ISSN 2229-5518

Approximations to Standard Normal Distribution Function

Ramu Yerukala and Naveen Kumar Boiroju

Abstract:

This paper presents three new approximations to the cumulative distribution function of standard normal distribution. The accuracy of the proposed approximations evaluated using maximum absolute error and the same is compared with the existing approximations available in the literature. The proposed approximations assure minimum of three decimal value accuracy and are simple to use and easily programmable.

Keywords: Normal distribution, Maximum absolute error and Box-plots.

1. Introduction

The most widely used probability distribution in statistical applications is the normal or Gaussian distribution function. The cumulative distribution function (cdf) of standard normal distribution is denoted by $\Phi(z)$ and is given by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx$$

The cdf of normal distribution mainly used for computing the area under normal curve and approximating the t, Chi-square, F and other statistical distributions for large samples. The cdf of normal distribution does not have a closed form. For this reason, a lot of works have been on the development of approximations and bounds for the cdf of normal distribution. Approximations are commonly used in many applications, where exact solutions are numerically involved, not tractable or for simplicity and convenience (Krishnamoorthy, 2014).

There are number of approximations for computing the cumulative probabilities of standard normal distribution at arbitrary level of accuracy available in the literature. Some of these approximations were previously studied by Polya (1945), Hart (1957 & 1966), Tocher (1963), Zelen and Severo (1964), Page (1977), Hammakar (1978), Lin (1989 & 1990), Norton (1989), Waissi and Rossin (1996), Byrc (2002) Aludaat and Alodat (2008), Winitzki (2008), Yerukala et al. (2011) and Choudhury (2014). We propose three new approximation functions to approximate the cdf of standard normal distribution. The accuracy of each approximation was assessed in terms of its maximum absolute error (Max. AE) when compared with the NORMSDIST () function for the values of $0 \le Z \le 5$.

2. Approximations to CDF of Normal Distribution

This section presents the historical review of different approximations to CDF of normal distribution available in the literature for positive values of z.

1. Polya (1945):

$$\Phi_1(z) \approx \frac{1}{2}\left\{1 + \left(1 - e^{\frac{-2z^2}{\pi}}\right)^{\frac{1}{2}}\right\}$$

2. Hart (1957):

$$\Phi_2(z) \approx \frac{1}{\sqrt{2\pi}}\left\{\frac{e^{\frac{-z^2}{\pi}}}{z + 0.8e^{-0.4z}}\right\}$$

3. Tocher (1963):

$$\Phi_3(z) \approx \frac{e^{2kz}}{1 + e^{2kz}}, \text{ where } k = \sqrt{\frac{2}{\pi}}.$$

4. Zelen and Severo (1964):

$$\Phi_4(z) \approx 1 - \left(a_1 t - a_2 t^2 + a_3 t^3\right)\left(\frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}}\right)$$

where $t = (1 + 0.33267z)^{-1}$, $a_1 = 0.4361836$, $a_2 = 0.1201676$ and $a_3 = 0.937298$.

5. Hart (1966):

$$\Phi_5(z) \approx 1 - \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}\, z}\left(1 - \frac{\frac{\sqrt{1+bz^2}}{1+az^2}}{P_0 z + \sqrt{P_0^2 z^2 + e^{\frac{-z^2}{2}}\frac{\sqrt{1+bz^2}}{1+az^2}}}\right)$$

where $a = \frac{1+\sqrt{1-2\pi^2+6\pi}}{2\pi}$, $b = 2\pi a^2$ and $P_0 = \sqrt{\frac{\pi}{2}}$.

6. Page (1977):

$$\Phi_6(z) \approx 0.5\{1 + \tanh(y)\}$$

where $y = \sqrt{\frac{2}{\pi}}\, z\left(1 + 0.044715z^2\right)$.

7. Hammakar (1978):

$$\Phi_7(z) \approx 1 - 0.5\left\{1 - \left(1 - e^{-y^2}\right)^{0.5}\right\}$$

where $y = 0.806z(1 - 0.018z)$.

8. Lin (1989):

$$\Phi_8(z) \approx 1 - 0.5\left(e^{-0.717z - 0.416z^2}\right)$$

9. Norton (1989):

$$\Phi_9(z) \approx \begin{cases} 1 - 0.5e^{-\left(\frac{z^2 + 1.2z^{0.8}}{2}\right)} & ;0 \le z \le 2.7 \\ \dfrac{1}{\sqrt{2\pi}\,z} e^{-\frac{z^2}{2}} & ;z > 2.7 \end{cases}$$

10. Lin (1990):

$$\Phi_{10}(z) \approx 1 - \frac{1}{1 + e^y}$$

where $y = 4.2\pi\left\{\frac{z}{9 - z}\right\}$, $0 \le z < 9$.

11. Waissi and Rossin (1996):

$$\Phi_{11}(z) \approx \frac{1}{1 + \exp\left(-\sqrt{\pi}\left(0.9z + 0.0418198z^3 - 0.0004406z^5\right)\right)}$$

12. Byrc (2002A):

$$\Phi_{12}(z) \approx \frac{(4-\pi)z + \sqrt{2\pi}(\pi - 2)}{(4-\pi)z^2\sqrt{2\pi} + 2\pi z + 2\sqrt{2\pi}(\pi - 2)} e^{\frac{-z^2}{2}}$$

13. Byrc (2002B):

$$\Phi_{13}(z) \approx \frac{z^2 + a_1 z + a_2}{\sqrt{2\pi}\,z^3 + b_1 z^2 + b_2 z + 2a_2} e^{\frac{-z^2}{2}}$$

where $a_1 = 5.575192695$, $a_2 = 12.77436324$, $b_1 = 14.38718147$ and $b_2 = 31.53531977$.

14. Aludaat and Alodat (2008):

$$\Phi_{14}(z) \approx 0.5 + 0.5\sqrt{1 - e^{-\sqrt{\frac{\pi}{8}}z^2}}$$

15. Winitzki (2008):

$$\Phi_{15}(z) \approx \frac{1}{2}\left\{1 + \left[1 - \exp\left\{\frac{-\left(\frac{z^2}{2}\right)\left(\frac{4}{\pi} + 0.147\frac{z^2}{2}\right)}{1 + 0.147\frac{z^2}{2}}\right\}\right]^{\frac{1}{2}}\right\}$$

16. Yerukala et al. (2011):

$$\Phi_{16}(z) \approx \begin{cases} 0.5 - 1.136H_1 + 2.47H_2 - 3.013H_3 & ;0 \le z \le 3.36 \\ 1 & ;z > 3.36 \end{cases}$$

where $H_1 = \tanh(-0.2695z)$, $H_2 = \tanh(0.5416z)$ and $H_3 = \tanh(0.4134z)$.

17. Yerukala (2012):

$$\Phi_{17}(z) \approx \begin{cases} 0.46375418 + 0.065687194H_{11} \\ -0.602383931H_{12} & ;0 \le z \le 3.6 \\ 1 & ;z > 3.6 \end{cases}$$

where $H_{11} = \tanh(1.280022196 - 0.720528073z)$, $H_{12} = \tanh(0.033142223 - 0.682842425z)$.

18. Choudhury (2014):

$$\Phi_{18}(z) \approx 1 - \frac{1}{\sqrt{2\pi}} \frac{e^{\frac{-z^2}{2}}}{0.226 + 0.64z + 0.33\sqrt{z^2 + 3}}$$

The accuracy of the above functions discussed in the section 4.

3. New Approximations to CDF of Normal Distribution

In this section, we present three new approximations to cdf of normal distribution. A new formula for standard normal distribution function $\Phi(z)$ is obtained using neural networks. Since $\Phi(z)$ is symmetric about zero and $\Phi(-z) = 1 - \Phi(z)$. It is sufficient to approximate only for all the values of $z \ge 0$. Hence, the proposed approximations are given only for the non-negative values of z. The following approximation developed using neural networks methodology (Yerukala, 2012). A neural networks model with an input layer consisting single input node, a hidden layer and an output layer with one node is considered. Input node takes the values of z from zero to five with an increment of 0.001, whereas, the output node represents the corresponding cumulative probability for a given z. The network is trained using backpropagation algorithm by taking the pair of observations as z value and its cumulative probability value p computed using NORMSDIST () function of MS Excel software. The hidden neurons are adjusted according to the sufficiently low training and testing error. The resulting approximation to cdf of standard normal distribution using neural networks is given below.

$$\Phi_{19}(z) \approx \frac{1}{1 + e^{-y}}; \quad 0 \le z \le 5$$

where $y = 0.125 + 3.611H_1 - 4.658H_2 + 4.982H_3$, $H_1 = \tanh(0.043 + 0.2624z)$, $H_2 = \tanh(-1.687 - 0.519z)$ and $H_3 = \tanh(-1.654 + 0.5044z)$.

Recently, researchers have mainly concentrated on developing different approximations both computationally tractable and sufficiently accurate by combining two or more existing approximations (Choudhury et al. 2007). We propose a combined approximation of two existing approximations F_1 and F_2 in the form of $wF_1 + (1-w)F_2$ where the weight w is determined using the least squares method over the range of z. The proposed approximation has the form,

$$\Phi_{20}(z) \approx w\Phi_5 + (1-w)\Phi_{13}; \quad z > 0 \text{ and } w = 0.268.$$

Third approximation is a simple modification to the Choudhury (2014) approximation and it has the form

$$\Phi_{21}(z) \approx 1 - \frac{\exp(-z^2/2)}{\left(\frac{44}{79} + \frac{8}{5}z + \frac{5}{6}\sqrt{z^2 + 3}\right)}$$

ISSN 2229-5518

Accuracy of the proposed approximations discussed in the following section.

4. Results and Discussion

This section presents the approximation errors of the listed functions in the above two sections. Each approximation is evaluated on the basis of maximum absolute error from NORMSDIST() function within the given range of z values. Table 1 presents the maximum absolute error and corresponding z value of each of the approximation. From this table, it is clear that the proposed approximation $\Phi_{20}(z)$ is more efficient than all the listed approximations and $\Phi_{19}(z)$ and $\Phi_{21}(z)$ are very good alternatives to the existing approximations. The error distribution of the approximations shows that, the proposed approximations have the error very close to zero (Figure 1). Maximum absolute error for $\Phi_{19}(z)$ and $\Phi_{21}(z)$ is observed at around the origin, whereas the same is observed for $\Phi_{20}(z)$ at z=1.9. The approximation $\Phi_{20}(z)$ have reduced the error about 86% and 60% respectively as compared with the $\Phi_5(z)$ and $\Phi_{13}(z)$. Approximation $\Phi_{20}(z)$ provides 5 to 9 correct decimals within the given range of z (Figure 2). The approximation $\Phi_{21}(z)$ is derived by improving the coefficients of the function proposed by Choudhury (2014), the resulting approximation not only improved the approximation accuracy but also simplified the formula at a great extent.

Finally, the proposed approximations are very simple and provides a minimum accuracy of three decimal places. Simplicity of these approximations enables their application over a wide range of analytical studies at a reasonable accuracy levels. These approximations are easy to programme and simple to use on a handheld calculator.

Table 1. Maximum absolute error and corresponding z value of the approximations

Function	Max. A E	Occurrence of Max. AE at Z=z
$\Phi_1(z)$	3.15E-03	1.64 to 1.66
$\Phi_2(z)$	4.30E-03	0.28 to 0.31
$\Phi_3(z)$	1.77E-02	1.7 to 1.77
$\Phi_4(z)$	1.15E-05	0.51 to 0.54
$\Phi_5(z)$	5.32E-05	1.02 to 1.06
$\Phi_6(z)$	1.79E-04	1.22 to 1.27 and 2.56 to 2.62
$\Phi_7(z)$	6.23E-04	0.33 and 0.34
$\Phi_8(z)$	6.59E-03	0.39
$\Phi_9(z)$	8.07E-03	0.87 to 0.89
$\Phi_{10}(z)$	6.69E-03	0.44 and 0.45
$\Phi_{11}(z)$	4.37E-05	1.14 and 1.15
$\Phi_{12}(z)$	7.18E-04	1.07 to 1.12
$\Phi_{13}(z)$	1.87E-05	1.47 to 1.57
$\Phi_{14}(z)$	1.97E-03	1.83 to 1.93
$\Phi_{15}(z)$	6.20E-05	2.19 to 2.23
$\Phi_{16}(z)$	1.25E-03	2.58 to 2.69
$\Phi_{17}(z)$	1.17E-04	3.67 and 3.68
$\Phi_{18}(z)$	1.93E-04	0.00
$\Phi_{19}(z)$	1.61E-04	0.00
$\Phi_{20}(z)$	7.54E-06	1.89 to 1.91
$\Phi_{21}(z)$	1.10E-04	0.06

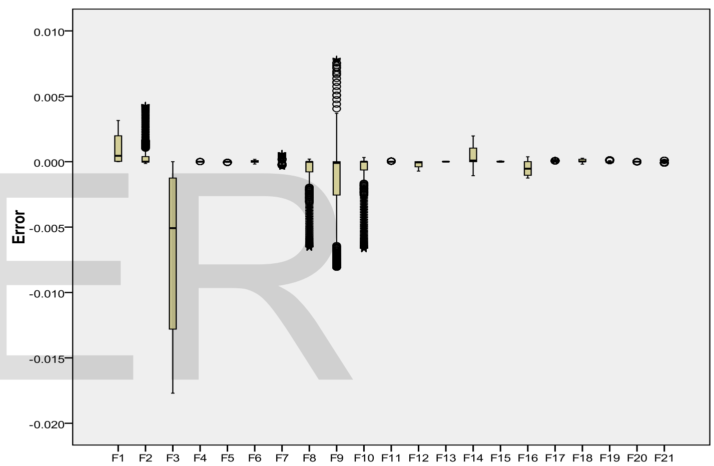

Figure 1. Error Distribution of the Approximations

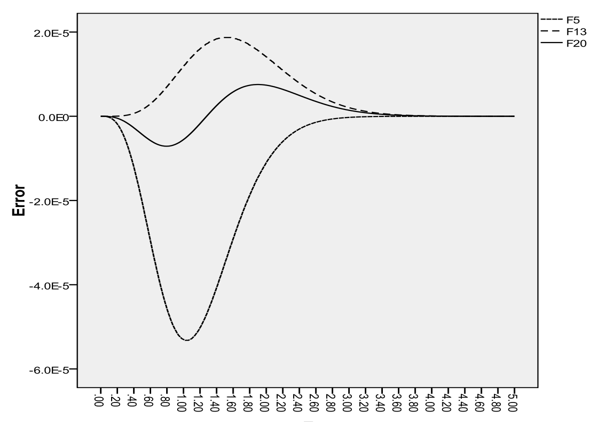

Figure 2. Error distribution of the approximations $\Phi_5(z)$, $\Phi_{13}(z)$ and $\Phi_{20}(z)$

Acknowledgements:

We are very thankful to Dr. M. Krishna Reddy, Professor (Rtd.), Department of Statistics, Osmania

University, Hyderabad for his valuable suggestions and encouragement.

References:

1. A. Choudhury and P. Roy, "A Fairly Accurate Approximation to the Area under Normal Curve," *Communications in Statistics-Simulation and Computation*, 38, pp. 1485-1492, 2009.
2. A. Choudhury, "A Simple Approximation to the Area under Standard Normal Curve", *Mathematics and Statistics* 2(3)147-149, 2014.
3. A. Choudhury, S. Ray, and P. Sarkar, "Approximating the Cumulative Distribution Function of the Normal Distribution," *Journal of Statistical Research*, 41(1), 59–67, 2007.
4. B.I. Yun, "Approximation to the Cumulative Normal Distribution using Hyperbolic Tangent based Function," *J. Korean Math. Soc.* 46, No. 6, pp. 1267-1276, 2009.
5. E. Page, "Approximating to the Cumulative Normal function and its Inverse for use on a Pocket Calculator," *Applied Statistics*, 26, 75-76, 1977.
6. G. Polya, "Remarks on computing the probability integral in one and two dimensions", *1st Berkeley Symp. Math. Statist. Prob.*, pp. 63-78, 1945.
7. G. R. Waissi, and D. F. Rossin, "A Sigmoid Approximation of the Standard Normal Integral," *Applied Mathematics and Computation* 77, pp 91-95, 1996.
8. H. C. Hammakar, "Approximating the Cumulative Normal Distribution and its Inverse," *Applied Statistics*, 27, 76-77, 1978.
9. J. T. Lin, "A Simpler Logistic Approximation to the Normal Tail Probability and its Inverse", *Appl. Statist.* 39, No.2, pp. 255-257, 1990.
10. J. T. Lin, "Approximating the Normal Tail Probability and its Inverse for use on a Pocket-Calculator," *Applied Statistics*, 38, 69-70, 1989.
11. K. D. Tocher, "The Art of Simulation", *English University Press*, London, 1963.
12. K. M. Aludaat and M. T. Alodat, "A Note on Approximating the Normal Distribution Function," *Applied Mathematical Sciences*, Vol. 2, no.9, 425-429, 2008.
13. K.. Krishnamoorthy, "Modified Normal-Based Approximation to the Percentiles of Linear Combination of Independent Random Variables with Applications," Available online at http://www.ucs.louisiana.edu/~kxk4695/CS-MNA-R.pdf, 2014
14. M. Zelen and N. C. Severo, "Probability Function (925-995) in Handbook of Mathematical Functions", Edited by M. Abramowitz and I. A. Stegun, Applied Mathematics Series, 55, 1964.
15. R. G. Hart, "A close approximation related to the error function (in Technical Notes and Short Paper)," *Math. Comput.* 20 (96) 600, 1966.
16. R. G. Hart, "A formula for the approximation of definite integrals of the normal distribution function," *Mathematical Tables and other Aids to Computation* 11(60), 1957.
17. R. M. Norton, "Pocket-Calculator Approximation for Area under the Standard Normal Curve," *the American Statistician*, Vol. 43, No. 1, 1989.
18. R. Yerukala, "Functional Approximation using Neural Networks," Unpublished thesis, Department of Statistics, Osmania University, Hyderabad, Telangana, India, 2012.
19. R. Yerukala, N. K., Boiroju and M. K. Reddy, "An Approximation to the CDF of Standard Normal Distribution," *International Journal of Mathematical Archive* 2(7):1077-1079, 2011.
20. S. Winitzki, "A Handy approximation for the Error Function and its Inverse," Available online at http://sites.google.com/site/winitzki/sergei-winitzkis-files, 2008 .
21. W. Bryc, "A Uniform Approximation to the Right Normal Tail Integral", *Applied Mathematics and Computation* 127, pp 365-374, 2002.

Ramu Yerukala, ANURAG Group of Institutions,
Hyderabad, Telangana, India-500 088.
Email: ramu20@gmail.com