrt{2} < Z_{\alpha/2} \iff \alpha < 0.157, \\ NPV_B > NPV_T &\iff f(\sqrt{\log n}) > f(Z_{\alpha/2}) \iff n < \exp(Z_{\alpha/2}^2), \\ NPV_B > NPV_A &\iff f(\sqrt{\log n}) > f(\sqrt{2}) \iff n < 8. \end{aligned}$$

6. Discussion

The positive predictive value (PPV) is a key index for evaluating treatment effects, such as efficacies of new drugs in a clinical trials. We proved in Theorem 1 that $PPV_A > PPV_T$ if and only if $\alpha > 0.157$. Statistical tests are usually designed to satisfy 5% level and 80% power in confirmative clinical trias. If this is the case, Theorem 1 shows that statistical tests are superior to the AIC. However, if it is not the case, for example, when sample sizes are small like a pilot study, one must increases the level of significance since otherwise probabilities of Type II error inflate; often the level of significance is raised up to 20%. If this is the case the theorem shows that the AIC is superior to the statistical test in the sense of the positive predictive value. Theorem 1 also shows that the decision based on AIC is superior to the decision based on BIC if and only if $1 < n < 8$.

Suppose that an animal experiment is conducted to determine whether certain chemical compound is carcinogenesis or not; namely, to determine whether the chemical

compound is not carcinogen (H_0) or the chemical compound is carcinogen (H_1). Type II errors should be more strictly controlled than Type I error in the determination, since number of cancer patients increase among the users of compound if it is falsely determined no carcinogen and released in the country. Thus NPV is more important than PPV for general public in such bioassays. Theorem 2 shows that the decision based on AIC is superior to the decision based on statistical test unless $\alpha > 0.157$ and that it is also superior to the BIC if $n \geq 8$.

Acknowledgement

The authors are very grateful to the reviewer and the editor of the BIC for their careful reading and comments.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the 2nd International Symposium on information Theory, Petrov, B. N., and Caski, F. (eds.), Akadimiai Kiado, Budapest: 267-281.*
- Nagano, T., Fujita, H., Tanaka, T. et al. (2012). randomized controlled trial comparing antioxidant-enriched enteral nutrition with immune-enhancing enteral nutrition after esophagectomy for cancer: a pilot study *Surgery Today, DOI 10.1007/s00595-012-0424-1.*
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6, 461-464.*
- Yanagawa, T., Kikuchi, Y. and Brown, K. G. (1997). No-observed-adverse-effect levels in severity data. *Journal of American Statistical Association, Vol. 92, 449-454.*

Received October 20, 2015

Revised November 23, 2015