

A Simple Approximation for Normal Distribution Function

Medhat Edous, Omar Eidous*

Department of Statistics, Yarmouk University, Irbid, Jordan

Copyright©2018 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This paper proposes an approximation to the standard normal distribution function. The introduced approximation formula is very simple and it has a very acceptable accurate. By comparing the proposed approximation with other existing approximations, it can be observed that the proposed one has a simple, easily computable formula and it gives a good accurate with maximum absolute error equals 0.000444.

Keywords Normal Distribution, Cumulative Distribution Function, Approximation

1. Introduction

The normal distribution function considered to be one of the most important continuous probability distribution. It plays a significant role in probability theory and mathematical statistics. The importance of the normal distribution comes from two reasons; the first one is that it has a wide range of applications in many areas of sciences because a lot of phenomena follow a normal distribution; another reason is that there is a theory that the average of the samples values takes the form of a normal distribution even so that the variable itself does not follow the normal distribution (central limit theorem).

The standard normal distribution is a normal distribution with mean equal zero and standard deviation equal one. If a random variable X is normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$, then its probability density function is,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$

The cumulative normal distribution function (c.d.f) is,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Since $\Phi(x)$ cannot be expressed in a closed form for all values of x and because the practical needs for computing

its values, then this provides the motivation for researchers and practitioners to approximate $\Phi(x)$.

The most basic expectation towards an approximation is that it is highly accurate. In the literature, there are many approximations to $\Phi(x)$ that focus mainly on the approximation accuracy. In general, these efforts have resulted in highly accurate functions, without taking some other features -like simplicity- of the approximation into account. In this paper, we will introduce a new formula with two terms to approximate $\Phi(x)$. It should be mentioned here that the proposed approximation is very simple, accurate and invertible.

Many efforts have been done to find an approximation for $\Phi(x)$, see for example, Johnson *et al.* [1], Polya [2] and Eidous and Al-Hanandeh [3]. Some of this work was done to replace the tables of cumulative standard normal probabilities that are found in most probability and statistics books. In these tables, numerical techniques were used to determine the cumulative standard normal probabilities since the integration of the normal density function cannot be done by elementary methods. Other work has been done for the purpose of curiosity. Let

$$\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

then

$$\begin{aligned} \Phi(x) &= 0.5 + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \\ &= 0.5 + \Phi^*(x), \quad x \geq 0. \end{aligned}$$

The problem of approximating $\Phi(x)$ is equivalent to the problem of approximating $\Phi^*(x)$. This paper proposes a simple approximation for $\Phi(x)$, $x \geq 0$ depends on the Polya [2]'s approximation. The new approximation has less maximum absolute error than that of Polya's approximation. The Polya's approximation of $\Phi(x)$ is,

$$\Phi_1(x) = 0.5 \left(1 + \sqrt{1 - e^{-\frac{2}{\pi}x^2}} \right),$$

with maximum absolute error equal 0.0031.

Eidous and Al-Salaman [4] gave the following approximation for $\Phi(x)$,

$$\Phi_2(x) = 0.5 \left(1 + \sqrt{1 - e^{-\frac{5}{8}x^2}} \right),$$

with maximum absolute error equal 0.0018.

Eidous and Al-Hanandeh [3] gave another approximation for $\Phi(x)$, which is given by,

$$\Phi_3(x) = 0.5 \left(1 + \sqrt{1 - e^{-\frac{81}{130}x^2}} \right).$$

The maximum absolute error of $\Phi_3(x)$ equals to 0.0016.

Lin [5] gave the following approximation for $\Phi(x)$,

$$\Phi_4(x) = \frac{1}{1 + e^{-4.2\pi x/(9-x)}}.$$

The maximum absolute error of $\Phi_4(x)$ is 0.0067. Bowling *et al.* [6]'s approximation of $\Phi(x)$ is,

$$\Phi_5(x) = \frac{1}{1 + e^{-1.702x}}.$$

With maximum absolute error equal 0.0095.

While the above approximations used one term in x to approximate $\Phi(x)$, there are many others approximations -in the literature- for $\Phi(x)$ with more than one term and give more accurate. See for example Eidous and Abu-Shareefa [7] at which most of these approximations have been collected.

2. The Proposed Approximation

This paper organizes and develops an approximation of the cumulative distribution function of the standard normal distribution $\Phi(x)$, which is defined as the following

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Going back to Johnson *et al.* [1], the Polya's approximation is originally represents an upper bound for $\Phi(x)$, that is,

$$\Phi(x) \leq 0.5 \left(1 + \sqrt{1 - e^{-\frac{2}{\pi}x^2}} \right) = \Phi_1(x).$$

Therefore, Polya used the above upper bound as an approximation to $\Phi(x)$. That is, $\Phi_1(x) \approx \Phi(x), \forall x \geq 0$ with maximum absolute error 0.0031. In order to improve the Polya's approximation, Eidous and Al-Salaman [4] proved that

$$\sqrt{1 - e^{-\frac{5}{8}x^2}} \leq \sqrt{1 - e^{-\frac{2}{\pi}x^2}}$$

and Eidous and Al-Hanandeh [3] showed that

$$\sqrt{1 - e^{-\frac{81}{130}x^2}} \leq \sqrt{1 - e^{-\frac{5}{8}x^2}} \leq \sqrt{1 - e^{-\frac{2}{\pi}x^2}}.$$

That is,

$$\Phi_3(x) \leq \Phi_2(x) \leq \Phi_1(x).$$

The usage of $\Phi_2(x)$ and $\Phi_3(x)$ as approximations of $\Phi(x)$ give more accurate than that of $\Phi_1(x)$, with maximum absolute errors 0.0018 and 0.0016 respectively.

To obtain a more accurate approximation for $\Phi(x)$, we suggest to change the coefficient of x^2 in the exponent term of the above approximation formulas. Instead of using a constant coefficient, we suggest to use a linear function in x as follows. Let $a = 0.647 - 0.021x$ then the proposed approximation of $\Phi(x)$ is,

$$\Phi_6(x) = 0.5 \left(1 + \sqrt{1 - e^{-ax^2}} \right).$$

In order to find the maximum absolute error of $\Phi_6(x)$, define the function $h(x)$ as the following,

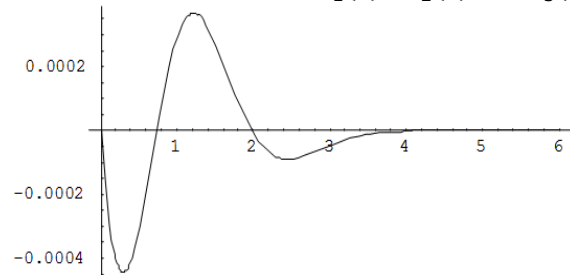
$$\begin{aligned} h(x) &= \Phi(x) - \Phi_6(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt - 0.5\sqrt{1 - e^{-ax^2}}, x \geq 0. \end{aligned}$$

The first derivation of $h(x)$ with respect to x is

$$\begin{aligned} h'(x) &= \frac{-x^2}{\sqrt{2\pi}} - \left[\frac{(0.3235 - 0.01575x)xe^{-(0.647x^2 - 0.021x^3)}}{\sqrt{1 - e^{-(0.647x^2 - 0.021x^3)}}} \right]. \end{aligned}$$

By equating $h'(x)$ with zero and solve it with respect to x , one obtain the two roots, $x = 0.29757$ and $x = 1.22578$. The values of $h(x)$ at $x = 0.29757$ and at $x = 1.22578$ are -0.000443987 and 0.000369022 respectively, which indicates that the maximum absolute error of $\Phi_6(x)$ is 0.000443987.

Graph (1) below plots the values of x and $h(x) = \Phi(x) - \Phi_6(x)$. The resulting curve represents the maximum error (in its magnitude) for $\Phi_6(x)$. It is clear that the absolute maximum error of $\Phi_6(x)$ is 0.000443987. By comparing this maximum error with that of the others approximations given in Section (1), we conclude that the absolute maximum error of $\Phi_6(x)$ is less than the absolute maximum errors of $\Phi_1(x), \Phi_2(x), \dots, \Phi_5(x)$.



Graph 1. The maximum error between the exact $\Phi(x)$ and its approximation $\Phi_6(x)$.

3. Comparison

We have carried out a comparison between our approximation $\Phi_6(x)$ and the other approximations $\Phi_1(x)$, $\Phi_2(x)$, $\Phi_3(x)$, $\Phi_4(x)$ and $\Phi_5(x)$. This comparison has been done for the values of x between 0.2 and 4.2 with step 0.4. The values of the different approximations are given in Table 1 below. The corresponding error between $\Phi(x)$ and the different approximations are also provided to compare the accuracy of them easily. From Table 1, we can see that our approximation $\Phi_6(x)$ is the best one for almost all values of x . Nearly, in all cases the associated error of $\Phi_6(x)$ is less than that of $\Phi_1(x)$, $\Phi_2(x)$, $\Phi_3(x)$, $\Phi_4(x)$ and $\Phi_5(x)$.

Table 1. The exact value of $\Phi(x)$ and the corresponding values of its approximations $\Phi_1(x)$, $\Phi_2(x)$, $\Phi_3(x)$, $\Phi_4(x)$, $\Phi_5(x)$ and $\Phi_6(x)$. The corresponding value of Error for each approximation is also given.

x	Normal $\Phi(x)$	Polya $\Phi_1(x)$	Eidous1 $\Phi_2(x)$	Eidous2 $\Phi_3(x)$	Lin $\Phi_4(x)$	Bowling $\Phi_5(x)$	Proposed $\Phi_6(x)$
0.2	0.579260	0.579283	0.578565	0.578446	0.579996	0.584288	0.579662
Error	-0.000024	0.000694	0.000814	-0.000535	-0.005028	0.000128	-0.000402
0.6	0.725747	0.726284	0.724435	0.724127	0.726497	0.735206	0.725932
Error	-0.000537	0.001312	0.001620	-0.000388	-0.009459	0.000484	-0.000185
1.0	0.841345	0.843119	0.840859	0.840481	0.841428	0.845796	0.841055
Error	-0.001774	0.000486	0.000864	0.000200	-0.004451	0.000881	0.000290
1.4	0.919243	0.922154	0.920191	0.919861	0.918996	0.915506	0.918912
Error	-0.002911	-0.000948	-0.000618	0.000382	0.003737	0.001051	0.000331
1.8	0.964070	0.967141	0.965834	0.965613	0.963883	0.955366	0.963970
Error	-0.003071	-0.001765	-0.001543	0.000218	0.008704	0.001025	0.000100
2.2	0.986097	0.988390	0.987709	0.987593	0.986045	0.976897	0.986161
Error	-0.002293	-0.001613	-0.001497	0.000045	0.009200	0.001096	-0.000064
2.6	0.995339	0.996608	0.996330	0.996282	0.995344	0.988170	0.995421
Error	-0.001270	-0.000991	-0.000943	-0.000015	0.007169	0.001248	-0.000082
3.0	0.998650	0.999187	0.999098	0.999082	0.998658	0.993976	0.998694
Error	-0.000537	-0.000447	-0.000432	-0.000013	0.004674	0.000932	-0.000044
3.4	0.999663	0.999841	0.999818	0.999814	0.999665	0.996942	0.999678
Error	-0.000178	-0.000155	-0.000151	-3.03343×10^{-6}	0.002721	-0.000600	-0.000015
3.8	0.999928	0.999975	0.999970	0.999969	0.999927	0.998449	0.999931
Error	-0.000047	-0.000042	-0.000041	4.68454×10^{-7}	0.001478	-0.003829	-3.01528×10^{-6}
4.2	0.999987	0.999997	0.999996	0.999996	0.999986	0.999214	0.999987
Error	-0.000010	-9.2734×10^{-6}	-9.13288×10^{-6}	6.23296×10^{-7}	0.000772	-0.008786	-2.53803×10^{-7}

REFERENCES

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

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

[3] Eidous and Al-Hanandeh (2018). A new one-Term Approximation to the Standard Normal Distribution. *Submitted*.

[4] Eidous, O. M. and Al-Salman, S. A. (2016). One-Term Approximation for Normal Distribution Function. *Mathematics and Statistics*, 4(1): 15-18.

[5] Lin, J. T. (1990). A simple Logistic approximation to the normal tail probability and its inverse. *Applied Statistics*, 39, 255-257.

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

[7] Eidous, O. M. and Abu-Shareefa, R. (2018). New approximations for standard normal distribution function. To be appear in *Communications in Statistics – Theory and Methods*.