International Journal of Mathematical Archive-2(7), July - 2011, Page: 1077-1079
Available online through www.ijma.info ISSN 2229 – 5046

AN APPROXIMATION TO THE CDF OF STANDARD NORMAL DISTRIBUTION

Ramu Yerukala*, Naveen Kumar Boiroju and M. Krishna Reddy

Department of Statistics, Osmania University, Hyderabad, India

Email: ramu20@gmail.com, nanibyrozu@gmail.com, reddymk54@gmail.com

(Received on: 20-06-11; Accepted on: 04-07-11)

ABSTRACT

In this paper, a new approximation function proposed for the cumulative distribution function of standard normal distribution using a feedforward neural network. Efficiency of the proposed function is measured using maximum absolute error, root mean squared error, mean absolute error, mean absolute percentage error and the results compared with the existing approximation functions in the literature.

Keywords: *Approximation, Artificial neural networks, Cumulative distribution function, Standard normal distribution.*

---

1. INTRODUCTION:

The normal distribution, also called the Gaussian distribution, is an important family of continuous probability distributions, applicable in many fields. The standard normal distribution is the normal distribution with a mean of zero and a variance of one [7]. The standard normal probability density function is given by

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right) \qquad -\infty < Z < \infty \tag{1.1}$$

And the cumulative probability distribution function (CDF) of the standard normal distribution is given by

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy \tag{1.2}$$

The cumulative distribution function of standard normal distribution has no closed form. Most of the statistical books give the values of $\Phi(z)$ in a tabular form and this table called the standard normal table. From this table we can find the probability under the normal curve. Several authors gave approximations for the CDF of standard normal distribution. We consider the approximations for the CDF have been proposed by Polya (1945) [9], Lin (1990) [8], Aludaat and Alodat (2008) [1], Bowling et. al., (2009) [2]. We propose a new function using neural networks to approximate the CDF of standard normal distribution. The results of the proposed function are evaluated using maximum absolute error (Max. AE), mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE).

2. REVIEW OF APPROXIMATIONS TO THE CDF:

In this section, we present some typical approximation formulas for the cumulative standard normal distribution:

(i) Polya (1945)

$$\Phi_1(z)\approx\frac{1}{2}\left[1+\left\{1-\exp\left(\frac{-2z^2}{\pi}\right)\right\}^{\frac{1}{2}}\right] \tag{2.1}$$

(ii) Lin (1990)

$$\Phi_2(y)\approx\left[1+\exp\{y\}\right]^{-1} \qquad \text{where } y=\frac{4.2\pi\, z}{(9-z)} \tag{2.2}$$

Corresponding author: Ramu Yerukala*, *E-mail: ramu20@gmail.com

(iii) Aludaat and Alodat (2008)

$$\Phi_3(z) \approx 0.5 + 0.5\sqrt{1 - \exp\left(-\sqrt{\frac{\pi}{8}}\,z^2\right)} \tag{2.3}$$

(iv) Bowling et. al., (2009), One Parameter Logistic Equation:

$$\Phi_4(z) \approx \frac{1}{1 + \exp\{-1.702\,z\}} \tag{2.4}$$

3. APPROXIMATION OF CDF USING NEURAL NETWORKS:

In this section, a new formula for calculating the function $\Phi(z)$ is obtained using neural networks. Since $\Phi(z)$ is symmetric about zero, it is sufficient to approximate only for all values of Z>0.

Feedforward neural networks are efficient nonparametric models for function approximation. It has been shown that a feed forward neural network, with a single hidden layer, can approximate any given continuous function on any compact subset to any degree of accuracy, providing that a sufficient number of hidden layer neurons are used [3]. Neural networks have been recently seen as attractive tools for developing efficient solutions for many real world problems in function approximation. Feedforward neural networks are known to be universal approximators of nonlinear functions, they are the most popular artificial neural networks structures for generating many input-output nonlinear mappings [4] [5] [6]. A feed forward neural network used to approximate the CDF of standard normal distribution. We have taken standard normal values from Z=0 to 3 with increment 0.01 as inputs and their corresponding $\Phi(z)$ values computed using NORMSDIST() function are used as the target of the neural network. The network structure contains an input layer with one neuron, a hidden layer with three neurons and an output layer with one neuron. Backpropogation learning method is used to train and test the network. PASW 18 software is used for the neural networks model building. The neural network model is obtained as:

$$\Phi_5(z) \approx 0.5 - 1.136\tanh(-0.2695z) + 2.47\tanh(0.5416z) - 3.013\tanh(0.4134z) \tag{3.1}$$

The proposed approximate function is validated by comparing it to the NORMSDIST () function in order to check the accuracy and range of validity. Performance of this function is evaluated using Max. AE, MAE, RMSE and MAPE. These results compared with those of existing functions in the following section.

4. COMPARISON OF FUNCTIONS:

We have carried out an exercise to determine the accuracy of these formulas. We have chosen the NORMSDIST () function of Microsoft Excel 2003 as the standard for computing the CDF of the standard normal distribution. The performance of the above functions is given in Table 1.

Table-1: Error measures for the approximating functions to CDF

Function	Max. AE	MAE	RMSE	MAPE
$\Phi_1(z)$	0.0031	0.0016	0.0019	0.1749
$\Phi_2(z)$	0.0067	0.0018	0.0030	0.2591
$\Phi_3(z)$	0.0020	0.0011	0.0012	0.1205
$\Phi_4(z)$	0.0095	0.0065	0.0070	0.7643
$\Phi_5(z)$	0.0013	0.0009	0.0009	0.0935

As shown in Table 1, the accuracy is quite high for the proposed method. Most of the probabilities using the proposed approximation coincide to the probabilities computed by the NORMSDIST () function. The maximum error is 0.0013 for the Z values between 0 and 3. It is also noted that the approximate function is relatively simple, compared to others available in the literature. We believe our approximation equation is easy to compute and highly useful for practitioners and users of standard normal distribution. The relation among the functions is observed with respect to root mean squared error is

$$RMSE(\Phi_5) < RMSE(\Phi_3) < RMSE(\Phi_1) < RMSE(\Phi_2) < RMSE(\Phi_4)$$

And the maximum absolute error observed for Φ_1 at z=1.65, Φ_2 at z=0.44, Φ_3 at z=1.88, Φ_4 at z=0.57, and Φ_5 at z=2.64. The minimum of maximum error is obtained from the proposed function; therefore the proposed function approximates CDF of standard normal distribution efficiently in the range of $0 \le Z \le 3$.

REFERENCES:

[1] Aludaat1, K. M. and Alodat, M. T. , A Note On Approximating The Normal Distribution Function, Applied Mathematical Sciences, Vol. 2, 9 (2008), 425 – 429.

[2] Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S. and Cho, B. R., A logistic approximation to the cumulative normal distribution, Journal of Industrial Engineering and Management, 2(1) (2009), 114-127.

[3] Chentouf, R., Jutten, C., Maignanb, M., Kanevsky, M., Incremental neural networks for function approximation, Nuclear Instruments and Methods in Physics Research - A 389 (1997), 268-270.

[4] Hassoun, M. H., Fundamentals of artificial neural networks, Prentice-Hall of India (2007).

[5] Haykin, S. S., “Neural Networks: A Comprehensive Foundation”, Upper Saddle River, N. J., Prentice Hall (1999).

[6] Hornik, K. Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators, Neural Networks, 2 (1989), 359-366.

[7] Johnson, N. I., Kotz, S. and Balakrishnan, N., Continuous univariate distributions. John Wiley & Sons (1994).

[8] Lin, J. T., A simpler logistic approximation to the normal tail probability and its inverse. Applied Statistics 39 (1990), 255–257.

[9] Polya, G., Remarks on computing the probability integral in one and two dimensions. Proceeding of the first Berkeley symposium on mathematical statistics and probability, (1945), 63-78.

****