ON COVERAGE PROBABILITY OF THE PREDICTION INTERVAL FOR NORMAL VARIABLE

Sa-aat Niwitpong*

Department of Applied Statistics, Faculty of Applied Science
King Mongkut's Institute of Technology North Bangkok,
Bangkok 10800, Thailand

ABSTRACT

This paper presents a coverage probability of a one-step-ahead prediction interval for a normal variable.This coverage probability is proved to be functionally independent of (μ, σ^2). Because of this functional independence, Monte Carlo simulation will include some variance reduction by setting $(\mu, \sigma^2) = (0, 1)$. This result will then valid for all possible values of (μ, σ^2). This leads to a great reduction in computation effort.

KEYWORDS: Coverage probability, prediction interval.

1. INTRODUCTION

Suppose $X = (X_1, X_2, ..., X_n)$ be independent and identically distributed random variables and $X_i \sim N(\mu, \sigma^2), i = 1, 2, 3, ..., n.$ A one-step-ahead prediction interval for X_{n+1}, where X_{n+1} is also a normally distributed with mean μ and variance σ^2 and is independent of X, is well-known see e.g., Bikel and Doksum [1]. This one-step-ahead prediction interval for X_{n+1} is very important in many applications, see e.g., Walpole et al. [4, pp. 241-243] and Bikel and Doksum [1, pp. 252-254].

As in Niwitpong [3], I have also derived a coverage probability of this prediction interval which is proved to be functionally independent of (μ, σ^2). This important result allows us to set μ equals zero and σ^2 equals one in Monte Carlo simulation and is valid for all possible parameter values of (μ, σ^2). This leads to a great reduction in computational effort. Section 2 reviews the method to construct prediction intervals for X_{n+1}. Section 3 gives the method to compute the coverage probability of a prediction interval for X_{n+1}. Section 4 presents Monte Carlo simulation results of the coverage probabilities of a prediction interval. The conclusion is in Section 5.

2. PREDICTION INTERVAL FOR A NORMAL VARIABLE

Bikel and Doksum [1] show that a (1 - α)100% prediction interval for X_{n+1} is

$$PI = \left[\bar{X} - t_{1-\frac{\alpha}{2}, n-1} s\sqrt{1+n^{-1}}, \bar{X} + t_{1-\frac{\alpha}{2}, n-1} s\sqrt{1+n^{-1}}\right] \quad (1)$$

* Corresponding Author Tel: 662 9132500-24 ext 4910; Fax 662 5856105
Email: snw@kmitnb.ac.th

where $t_{1-\frac{\alpha}{2}}$ is a $(1-\frac{\alpha}{2})th$ quantile of the t distribution,

$\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

In the next section, we prove that the coverage probability of a prediction interval for X_{n+1} is not depend on (μ, σ^2).

3. THE COVERAGE PROBABILITY OF PREDICTION INTERVAL

The unconditional coverage probability of PI for X_{n+1} in (1) is

$$\Pr(X_{n+1} \in PI) = \Pr\left[\bar{X} - t_{1-\frac{\alpha}{2},(n-1)} s\sqrt{1+n^{-1}} \le X_{n+1} \le \bar{X} + t_{1-\frac{\alpha}{2},(n-1)} s\sqrt{1+n^{-1}}\right]$$

$$= \Pr\left[(\bar{X}-\mu) - t_{1-\frac{\alpha}{2},n-1} s\sqrt{1+n^{-1}} \le X_{n+1} - \mu \le (\bar{X}-\mu) + t_{1-\frac{\alpha}{2},n-1} s\sqrt{1+n^{-1}}\right]$$

$$= \Pr\left[\frac{(\bar{X}-\mu)}{\sigma} - t_{1-\frac{\alpha}{2},n-1} \frac{s}{\sigma}\sqrt{1+n^{-1}} \le \frac{X_{n+1}-\mu}{\sigma} \le \frac{(\bar{X}-\mu)}{\sigma} + t_{1-\frac{\alpha}{2},n-1} \frac{s}{\sigma}\sqrt{1+n^{-1}}\right]$$

$$= \Pr\left[\bar{Z} - t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}} \le Z_{n+1} \le \bar{Z} + t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}}\right]$$

$$= \Phi\left[\bar{Z} + t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}}\right] - \Phi\left[\bar{Z} - t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}}\right] \qquad (2)$$

where $\frac{\bar{X}-\mu}{\sigma} = \frac{n^{-1}\sum_{i=1}^{n} X_i - \mu}{\sigma} = \frac{n^{-1}\left[\sum_{i=1}^{n} X_i - n\mu\right]}{\sigma} = \frac{\sum_{i=1}^{n}\left[(X_i-\mu)\sigma^{-1}\right]}{n} = \frac{\sum_{i=1}^{n} Z_i}{n} = \bar{Z}$,

$s_Z^2 = \frac{1}{\sigma^2(n-1)}\sum_{i=1}^{n}\left[X_i - \bar{X}\right]^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left[\sigma^{-1}(X_i-\mu) - \sigma^{-1}(\bar{X}-\mu)\right]^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left[Z_i - \bar{Z}\right]^2$,

and $\Phi(.)$ is a cumulative standard normal distribution.

The coverage probability of PI is therefore independent of (μ, σ^2). These important results allow us to set $(\mu, \sigma^2) = (0, 1)$ in Monte Carlo simulation. This leads to a great reduction in computation of the coverage probability of prediction interval PI.

4. MONTE CARLO SIMULATION

In this section, the coverage probabilities of PI is computed using Monte Carlo simulation. Now, we set the indicator $I_{PI}(X_{n+1})$ defined by $I_{PI}(X_{n+1}) = 1$ if $X_{n+1} \in PI$ and 0 otherwise. Therefore, from (2), we have (see, Klimov [2, p. 30]).

$$\Pr(X_{n+1} \in PI) = E(I_{PI}(X_{n+1})),$$

$$= E\left[\Phi\left[\bar{Z} + t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}}\right] - \Phi\left[\bar{Z} - t_{1-\frac{\alpha}{2},n-1} s_Z\sqrt{1+n^{-1}}\right]\right]. \qquad (3)$$

Now $P(X_{n+1} \in PI)$ can be estimated using Monte Carlo simulation. We set $\{W_i\}$ (i = 1, 2,..., M.) to be independent and identically distributed random variables having the same distribution of (2). We estimate $P(X_{n+1} \in PI)$ in (3) by

$$\frac{\sum_{i=1}^{M} W_i}{M}.$$

All simulations were performed using programs written in **R** with M =10000 and $\alpha = 0.05, n = 30, 50, 100, 250, \mu = \{0, 5, 10, 50, -5, -10, -50\}$ and $\sigma^2 = \{1, 5, 10, 50\}$. The estimated coverage probabilities of a prediction interval PI are reported in Table 1. From Table 1, the estimated coverage probabilities are very closed to a nominal value of 0.95 for all sample sizes and parameter values of (μ, σ^2) considered here. Therefore, we can set only $\mu = 0$ and $\sigma^2 = 1$ in Monte Carlo simulation. This result is valid for the parameter values of (μ, σ^2). This leads to a great reduction in computation effort.

5. CONCLUSION

The coverage probability of a one-step-ahead prediction interval for a normal variable is proved to be functionally independent of (μ, σ^2). A high speed computer simulation can be carried out by setting only $\mu = 0$ and $\sigma^2 = 1$.

Table1. The estimated coverage probabilities of a prediction interval PI for M = 10000 and $\alpha = 0.05$.

n	μ	σ^2	Coverage Probability
30	0	1	0.9506
50			0.9505
100			0.9499
250			0.9500
30	5	5	0.9504
50			0.9499
100			0.9501
250			0.9500
30	10	10	0.9502
50			0.9493
100			0.9507
250			0.9497
30	50	50	0.9509
50			0.9498
100			0.9500
250			0.9499
30	-5	5	0.9500
50			0.9497
100			0.9505
250			0.9499
30	-10	10	0.9503
50			0.9497
100			0.9497
250			0.9501
30	-50	50	0.9499
50			0.9503
100			0.9498
250			0.9502

6. ACKNOWLEDGEMENTS

The author is grateful to Dr. Gareth Clayton and the referees whose comment substantially improved the presentation of the paper.

REFERENCES

[1] Bikel, P. G. and Doksum, K. A. **2001** *Mathematical Statistics: Basic Ideas and Selected Topics Volume 1.* New Jersey, Prentice Hall.

[2] Kilmov, G. **1986** *Probability Theory and Mathematical Statistics.* Moscow, MIR Publishers Moscow.

[3] Niwitpong, S. **2005** Prediction Interval for an AR(1) Process Using Combined Predictors, *Thailand Statistician 3,* 3-11.

[4] Walpole et al. **2002** *Probability and Statistics for Engineers and Scientists.* New Jerseys, Prentice Hall.

Program R for computing the coverage probability and the expected length of a one-step-ahead prediction interval for Normal variable

```
pi.cov <- function(mu, sigma, n, M)
{
# Written by Sa-aat Niwitpong in November, 2005.
# This function computes a coverage probability of a prediction interval for Y[n+1]
# where Y[n+1] is a one-step-ahead predictor of Y[i] ~N(mu,sigma^2 ), i =1, 2, ### #  3,..,n.
# The expected length of this prediction interval is also included.
#  cov and ex denote the coverage probability and its expected length of this prediction # # interval
      cov <- rep(0, M)
      ex <- rep(0, M)
       for(i in 1:M) {
            y <- rnorm(n, mu, sigma)
            z <- (y - mu)/sigma
            z.mean <- mean((y - mu)/sigma)
            s.z <- var(z)
#.......................................................................................................
# Coverage probability for a PI and its expected length using a standard method.
#.......................................................................................................
           cov[i] <- pnorm(z.mean + qt(0.975, (n - 1)) * sqrt(1/n + 1) * sqrt(s.z)) -            pnorm(z.mean
+ qt(0.025, (n - 1)) * sqrt(1/n + 1) * sqrt(s.z))
            ex[i] <- 2 * qt(0.975, n - 1) * sqrt(1/n + 1) * sqrt(s.z)      #
      }
      cov.mean <- mean(cov)
      ex.mean <- mean(ex)
      out <- cbind(cov.mean, ex.mean)
      out
}
```