Thailand Statistician
July 2006; 4: 85-92
http://statassoc.or.th
Contributed paper

Prediction Intervals for a Single Future Value of Normal and Non-Normal Variables

Sa-aat Niwitpong* and Penpak Pinkoompee
Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's Institute of Technology North Bangkok,10800, Thailand.
***Author for correspondence**, e-mail: snw@kmitnb.ac.th

Received: 1 August 2006
Accepted: 20 September 2006

Abstract

This paper investigates one-step-ahead prediction intervals for normal and non-normal variables. We propose studentized bootstrap and bootstrap percentile-t methods to construct one-step-ahead prediction intervals for normal and non-normal variables compared to a standard method. A minimum coverage probability 1- α and their expected lengths are used to select a preferable prediction interval. Monte Carlo simulation results show that not only does the bootstrap percentile-t give a better coverage probability but also give a shorter expected length.

Keywords: bootstrap, coverage probability, expected length.

1. Introduction

Our aim in this paper is to construct one-step-ahead prediction intervals for normal and non-normal variables. They are compared by a minimum coverage probability 1- α and an expected length. A one-step-ahead prediction interval for normal variable using a standard method can be found in e.g., Bikel and Doksum [1], Niwitpong [5] and Devore [2]. Niwitpong also proved that the coverage probability of a one-step-ahead prediction interval for a normal variable is close to 0.95 and is not dependent on the parameter values (μ, σ^2) . However, for one-step-ahead prediction intervals for non-normal variables, there has not been much research in this area. In this paper, we propose to use bootstrap techniques (Efron and Tibshirani [3]); studentized bootstrap and bootstrap percentile-t. They are presented in section 2. The simulation and numerical results are presented in section 3. The conclusion is presented in section 4.

2. A one-step-ahead prediction interval for X_{n+1}

Let $X_1, X_2, ..., X_n$ are independent identically distributed as $N(\mu, \sigma^2)$. Our aim is to construct a one-step-ahead prediction interval for X_{n+1}. Three methods are considered:

2.1 Standard method
2.2 Studentized bootstrap
2.3 Bootstrap percentile-t.

For the case where X_{n+1} is also normally distributed with mean μ and variance σ^2, their coverage probabilities are very closed to 1- α for all values of the (μ, σ^2) (Niwitpong [5]). For X_{n+1} which is non-normally distributed, we propose to construct a one-step-ahead prediction interval for X_{n+1} using methods 2.2 and 2.3. Descriptions of these methods are as follows.

2.1 Prediction interval for $\theta = X_{n+1}$ by the standard method

Bikel and Doksum [1] and Devore [2] explained the standard method for constructing a one-step-ahead prediction interval for X_{n+1}, which is also assumed to be normally distributed with mean μ and variance σ^2. A class of prediction unbiased predictor for $\hat{Y} = \hat{X}_{n+1}$ is $\bar{X}$ and the resulting prediction error is $\bar{X} - X_{n+1}$. The expected value of the prediction error is

$$E(\bar{X} - X_{n+1}) = E(\bar{X}) - E(X_{n+1}) = \mu - \mu = 0 .$$

Since X_{n+1} is independent of $X_1, X_2, ..., X_n$ and therefore independent of $\bar{X}$, the variance of the prediction error is therefore

$$Var(\bar{X} - X_{n+1}) = Var(\bar{X}) + Var(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right).$$

Moreover, the prediction error is a linear combination of independent normally distributed random variables, we therefore have $\bar{X} - X_{n+1} \sim N\left(0, [n^{-1} + 1]\sigma^2\right)$ .Thus

$$Z = \frac{(\overline{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}}$$

has a standard normal distribution and is independent of $V = (n-1)s^2 / \sigma^2$, which has a χ^2_{n-1} distribution. Devore has described that replacing σ by the sample standard deviation s (of $X_1, X_2, \ldots, X_n$) results in

$$T = \frac{Z}{\sqrt{\frac{V}{n-1}}} = \frac{\overline{X} - X_{n+1}}{s\sqrt{1 + \frac{1}{n}}}$$ ~ *t* distribution with *n* - 1 df.

Writing the acceptance for *T* region for 2 sided alternative test, then manipulating this formula as $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is manipulated for a confidence interval for μ, gives a one-step-ahead prediction interval for X_{n+1} gives the following result

$$\left[\overline{X} - t_{1-\frac{\alpha}{2},(n-1)} \cdot s\sqrt{1+n^{-1}} \;\; , \;\; \overline{X} + t_{1-\frac{\alpha}{2},(n-1)} \cdot s\sqrt{1+n^{-1}}\right] \qquad (1)$$

2.2 Prediction interval for $\theta = X_{n+1}$ by the studentized bootstrap method

Let $\theta = X_{n+1}$ is used to assign an approximate prediction interval to parameter θ of interest. Suppose that we are in the one-sample situation where the data are obtained by random sampling from an unknown distribution *F*, $F \rightarrow x = (x_1, x_2, \ldots, x_n)$, $\hat{\theta} = \hat{X}_{n+1}$ is an estimator of θ and $se(\hat{\theta})$ is the standard error of $\hat{\theta}$. Under most circumstances it turns out that as the sample size *n* grows large, the distribution of $\hat{\theta}$ becomes more and more normal, with mean near θ and variance $s\hat{e}^2$, written as $\hat{\theta} \sim N(\theta, s\hat{e}^2)$ or equivalently $Z = \frac{\hat{\theta} - \theta}{s\hat{e}} \sim N(0,1)$.

Let $z_{\frac{\alpha}{2}}$ donotes the $100 \cdot \frac{\alpha}{2}$ *th* percentile point of the standard normal distribution.

The $100 \cdot (1-\alpha)\%$ one-step-ahead prediction interval for $\theta = X_{n+1}$ is

$$\left[\hat{\theta}-Z_{(1-\frac{\alpha}{2})}\cdot s\hat{e}\ ,\ \hat{\theta}+Z_{(1-\frac{\alpha}{2})}\cdot s\hat{e}\right] \quad \text{...(2)}$$

Efron and Tibshirani [3, p.47] described the bootstrap algorithm for estimating standard errors, $s\hat{e}$, by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications and estimating the standard error of estimate of $\hat{\theta}$ by the empirical standard deviation of the replications. The result is called the bootstrap estimate of standard error, denoted by $s\hat{e}$, where *B* is the number of bootstrap samples used. The following steps are used to compute $s\hat{e}$;

2.2.1 Select *B* independent bootstrap samples $x^{*1}, x^{*2}, \ldots, x^{*B}$, each consisting of *n* data values drawn with replacement from X .

2.2.2 Evaluate the bootstrap estimate corresponding to each bootstrap sample,

$$\hat{\theta}^*(b) = s(x^{*b}) \quad b = 1,2,\ldots,B.$$

2.2.3 Estimate the standard error $se(\hat{\theta})$ by the sample standard deviation of the *B* replications;

$$s\hat{e} = \left\{ \frac{\sum_{b=1}^{B}\left[\hat{\theta}^*(b)-\hat{\theta}^*(\cdot)\right]^2}{B-1} \right\}^{1/2},$$

where $\hat{\theta}^*(\cdot) = \dfrac{\sum_{b=1}^{B}\hat{\theta}^*(b)}{B}$.

2.3 Prediction interval for $\theta = X_{n+1}$ by bootstrap percentile-t method

Let $\theta = X_{n+1}$ is used to assign an approximate prediction interval to parameter θ of interest. Suppose that we are in the one-sample situation where the data are obtained by random sampling from an unknown distribution *F*, $F \rightarrow x = (x_1, x_2, \ldots, x_n)$, $\hat{\theta} = \hat{X}_{n+1}$ is an estimator of θ and $se(\hat{\theta})$ is the standard error of $\hat{\theta}$.

2.3.1 Let the prediction error in each bootstrap is $e^{*b} = x_{n+1}^{*b} - \overline{x}^{*b}$,

$$b = 1,2,\ldots,B$$

2.3.2 Compute $t^{*b} = \dfrac{e^{*b}}{\hat{se}}$ and let $t_{\frac{\alpha}{2}}$ indicate $100 \cdot \dfrac{\alpha}{2}$ *th* percentile point of the *t*-distribution.

2.3.3 The prediction interval $100 \cdot (1-\alpha)\%$ for $\theta = X_{n+1}$ is

$$\left[\hat{\theta} - t^{*}_{(1-\frac{\alpha}{2})} \cdot \hat{se} \ , \ \hat{\theta} + t^{*}_{(1-\frac{\alpha}{2})} \cdot \hat{se} \right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

where $\hat{se}$ is computed from section 2.2.

3. Simulation and Numerical Results

In this section, the coverage probabilities for prediction intervals 1-3 are computed using Monte Carlo simulation. A minimum coverage probability 1- α and their expected lengths are used to select a preferable prediction interval. The prediction level is 95% and four distributions were used following Kakizawa [4] namely

3.1 the standard normal distribution

3.2 the contaminated normal distribution with density is

$$0.3 \times \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2} x^2 \right) + 0.7 \times \frac{1}{(20\pi)^{1/2}} \exp\left(-\frac{1}{20} x^2 \right)$$

3.3 the *t* distribution with degree of freedom *n*-1

3.4 the exponential distribution with $\lambda = 1$.

We study four sample sizes 30, 50, 100 and 250. For each sample size and distribution, M=10,000 simulation runs were used to evaluate the bootstrap replication corresponding to each bootstrap sample B=1,000 bootstrap replications were conducted. Numerical results are presented in Tables 1 and 2.

Table 1. The estimated coverage probabilities for each method of constructing a one-step-ahead prediction interval for X_{n+1}.

Distribution	Sample size	Standard Method			Studentized Bootstrap			Bootstrap Percentile -t		
		Lower tailed (2.5%)	Coverage (95%)	Upper tailed (2.5%)	Lower tailed (2.5%)	Coverage (95%)	Upper tailed (2.5%)	Lower tailed (2.5%)	Coverage (95%)	Upper tailed (2.5%)
Normal	30	0.0246	0.9503	0.0251	0.0357	0.9285	0.0358	0.0024	0.9954	0.0022
	50	0.0247	0.9505	0.0248	0.0306	0.9378	0.0316	0.0017	0.9967	0.0016
	100	0.0250	0.9500	0.0250	0.0288	0.9429	0.0283	0.0011	0.9979	0.0010
	250	0.0250	0.9500	0.0250	0.0257	0.9485	0.0258	0.0010	0.9980	0.0010
Contaminated Normal	30	8.5-e05	0.9998	8.1-e05	0.0342	0.9293	0.0365	0.0021	0.9957	0.0022
	50	3.9-e05	0.9999	4.1-e05	0.0300	0.9359	0.0341	0.0022	0.9957	0.0021
	100	1.7-e05	0.9999	1.7-e05	0.0283	0.9407	0.0310	0.0028	0.9958	0.0024
	250	9.3-e06	0.9999	9.3-e06	0.0268	0.9469	0.0263	0.0018	0.9960	0.0022
t	30	0.0259	0.9480	0.0261	0.0370	0.9260	0.0370	0.0038	0.9931	0.0031
	50	0.0257	0.9486	0.0257	0.0335	0.9337	0.0328	0.0016	0.9965	0.0019
	100	0.0254	0.9493	0.0253	0.0260	0.9446	0.0294	0.0012	0.9970	0.0018
	250	0.0251	0.9497	0.0252	0.0252	0.9487	0.0261	0.0010	0.9977	0.0013
Exponential	30	0	0.9414	0.0586	0.0012	0.9274	0.0714	0.0384	0.9606	0.0010
	50	0	0.9444	0.0556	0.0001	0.9341	0.0658	0.0355	0.9636	0.0009
	100	0	0.9456	0.0544	0	0.9405	0.0595	0.0333	0.9659	0.0008
	250	0	0.9473	0.0527	0	0.9463	0.0537	0.0311	0.9683	0.0006

Table 2. The estimated expected lengths of a one-step-ahead prediction interval for X_{n+1}.

Distribution	Sample size	Standard	Studentized Bootstrap	Bootstrap Percentile -t
Normal	30	4.1185	3.8484	1.8272
	50	4.0454	3.8799	1.7341
	100	3.9771	3.8997	1.6692
	250	3.9438	3.9115	1.6142
Contaminated Normal	30	9.2141	8.6123	4.1135
	50	9.0245	8.6616	3.9278
	100	8.8897	8.7077	3.7662
	250	8.8085	8.7418	3.5076
t	30	4.2640	3.9776	1.8894
	50	4.1187	3.9555	1.7639
	100	4.0169	3.937	1.6724
	250	3.9571	3.9276	1.6228
Exponential	30	4.0361	3.7040	1.8318
	50	3.9993	3.8021	1.6855
	100	3.9426	3.8564	1.5804
	250	3.9343	3.8863	1.5201

4. Conclusion

From the results in section 3, we can draw the following conclusions.

4.1 One-step-ahead prediction intervals for X_{n+1} when using the standard method and bootstrap percentile-t have minimum coverage probabilities of 0.95 when samples are from the normal and contaminated normal distributions.

4.2 Only the one-step-ahead prediction interval for X_{n+1} when using the bootstrap percentile-t method has a minimum coverage probability of 0.95 when samples are from the *t* and exponential distributions.

4.3 Most bootstrap percentile-t intervals have coverage probabilities more than standard method and when the sample size is large, coverage probability is large.

4.4 In this case study, bootstrap methods have expected lengths less than the standard method and when the sample size increase, the expected length decrease.

5. Acknowledgements

The authors are grateful to Dr. Gareth Clayton and the referees for their helpful comments.

Reference

[1] Bikel, Pl G. & Doksum, K. A. *Mathematical Statistics: Basic Ideas and Select Topic Volume I.* New Jersey : Prentice Hall.,2001

[2] Devore, L. *Probability and Statistics for Engineering and the Sciences*. New York, Duxbury Press, 2001.

[3] Efron, B. & Tibshirani R.J. *An Introduction to the Bootstrap*. New York: Chapman & Hall., 1994

[4] Kakizawa, Y. Valid Edgeworth Expansions of Some Estimators and Bootstrap Confidence Intervals in First-order Autoregression. *Journal of Time Series Analysis*, **20**; 1999: 343-359.

[5] Niwitpong, S. On Coverage Probability of the Prediction Interval for Normal Variable". Proceedings of KMITL International Conference on Science and Applied Science. : Bangkok , Thailand 8th-10th March, 2006.