Journal of the Korean
Statistical Society
Vol. 21, No. 2, 1992

On the Model Selection Criteria in Normal Distributions†

Han-Yeong Chung[1] and Kee-Won Lee[2]

ABSTRACT

A model selection approach is used to find out whether the mean and the variance of a unique sample are different from the pre-specified values. Normal distribution is selected as an approximating model. Kullback-Leibler discrepancy comes out as a natural measure of discrepancy between the operating model and the approximating model. Several estimates of selection criterion are computed including AIC, TIC, and a couple of bootstrap versions. Nonparametric, modified nonparametric and parametric bootstrap estimator of the selection criterion are considered according to the way of resampling. It is shown that a closed form expression is available for the parametric bootstrap estimated criterion. A Monte-Carlo study is provided to give a formal comparison when the operating family itself is normally distributed.

1. INTRODUCTION

Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with distribution function F such that the mean of F, μ_F and the variance of F, σ_F^2, , and

[1] Department of Statistics, Hallym University, Chunchon, 200-702, Korea.
[2] Department of Statistics, Hallym University, Chunchon, 200-702, Korea.
† This research was supported by a Hallym University grant.

the fourth moment of F exist. For this sample from F, we want to know whether the mean and the variance remain the same as some precribed values.

In this paper a model selection approach is used to find out which submodel gives the best fit to the observed values. The distribution F described above is therefore an operating model supposed to generate our observations. Normal distribution is selected as a popular choice of an approximating model. Then the problem is to choose the best fitting submodel based on some reasonable selection criterion among the following four possible candidates; $N(\mu_0, \sigma_0^2)$, $N(\mu, \sigma_0^2)$, $N(\mu_0, \sigma^2)$, and $N(\mu, \sigma^2)$, where μ_0, σ_0^2 are some prescribed values and μ, σ^2 are unknown.

In this setting, Kullback-Leibler discrepancy comes out as a natural measure of discrepancy between the operating model and the approximating model. Let $g(\cdot, \theta)$ be the probability density function of the approximating model indexed by the parameter θ. Our selection criterion based on the expected overall discrepancy is given by

$$E\Big\{ - E_F \log g(Z, \hat{\theta})\Big\}, \tag{1.1}$$

where Z is a random variable distributed according to F but independent of the random sample already observed, and $\hat{\theta}$ is the minimum discrepancy estimator minimizing the empirical discrepancy, which in this case is given by $-\sum_{i=1}^{n} \log g(X_i, \theta)/n$. Thus $\hat{\theta}$ is algebraically the same as the maximum likelihood estimator under the approximating model. But as indicated in Linhart and Zucchini(1986), it has different asymptotic properties due to different probability structure.

Except for some trivial cases, it is not easy to derive (1.1) explicitly. By a clear recognition that unreserved maximization of the likelihood results in an unsatisfatory choice between models that differ appreciably in their dimensionality, Akaike(1973) introduced an information theoretic model selection criterion, AIC, which is given by

$$-2\sum_{i=1}^{n} \log g(X_i, \hat{\theta}) + 2p, \tag{1.2}$$

where p denotes the number of parameters estimated. In fact, AIC aims at the estimation of the quantity $2n \times$ (1.1). So, in our study, $AIC/2n$ will be tabulated for the purpose of comparison with other estimators of the selection criterion.

Another asymptotic criterion, which we call TIC, to refine AIC is derived and fully explained in Linhart and Zucchini(1986). When the approximating model $g(\cdot, \theta)$ is quite different from the operating model F, the minimum discrepancy estimator $\hat{\theta}$ has a different asymptotic distribution from the usual maximum likelihood estimator under the approximating model. This fact plays an important role in the derivation of TIC, which is given by

$$-\sum_{i=1}^{n} \log g(X_i, \hat{\theta})/n + \mathrm{tr}(\Omega_n^{-1}\Sigma_n)/n, \tag{1.3}$$

where $\Omega_n = -\sum_{i=1}^{n} \nabla^2 \log g(X_i, \theta)/n \mid_{\theta=\hat{\theta}}$,

and $\Sigma_n = \sum_{i=1}^{n} \{\nabla \log g(X_i, \theta)\}\{\nabla \log g(X_i, \theta)\}^t/n \mid_{\theta=\hat{\theta}}$.

The bootstrap, which is one of a very well-known resampling methods initiated by Efron(1979), can also be applied to our problem to give several bootstrap estimators of (1.1). A key idea in using the bootstrap method is to use $\hat{\theta}^*$ and $\hat{F}_n$ in place of $\hat{\theta}$ and F, the distribution of future observation Z, in the expression (1.1), where $\hat{\theta}^*$ is a bootstrap version of $\hat{\theta}$ computed from the bootstrap sample and F_n is the empirical cumulative distribution function based on the original sample.

Efron(1983) suggested a number of nonparametric methods to estimate the prediction error in case of logistic regression model including cross-validation and several different versions of bootstrap. Then the idea was further explored in the context of generalized linear models in Efron(1986). Efron(1983) proposed double bootstrap and a couple of bias adjusted bootstrap error rate estimates as improved ones over the naive bootstrap estimates. His concern was rather restricted to the so called nonparametric way of bootstrap and its improvement.

In this paper sophisticated ways of boostrap are not considered just for the purpose of an easy implementation. Three versions of naive bootstrap estimators are considered according to the method of resampling; the nonparametric bootstrap, the modified nonparametric bootstrap, and the parametric bootstrap.

In section 2, each estimator of selection criterion is derived. AIC will be modified in order to be comparable with other criteria. It will be argued that the parametric bootstrap estimated criterion can be derived explicitly. Whenever a closed form expression is not available, a stochastic approximation to that estimator of selection criterion is introduced. The algorithm will be explained in detail.

In section 3, a stream of numerical work is given. First, an illustrating example is provided to show its applicability. Next, bias and the mean squared error for the submodel (μ, σ_0^2) are derived explicitly. Then, it is shown that the expected overall discrepancy can be computed explicitly when F is itself normally distributed. Furthermore expected values and the mean squared errors of AIC and parametric bootstrap estimator of the selection criterion are derived explicitly in this particular case. Finally a Monte-Carlo study is provided to make sure the derivation and to compare each criterion in terms of the mean squared error.

2. THE ESTIMATED CRITERIA

In this section each estimator of the selection criterion will be derived explicitly for each submodel. For the bootstrap estimator of the selection criterion, it will be shown that a closed form expression is available for the parametric bootstrap estimated criterion. Simple expansion argument shows that all these efforts are aiming at the estimation of the first order correction term. Stochastic approximation algorithms are introduced whenever a closed form expression is not available.

2.1. AIC

AIC is given by the formula (1.2). The following summary of $AIC/2n$ can be easily checked.

θ	p	$AIC/2n$
(μ_0, σ_0^2)	0	$1/2(\log 2\pi + \log \sigma_0^2 + \tilde{\sigma}^2/\sigma_0^2)$
(μ, σ_0^2)	1	$1/2(\log 2\pi + \log \sigma_0^2 + \hat{\sigma}^2/\sigma_0^2 + 2/n)$
(μ_0, σ^2)	1	$1/2(\log 2\pi + \log \tilde{\sigma}^2 + 1 + 2/n)$
(μ, σ^2)	2	$1/2(\log 2\pi + \log \hat{\sigma}^2 + 1 + 4/n)$

In this table, $\hat{\mu}$ and $\hat{\sigma}^2$ are the usual maximum likelihood estimator of μ and σ^2 based on the random sample of size n from normal distribution with mean μ and variance σ^2. Also, $\tilde{\sigma}^2$ denotes the maximum likelihood estimator of σ^2 when $\mu = \mu_0$ is known. That is, $\hat{\mu} = \sum_{i=1}^{n} X_i/n$, $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \hat{\mu})^2/n$, and $\tilde{\sigma}^2 = \sum_{i=1}^{n}(X_i - \mu_0)^2/n$.

2.2. TIC

In order to derive the asymptotic criterion of the form (1.3), we just have to put the trace term in place of p/n in the above table. To get the trace term we need to differentiate the log-likelihood with respect to the corresponding parameters.

Let $\hat{\mu}_4 = \sum_{i=1}^{n}(X_i - \hat{\mu})^4/n$, $\tilde{\mu}_4 = \sum_{i=1}^{n}(X_i - \mu_0)^4/n$, then the following can be easily checked. Note that the trace term is approximately the number of parameters to be estimated when the approximating model is really close to the operating model.

θ	$\mathrm{tr}(\Omega_n^{-1}\Sigma_n)$	TIC
(μ_0, σ_0^2)	0	$1/2(\log 2\pi + \log \sigma_0^2 + \tilde{\sigma}^2/\sigma_0^2)$
(μ, σ_0^2)	$\hat{\sigma}^2/\sigma_0^2$	$1/2\{\log 2\pi + \log \sigma_0^2 + \hat{\sigma}^2/\sigma_0^2(1 + 2/n)\}$
(μ_0, σ^2)	$(\tilde{\mu}_4 - \tilde{\sigma}^4)/2\tilde{\sigma}^4$	$1/2\{\log 2\pi + \log \tilde{\sigma}^2 + 1 + (\tilde{\mu}_4 - \tilde{\sigma}^4)/n\tilde{\sigma}^4\}$
(μ, σ^2)	$(\hat{\mu}_4 + \hat{\sigma}^4)/2\hat{\sigma}^4$	$1/2\{\log 2\pi + \log \hat{\sigma}^2 + 1 + (\hat{\mu}_4 + \hat{\sigma}^4)/n\hat{\sigma}^4\}$

2.3. Nonparametric Bootstrap

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, the distribution of future observation Z, is independent of the original sample observed in the expression (1.1). This implies that Z is independent of $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$. This idea will be clarified when we construct the bootstrap estimator of (1.1) for each submodel.

For (μ_0, σ_0^2), it can be easily checked that the bootstrap estimator of the criterion remains the same as AIC or TIC. Furthermore, in this case the expression (1.1) becomes

$$E_F - \log g(Z) = 1/2\{\log 2\pi + \log \sigma_0^2 + E_F(Z - \mu_0)^2/\sigma_0^2\}. \quad (2.3.1)$$

Therefore, in this case AIC, TIC, and the bootstrap estimator of the criterion are all unbiased.

Remark. $E_F(Z - \mu_0)^2$ is not in general equal to σ_F^2.

For (μ, σ_0^2), (1.1) is given explicitly by

$$1/2\{\log 2\pi + \log \sigma_0^2 + (1 + 1/n)\sigma_F^2/\sigma_0^2\}. \quad (2.3.2)$$

Therefore, in this case bias and the mean squared error of each estimated criterion can be derived explicitly. Easily the nonparametric bootstrap estimator of the criterion is derived as

$$1/2\{\log 2\pi + \log \sigma_0^2 + (1 + 1/n)\hat{\sigma}^2/\sigma_0^2\}. \quad (2.3.3)$$

For (μ_0, σ^2), by noting that F is independent of the distribution of the original sample, it can be easily checked that

$$E_F - \log g(Z, \hat{\theta}) = 1/2\{\log 2\pi + \log \tilde{\sigma}^2 + E_F(Z - \mu_0)^2/\tilde{\sigma}^2\}. \quad (2.3.4)$$

Now, plug in $\tilde{\sigma}^{*2}$ in place of $\tilde{\sigma}^2$ in (2.3.4), which is computed from the bootstrap sample. The bootstrap sample $(X_1^*, \cdots, X_n^*)$ is obtained from the original sample $(X_1, \cdots, X_n)$ in a nonparametric way, that is, by sampling with replacement. Formally $\tilde{\sigma}^{*2} = \sum_{i=1}^{n}(X_i^* - \mu_0)^2/n$. Also plug in F_n, the empirical distribution function based on the original sample, in place of F. Then,

$$E_{F_n} - \log g(Z, \hat{\theta}^*) = 1/2\left\{\log 2\pi + \log \tilde{\sigma}^{*2} + \sum_{i=1}^{n}(X_i - \mu_0)^2/n\tilde{\sigma}^{*2}\right\}. \quad (2.3.5)$$

Taking expectation with respect to $\hat{\theta}^*$, denoted by $E^*(\cdot)$, we obtain a bootstrap estimator of the criterion given by

$$E^* 1/2(\log 2\pi + \log \tilde{\sigma}^{*2} + \tilde{\sigma}^2/\tilde{\sigma}^{*2}). \quad (2.3.6)$$

As we can see, a closed form expression is not easy to derive. A nice feature of bootstrap method is that the basic resampling procedure can be repeated, and we may easily turn to the stochastic approximation to the expression (2.3.6). Repeating the resampling procedure for a large number of times, say B times, we obtain B of (2.3.5)'s. By the weak law of large numbers, the average value of these B bootstrap estimated criteria approximates (2.3.6). Formally the stochastic approximation to (2.3.6) can be written as,

$$1/B\sum_{j=1}^{B} 1/2\{\log 2\pi + \log \tilde{\sigma}_j^{*2} + \tilde{\sigma}^2/\tilde{\sigma}_j^{*2}\}. \quad (2.3.7)$$

Note that we may write

$$\log \tilde{\sigma}^{*2} = -\log\{1 + (\tilde{\sigma}^2/\tilde{\sigma}^{*2} - 1)\} + \log \tilde{\sigma}^2. \quad (2.3.8)$$

Recall that $\tilde{\sigma}^{*2}/\tilde{\sigma}^2 - 1$ is $O_p^*(n^{-1/2})$, and that the function $h(x) = 1/x$ is continuous at $x = 1$, it can be checked that (2.3.6) is equivalent to

$$1/2(\log 2\pi + \log \tilde{\sigma}^2 + 1) + O_p(1/n), \quad (2.3.9)$$

by expanding $\log(1 + x)$ near 0.

Remark. The distribution of (2.3.5) can be approximated by the empirical cumulative distribution function based on B bootstrap estimated criteria by Glivenko-Cantelli theorem.

Finally for (μ, σ^2), we have

$$E_F - \log g(Z, \hat{\theta}) = 1/2\{\log 2\pi + \log \hat{\sigma}^2 + E_F(Z - \hat{\mu})^2/\hat{\sigma}^2\}. \quad (2.3.10)$$

Plug in $\hat{\theta}^* = (\hat{\mu}^*, \hat{\sigma}^{*2})$ in place of $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$. Formally the bootstrap versions of $(\hat{\mu}, \hat{\sigma}^2)$ are given by $\hat{\mu}^* = \sum_{i=1}^{n} X_i^*/n$, and $\hat{\sigma}^{*2} = \sum_{i=1}^{n}(X_i^* - \hat{\mu}^*)^2/n$. Also plug in F_n in place of F. Then

$$E_{F_n} - \log g(Z, \hat{\theta}^*) = 1/2\left\{\log 2\pi + \log \hat{\sigma}^{*2} + \sum_{i=1}^{n}(X_i - \hat{\mu}^*)^2/n\hat{\sigma}^{*2}\right\}. \quad (2.3.11)$$

Again, take expectation with respect to $\hat{\theta}^*$, we obtain the bootstrap estimated criterion given by

$$E^* 1/2\left\{\log 2\pi + \log \hat{\sigma}^{*2} + \sum_{i=1}^{n}(X_i - \hat{\mu}^*)^2/n\hat{\sigma}^{*2}\right\}. \quad (2.3.12)$$

A stochastic approximation to (2.3.12) is available by repeating the basic resampling procedure, which can be written as

$$1/B\sum_{j=1}^{B} 1/2\left\{\log 2\pi + \log \hat{\sigma}_j^{*2} + \sum_{i=1}^{n}(X_i - \hat{\mu}_j^*)^2/n\hat{\sigma}_j^{*2}\right\}. \quad (2.3.13)$$

Note that

$$\sum_{i=1}^{n}(X_i - \hat{\mu}^*)^2/n\hat{\sigma}^{*2} = \hat{\sigma}^2/\hat{\sigma}^{*2} + (\hat{\mu}^* - \hat{\mu})^2/\hat{\sigma}^{*2}, \quad (2.3.14)$$

and that

$$(\hat{\mu}^* - \hat{\mu})^2/\hat{\sigma}^{*2} = O_p^*(1/n). \quad (2.3.15)$$

Therefore another expansion argument can be used to show that (2.3.12) is equivalent to

$$1/2(\log 2\pi + \log \hat{\sigma}^2 + 1) + O_p(1/n). \quad (2.3.16)$$

So, up to the leading term, nonparametric bootstrap is equivalent to AIC.

2.4. Modified Nonparametric Bootstrap

A key idea in using the resampling method in the estimation of (1.1) is that the resampled version $\hat{\theta}^*$ will be close to $\hat{\theta}$. A slight modification of the bootstrap algorithm in section 2.3 aims at making the bootstrap sample as similar to the approximating model as possible, since θ is used to describe g not F. For (μ_0, σ_0^2), all the parameters are fully specified. So start with (μ, σ_0^2).

For (μ, σ_0^2), we have a restriction that the variance of the approximating model is σ_0^2, while the variance of X_i^* is $\hat{\sigma}^2$. Let the modified bootstrap sample be denoted by $(X_1^{**}, \cdots, X_n^{**})$. These are obtained from the bootstrap sample as follows;

$$X_i^{**} = (\sigma_0/\hat{\sigma})X_i^* + (1 - \sigma_0/\hat{\sigma})\hat{\mu}. \tag{2.4.1}$$

With this modification, the variance of modified bootstrap sample becomes σ_0^2, while the expected value of it stays the same. Then compute the modified bootstrap version of the estimated parameter and the modified bootstrap estimated criterion as usual. The following modified bootstrap estimated criterion will be obtained;

$$1/2(\log 2\pi + \log \sigma_0^2 + \hat{\sigma}^2/\sigma_0^2 + 1/n). \tag{2.4.2}$$

It will be seen in section 3 that this kind of modification in general reduces the mean squared error of the estimated criterion.

For (μ_0, σ^2), the expectation is specified as μ_0. Hence modify the bootstrap sample in such a way that its mean becomes μ_0. Modify the bootstrap sample as follows;

$$X_i^{**} = X_i^* - (\hat{\mu} - \mu_0). \tag{2.4.3}$$

Then the expected value of the modified bootstrap sample becomes μ_0 as specified. With this modified bootstrap sample the derivation of the estimated criterion similar to (2.3.6) and its stochastic approximation, (2.3.7) are straightforward. Again, (2.3.6) under modification leads to (2.3.9) asymptotically. For (μ, σ^2), no specific restrictions are made regarding the parameters of approximating model. So, no modification is needed.

Remark. What if we modify the original sample rather than the bootstrap sample initially obtained? It does not seem to give such a good result as can be easily checked. The reason seems that g is introduced as an approximating model, and it may or may not cover the true random mechanism generating the original sample. Even when we give some restrictions to a submodel, it may not hold true of the operating model. For example, in case of (μ, σ_0^2), there is a restriction that the variance of the operating model should be σ_0^2, while the operating model does not specify the size of variance. Therefore the situation is different from that of bootstrapping linear regression model with fixed covariate case where we draw the bootstrap sample from the recentered residuals to mimic the behaviour of the error distribution.

2.5. Parametric Bootstrap

What if we resample from the fitted approximation model to compute the bootstrap version of $\hat{\theta}$? This idea will be explored in the present section. Again the motivation of this type of resampling is that θ is a parameter indexing the approximating model. Therefore we may obtain the bootstrap sample from the fitted approximating model $g(\cdot, \hat{\theta})$ instead of just drawing with replacement from the original sample.

The following gives an algorithm of the procedure;

Step 1. Draw a bootstrap sample of size n from the fitted approximating model, that is, draw $X_1^*, \cdots, X_n^*$ from $N(\hat{\mu}, \sigma_0^2)$ for (μ, σ_0^2), from $N(\mu_0, \tilde{\sigma}^2)$ for (μ_0, σ^2), and from $N(\hat{\mu}, \hat{\sigma}^2)$ for (μ, σ^2).

Step 2. From the bootstrap sample obtained in the previous step, compute the bootstrap versions of $\hat{\theta}$, that is, compute $\hat{\mu}^* = \sum_{i=1}^{n} X_i^*/n$ for (μ, σ_0^2), $\tilde{\sigma}^{*2} = \sum_{i=1}^{n}(X_i^* - \mu_0)^2/n$ for (μ_0, σ^2), and $\hat{\mu}^* = \sum_{i=1}^{n} X_i^*/n$, $\hat{\sigma}^{*2} = \sum_{i=1}^{n}(X_i^* - \hat{\mu}^*)^2/n$ for (μ, σ^2).

Step 3. Compute the bootstrap version for the selection criterion similar to (2.3.3) for (μ, σ_0^2), similar to (2.3.6) for (μ_0, σ^2), and similar to (2.3.10) for (μ, σ^2).

One important feature of parametric bootstrap is that its analytic properties can be easily derived. Estimated criterion itself can be computed explicitly even for (μ_0, σ^2), and for (μ, σ^2).

For (μ_0, σ^2), bootstrap estimated criterion is expressed as (2.3.6). To compute $E^* \log \tilde{\sigma}^{*2}$, note that $n\tilde{\sigma}^{*2}/\tilde{\sigma}^2$ has a chi-square distribution with n degrees of freedom. For a chi-square distribution with ν degrees of freedom, say V, the cumulant generating function of $\log V$ is given by

$$K(t) = t \log 2 + \log \Gamma(\nu/2 + t) - \log \Gamma(\nu/2). \tag{2.5.1}$$

Therefore the first derivative and the second derivative of $K(t)$ are given by, .

$$K'(t) = \log 2 + \Psi(\nu/2 + t), \tag{2.5.2}$$

$$K''(t) = \Psi'(\nu/2 + t), \tag{2.5.3}$$

where $\Psi(\cdot)$ and $\Psi'(\cdot)$ are *digamm* function and *trigamma* function respectively. Therefore the expectation and variance of $\log V$ are given by

$$E(\log V) = \log 2 + \Psi(\nu/2), \tag{2.5.4}$$

$$Var(\log V) = \Psi'(\nu/2). \tag{2.5.5}$$

By (2.5.4), it can be easily checked that

$$E^* \log \tilde{\sigma}^{*2} = \log \tilde{\sigma}^2 + \Psi(n/2) - \log(n/2). \tag{2.5.6}$$

Furthermore, $E(\tilde{\sigma}^2/n\tilde{\sigma}^{*2}) = 1/(n-2)$. Therefore, under parametric bootstrap, (2.3.6) can be calculated explicitly as

$$1/2\{\log 2\pi + \log \tilde{\sigma}^2 + \Psi(n/2) - \log(n/2) + n/(n-2)\}. \tag{2.5.7}$$

For (μ, σ^2), by a similar argument, (2.3.10) under parametric bootstrap can be calculated explicitly as

$$1/2[\log 2\pi + \log \hat{\sigma}^2 + \Psi\{(n-1)/2\} - \log(n/2) + (n+1)/(n-3)]. \tag{2.5.8}$$

Remark. No further modification is needed as in case of nonparametric bootstrap, since all the parameters are fully specified in the resampling procedure.

3. A NUMERICAL STUDY

3.1 An Illustrating Example

Sakamoto, Ishiguro, and Kitagawa(1986) pp. 143 gives an example regarding the diameters of ball bearings produced from a certain machine. It has been known that the diameters of the products are distributed as a normal distribution with mean 1 cm and the standard deviation 0.01 cm when the machine is operating normally. One day 20 ball bearings were randomly chosen and their diameters were measured. The following summary statistics were obtained.

$$\hat{\mu} = 0.99985, \hat{\sigma}^2 = 0.000168. \tag{3.1.1}$$

In this example, we don't really have to specify the operating model in detail. So, drop the assumption and proceed as in section 2. Usually it is easy to correct the mean of the diameters but if it turns out that the standard deviation becomes larger than specified, it is much more troublesome. The following is a summary of the estimators.

	$AIC/2n$	TIC	NB	MNB	PB
$(1, 0.01^2)$	−2.8475	−2.8475	−2.8475	−2.8475	−2.8475
$(\mu, 0.01^2)$	−2.7976	−2.7637	−2.8057	−2.8226	−2.8057
$(1, \sigma^2)$	−2.8776	−2.8796	−2.8995	−2.8904	−2.8918
(μ, σ^2)	−2.8277	−2.8296	−2.8746	−2.8680	−2.8744

100 bootstrap samples are generated to get the stochastic approximation versions. Standard errors for the stochastic approximations are ranging from 0.003 to 0.008. For (μ, σ^2), it won't be necessary to compute the modified version of the criterion, but another stochastic approximation is done for the purpose of comparison.

All the estimators indicate that $(1, \sigma^2)$ fits the best, that is, the mean of the diameters stayed the same but the standard deviation increased. With this set of data we may conclude that the process does not seem to be operating normally.

3.2. Bias and MSE for (μ, σ_0^2)

Formal comparison of the various bootstrap methods will be made in this section including AIC and TIC in terms of the magnitude of bias and mean squared error, abbreviated as MSE for the case of (μ, σ_0^2), since an analytic comparison can be made easily. All the estimated criteria have $1/2(\log 2\pi + \log \sigma_0^2 + \hat{\sigma}^2/\sigma_0^2)$ in common, which can be regarded as the maximum average log-likelihood. Denote the maximum average log-likelihood by $l_{\max}$. Then the expected value of $l_{\max}$, $E\ l_{\max}$ is given by $1/2\{\log 2\pi + \log \sigma_0^2 + (1 - 1/n)\sigma_F^2/\sigma_0^2\}$.

Note that the difference between the estimated criterion and the expected overall discrepancy can be written as the sum;

$$(E\hat{O}D - l_{\max}) + (l_{\max} - E\ l_{\max}) + (E\ l_{\max} - EOD), \tag{3.2.1}$$

where EOD denotes the expected overall discrepancy and $E\hat{O}D$ denotes the estimated criterion. It is easy to check that $l_{\max} - E\ l_{\max} = 1/2\{\hat{\sigma}^2/\sigma_0^2 - (1-1/n)\sigma_F^2/\sigma_0^2\}$, and $E\ l_{\max} - EOD = -\sigma_F^2/n\sigma_0^2$. Then the bias calculation is straightforward and will be given below.

MSE is given by $E(E\hat{O}D - EOD)^2$, and is the sum of the variance of that estimated criterion and the square of the bias. Furthermore we know that the variance of the sample variance with devisor n is given by;

$$Var(\hat{\sigma}^2) = (\mu_{F,4} - \sigma_F^2)/n - 2(\mu_{F,4} - 2\sigma_F^2)/n^2 + (\mu_{F,4} - 3\sigma_F^2)/n^3, \quad (3.2.2)$$

where $\mu_{F,4}$ is the 4th central moment of F, that is, $E_F(Z - \mu_F)^4$. The following gives a summary of the results.

	$\hat{EOD} - l_{\max}$	$Bias$	$Var(\hat{EOD})$
AIC	$1/n$	$1/n - 1/n(\sigma_F^2/\sigma_0^2)$	$Var(\hat{\sigma}^2)/4\sigma_0^2$
TIC	$2\hat{\sigma}^2/n\sigma_0^2$	$-1/n^2(\sigma_F^2/\sigma_0^2)$	$(1+2/n)^2 Var(\hat{\sigma}^2)/4\sigma_0^2$
NB	$\hat{\sigma}^2/n\sigma_0^2$	$-(1/n + 1/n^2)(\sigma_F^2/2\sigma_0^2)$	$(1+1/n)^2 Var(\hat{\sigma}^2)/4\sigma_0^2$
MNB	$1/2n$	$1/2n - 1/n(\sigma_F^2/\sigma_0^2)$	$Var(\hat{\sigma}^2)/4\sigma_0^2$
PB	$\hat{\sigma}^2/n\sigma_0^2$	$-(1/n + 1/n^2)(\sigma_F^2/2\sigma_0^2)$	$(1+1/n)^2 Var(\hat{\sigma}^2)/4\sigma_0^2$

In this summary, NB denotes the nonparametric bootstrap, MNB denotes the modified nonparametric bootstrap, and PB denotes the parametric bootstrap estimator of the criterion. From the summary it is observed that all the estimators except TIC have bias of order $O(1/n)$, and all the estimators have MSE of order $O(1/n)$.

It seems that the effect of using the parametric structure of the approximating family in drawing the bootstrap sample does not reduce either the magnitude of bias nor the MSE. But when the size of the original sample is small, it frequently happens that most of the resample are identical so that the bootstrap version of the standard deviation is nearly zero. A parametric bootstrap is a definite alternative in that case.

A possible improvement in terms of bias reduction may be achieved by the more sophisticated bias-corrected versions of the boostrap method as suggested in Efron(1983). But our concern is restricted to naive ones just for the purpose of an easy implementation.

3.3. Normal Operating Family

Suppose that the operating model F itself is normally distributed. For the sake of computational simplicity, assume that the mean is 0 and the variance is 1. In this case EOD can be derived explicitly. Furthermore, for $(\mu, 1)$ all the biases and MSEs can be computed exactly. Besides, expected values and MSEs of AIC and PB can be derived explicitly.

3.3.1. Derivation of EOD

Recall that $n\tilde{\sigma}^2$ and $n\hat{\sigma}^2$ are distributed as the chi-square distribution with n and $(n-1)$ degrees of freedom respectively. By (2.5.4), we have

$$E \log \tilde{\sigma}^2 = \Psi(n/2) - \log(n/2), \tag{3.3.1}$$

$$E \log \hat{\sigma}^2 = \Psi\{(n-1)/2\} - \log(n/2), \tag{3.3.2}$$

(2.3.1), (2.3.2), (2.3.4), and (2.3.10) are the keys to compute EOD. Apply (3.3.1) and (3.3.2) with the assumption that $\sigma_F^2 = \sigma_0^2 = 1$, then the following summary will be obtained. All the submodels have $1/2(\log 2\pi + 1)$ in common, so we subtract it from the entries.

θ	$EOD - 1/2(\log 2\pi + 1)$
$(0, 1)$	0
$(\mu, 1)$	$1/2n$
$(0, \sigma^2)$	$1/2\{\Psi(n/2) - \log(n/2) + 2/(n-2)\}$
(μ, σ^2)	$1/2[\Psi\{(n-1)/2\} - \log(n/2) + 4/(n-3)]$

3.3.2. Expected Values of AIC and PB

The following gives a summary of the expected values $E(AIC/2n)$ and $E(PB)$.

θ	$E(AIC/2n) - 1/2(\log 2\pi + 1)$
$(0, 1)$	0
$(\mu, 1)$	$1/2n$
$(0, \sigma^2)$	$1/2\{\Psi(n/2) - \log(n/2) + 2/n\}$
(μ, σ^2)	$1/2[\Psi\{(n-1)/2\} - \log(n/2) + 4/n]$

θ	$E(PB) - 1/2(\log 2\pi + 1)$
$(0, 1)$	0
$(\mu, 1)$	$-1/2n^2$
$(0, \sigma^2)$	$1/2\{\Psi(n/2) - \log(n/2) + 1/(n-2)\}$
(μ, σ^2)	$1/2[\Psi\{(n-1)/2\} - \log(n/2) + 2/(n-3)]$

3.3.3. MSE of AIC and PB

MSE of the estimated criterion is given by, $E(E\hat{O}D - EOD)^2$. When we derive the MSEs of AIC and PB, (2.5.5) is used repeatedly. The following summary of the results will be obtained.

For (μ, σ_0^2), it is easy to see that AIC gives a smaller amount of MSE than PB, but the difference is of order $O(1/n^2)$. It will be seen in section 3.3.4 that the MSE of PB is smaller when $n = 5$ for the other two cases.

θ	AIC
$(0,1)$	$1/2n$
$(\mu,1)$	$1/2n - 1/2n^2$
$(0,\sigma^2)$	$1/4[\Psi'(n/2) + 4\{1/n - 1/(n-2)\}^2]$
(μ,σ^2)	$1/4[\Psi'\{(n-1)/2\} + 16\{1/n - 1/(n-1)\}^2]$

θ	PB
$(0,1)$	$1/2n$
$(\mu,1)$	$1/4(2/n + 3/n^2 - 1/n^4)$
$(0,\sigma^2)$	$1/4[\Psi'(n/2) + \{\Psi(n/2) - \log(n/2)\}^2]$
(μ,σ^2)	$1/4[\Psi'\{(n-1)/2\} + \{\Psi((n-1)/2) - \log(n/2)\}^2]$

3.3.4. A Numerical Study

From the manual of *Mathematica* and from Johnson and Kotz(1969), the following recursive relation and the following formulae are obtained.

$$\Psi(\nu + 1) = \Psi(\nu) + \nu^{-1}, \qquad (3.3.3)$$

$$\Psi(n) = \sum_{k=1}^{n-1} 1/k - \gamma, \qquad (3.3.4)$$

$$\Psi'(\nu) = \sum_{k=0}^{\infty} 1/(\nu + k)^2, \qquad (3.3.5)$$

where n is a positive integer and γ is the Euler's constant, which is around 0.577216. There are also remarkably good approximation formulae for $\Psi(\cdot)$ and $\Psi'(\cdot)$, that is,

$$\Psi(\nu) \approx \log(\nu - 1/2), \qquad (3.3.6)$$

$$\Psi'(\nu) \approx (\nu - 1/2)^{-1}. \tag{3.3.7}$$

Remark. From these approximation formulae, it is easy to see that both AIC and PB have MSEs of order $O(1/n)$.

With these formulae and the fact that $\Psi(1/2) = -\gamma - 2\log 2 = -1.963510$, the following summary for the case when $n = 5, 10, 20$ is obtained.

n	$\Psi(n/2)$	$\Psi\{(n-1)/2\}$	$\Psi'(n/2)$	$\Psi'\{(n-1)/2\}$
5	0.7032	0.4228	0.4904	0.6449
10	1.5061	1.3889	0.2213	0.2487
20	2.2572	2.1977	0.1052	0.1110

Table 1 contains the the summary of EOD, $E(AIC/2n)$, $E(PB)$, $MSE(AIC)$, and $MSE(PB)$ for $n = 5, 10, 20$. As expected in section 3.3.3, it is observed that AIC gives a smaller amount of MSE for $(\mu, 1)$, but the difference gets smaller as n increases. It is also observed that for $(0, \sigma^2)$, and for (μ, σ^2), PB gives a smaller amount of MSE when $n = 5$, but the situation gets reversed as n increases. It is clear from the table that the MSEs are of order $O(1/n)$.

3.4. Monte-Carlo Study

In this section a Monte Carlo study is provided to give a formal justification of the results given in section 3.3.3 and to give a numerical comparison of each method when a closed form expression is not available. For this study the standard normal distribution is taken as the operating model. Sample sizes are set at 5, 10, and 20 respectively. To get the stochastic approximation to the bootstrap estimators of the selection criterion, 100 bootstrap samples are drawn. This simulation study is run on Micro Vax II, and the total CPU time used is about 1 hour and a half.

Table 1. Exact values of EOD, $E(AIC/2n)$, $E(PB)$, $MSE(AIC)$, and $MSE(PB)$ for $n = 5, 10, 20$ when F is standard normal.

θ	n	EOD	$E(AIC/2n)$	$E(PB)$	$MSE(AIC)$	$MSE(PB)$
$(0,1)$	5	1.4189	1.4189	1.4189	0.1000	0.1000
	10	1.4189	1.4189	1.4189	0.0500	0.0500
	20	1.4189	1.4189	1.4189	0.0250	0.0250
$(\mu,1)$	5	1.5189	1.5189	1.3989	0.0800	0.1296
	10	1.4689	1.4689	1.4139	0.0450	0.0575
	20	1.4439	1.4439	1.4177	0.0238	0.0269
$(0,\sigma^2)$	5	1.6457	1.5124	1.5392	0.1404	0.1340
	10	1.4923	1.4673	1.4585	0.0560	0.0580
	20	1.4518	1.4462	1.4291	0.0263	0.0268
(μ,σ^2)	5	2.1722	1.5722	1.9254	0.5212	0.2221
	10	1.5944	1.5087	1.4841	0.0640	0.0743
	20	1.4665	1.4665	1.4317	0.0278	0.0305

To assess the overall sampling behaviour of each estimator, 1000 replications are made. Table 2, Table 3, and Table 4 give the average values of the estimated criteria so obtained and the MSE of each estimated criterion. The results are given right below the average value of the selection criteria. The standard errors of the average values are around 0.01, 0.007, and 0.005 for $n = 5$, $n = 10$, and $n = 20$ respectively.

Table 2. The average value of the estimated criterion and the estimated MSE when $n = 5$.

θ		$AIC/2n$	TIC	NB	MNB	PB
$(0,1)$	$\hat{EOD}$	1.4171	1.4171	1.4171	1.4171	1.4171
	MSE	0.1009	0.1009	0.1009	0.1009	0.1009
$(\mu,1)$	$\hat{EOD}$	1.5252	1.4877	1.4064	1.4252	1.4064
	MSE	0.0831	0.1637	0.1321	0.0918	0.1322
$(0,\sigma^2)$	$\hat{EOD}$	1.5097	1.4221	1.7238	1.8571	1.5357
	MSE	0.1405	0.1753	3.0829	3.1510	0.1406
(μ,σ^2)	$\hat{EOD}$	1.5773	1.4766	8.1841	8.1841	1.9308
	MSE	0.5164	0.6478	524.1161	524.1161	0.3802

Table 3. The average value of the estimated criterion and the estimated MSE when $n = 10$.

θ		$AIC/2n$	TIC	NB	MNB	PB
$(0,1)$	$E\hat{O}D$	1.4277	1.4277	1.4277	1.4277	1.4277
	MSE	0.0518	0.0518	0.0518	0.0518	0.0518
$(\mu,1)$	$E\hat{O}D$	1.4776	1.4693	1.4234	1.4276	1.4234
	MSE	0.0465	0.0669	0.0582	0.0481	0.0582
$(0,\sigma^2)$	$E\hat{O}D$	1.4742	1.4479	1.4447	1.4683	1.4477
	MSE	0.0582	0.0614	0.0659	0.0679	0.0600
(μ,σ^2)	$E\hat{O}D$	1.5174	1.4900	1.5334	1.5334	1.4935
	MSE	0.0694	0.0756	0.0908	0.0908	0.0739

Table 4. The average value of the estimated criterion and the estimated MSE when $n = 20$.

θ		$AIC/2n$	TIC	NB	MNB	PB
$(0,1)$	$E\hat{O}D$	1.4299	1.4299	1.4299	1.4299	1.4299
	MSE	0.0273	0.0273	0.0273	0.0273	0.0273
$(\mu,1)$	$E\hat{O}D$	1.4559	1.4546	1.4302	1.4309	1.4302
	MSE	0.0250	0.0302	0.0276	0.0250	0.0276
$(0,\sigma^2)$	$E\hat{O}D$	1.4531	1.4466	1.4321	1.4362	1.4333
	MSE	0.0276	0.0277	0.0280	0.0283	0.0281
(μ,σ^2)	$E\hat{O}D$	1.4786	1.4712	1.4449	1.4449	1.4436
	MSE	0.0283	0.0286	0.0302	0.0302	0.0301

With regards to AIC and PB, the values from Monte-Carlo study agree very well with the theoretical values obtained in the previous section. For $(\mu, 1)$ the closed form expression obtained is computed directly without bothering stochastic approximation. Therefore, NB and PB give the same average values.

As expected, NB shows an extremely wild behaviour when $n = 5$. In fact, it was observed that there are some cases where the bootstrap version of the standard deviation is so small that the logarithm of it blew up. In that case, another bootstrap sample is drawn to go on. The PB can be an alternative to overcome this kind of problem.

As n increases it can be observed that the MSE decreases in an order of $O(1/n)$. PB gives smaller MSE in many of cases when $n = 5$. But as n increases AIC seems to give the smallest MSE among the estimated criteria.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Professor Sakamoto for the permission to use the data set given in section 3. They are also grateful to referees for their valuable suggestions on improving this paper. Many thanks are also due to Mr. Jin-Ho Park for computational help.

REFERENCES

(1) Efron, B.(1979). Bootstrap Methods. *Annals of Statistics,* 7, 1-26.

(2) Efron, B.(1983). Estimating the Error Rate of a Prediction Rule: Improvements on Cross-Validation. *Journal of the American Statistical Association,* 78, 316-331.

(3) Efron, B.(1986). How Biased is the Apparent Error Rate of a Prediction Rule ? *Journal of the American Statistical Association,* 81, 461-470.

(4) Johnson, N.L. and Kotz, S.(1969). *Distributions in Statistics: Discrete Distributions,* Houghton Mifflin Co, Boston.

(5) Linhart, H. and Zucchini, W.(1986). *Model Selection,* Wiley, New-York.

(6) Sakamoto, Y., Ishiguro, M., and Kitagawa, G.(1986). *Akaike Information Criterion Statistics,* D. Reidel Publishing Co, Dordrecht/Tokyo.