Approximations for CDF of Standard Normal $\Phi(z)$

Omar M. Eidous[*]

omarm@yu.edu.jo

Areen Bani-Salameh

Abstract

In this paper, new approximations for the standard normal cumulative distribution function $\Phi(z)$ have been proposed. We focused on a set of approximations that take the following form,

$$0.5\left(1+\sqrt{1-e^{-\mathrm{A(z)}}}\right),$$

where $\mathrm{A(z)}$ is a polynomial function in z consisting of only two terms. The accuracy of the proposed approximations was compared with the accuracy of some existing approximations in the literature that take the same form. The comparison relied on calculating two well-known measures: the maximum absolute error (MAAE) and the mean absolute error (MEAE). The numerical results showed that the proposed approximations outperformed the approximations that were studied and belongs to family of approximations given in the above formula.

Key words: Normal distribution, Approximations, Cumulative distribution function, Maximum absolute error (MAAE), Mean absolute error (MEAE).

[*] Department of Statistics, Faculty of Science, Yarmouk University, Irbid – Jordan.

Received: 11/9/2024 . Accepted: 13/1 /2025 .

تقريبات لدالة التوزيع التراكمي للمتغير الطبيعي القياسي $\Phi(z)$

عمر اعدوس*

عرين بني سلامة

ملخص

في هذا البحث، تم اقتراح تقريبات جديدة لدالة التوزيع التراكمي الطبيعي القياسي $\Phi(z)$. ركزنا على مجموعة من التقريبات التي تأخذ الشكل التالي،

$$0.5\left(1+\sqrt{1-e^{-A(z)}}\right)$$

حيث $A(z)$ هي دالة متعددة الحدود تتكون من حدين فقط .تمت مقارنة دقة التقريبات المقترحة بدقة بعض التقريبات الموجودة في الأدبيات والتي تأخذ نفس الشكل.السابق. اعتمدت المقارنة على حساب مقياسين معروفين وهما: الخطأ المطلق الأقصى (MAAE) ومتوسط الخطأ المطلق (MEAE) وقد اظهرت النتائج العددية أن التقريبات المقترحة تفوقت على التقريبات التي تم دراستها والتي تنتمي إلى عائلة التقريبات المذكورة في الصيغة أعلاه.

الكلمات المفتاحية: التوزيع الطبيعي، التقريبات، دالة التوزيع التراكمي، الخطأ المطلق الأقصى (MAAE)، متوسط الخطأ المطلق (MEAE)

* قسم الإحصاء، كلية العلوم، جامعة اليرموك، إربد– الأردن.

تاريخ تقديم البحث: 2023/9/11.   تاريخ قبول البحث: 2025/1/13م .

Introduction:

The standard normal distribution (Gaussian distribution) is one of the most famous continuous distribution and holds particular importance in statistical theory. Due to the importance of this distribution, it has been studied from various aspects, including attempts to find approximations (see, Boiroju and Rao, 2014; Soranzo and Epure, 2014; Eidous and Abu-Shareefa, 2020; Hanandeh and Eidous, 2022; Eidous and Ananbeh, 2022 and Soranzo et al., 2023) or precise upper and lower bounds (see, Bercu, 2020; Eidous, 2022; Lipoth et al., 2022; Eidous, 2023 and Ananbeh and Eidous, 2024) for its cumulative distribution function $\Phi(z)$, as its exact formula cannot be determined. The normal distribution has many applications in most scientific fields, including, for example, engineering, genetics, psychology, biology, medicine, physics, reliability analysis, and statistics. Most books on mathematical statistics or probability have addressed the study of this distribution (see for example, Johnson et al., 1995). If the random variable Z follows a standard normal distribution, then its probability density function (pdf) is,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty.$$

Based on this formula, the probability that the random variable Z is less than or equal to a certain value z represents the cumulative distribution function (CDF) of the standard normal distribution, and is given by,

$$\Phi(z) = P(Z < z) \qquad = \int_{-\infty}^{z} \varphi(t)dt$$

$$= 0.5 + \int_{0}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt . \qquad z \geq 0$$

Since $\Phi(z)$ is symmetric around $z = 0$, then $\Phi(z) = 1 - \Phi(-z)$ for $z < 0$. The above integral has no closed form solution, so numerical techniques are used to calculate its value. There is a large literature body for approximating $\Phi(z)$ based on their accuracy level (Eidous and Abu-Hawwas, 2021; Eidous and Al-Rawash, 2021 and Soranzo et al., 2023).

The two useful measures for computing the accuracy of any approximation of $\Phi(z)$ are the maximum absolute error $(MAAE)$ and the mean absolute error $(MEAE)$ (see, Eidous and Abu-Shareefa, 2020), which are given as follows:

If $\widehat{\Phi}(z)$ is the approximation of $\Phi(z)$ then the $MAAE$ of $\widehat{\Phi}(z)$ is,

$$MAAE\left(\widehat{\Phi}(z)\right) = \max_z\left|\Phi(z) - \widehat{\Phi}(z)\right|$$

$$= \max_z\left|\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt - \widehat{\Phi}(z)\right|.$$

and the $MEAE$ of $\widehat{\Phi}(z)$ is,

$$MEAE\left(\widehat{\Phi}(z)\right) = \frac{\sum_z\left|\Phi(z) - \widehat{\Phi}(z)\right|}{n}.$$

The symbol n represents the number of z values chosen to calculate $MEAE$. For example, if we want to calculate $MEAE$ for $z = 0(0.001)5$ then $n = 5001$, and the z values are: 0, 0.001, 0.002, …, 5.

In this paper, new two approximations of the form $0.5\left(1+\sqrt{1-e^{-A(z)}}\right)$ for $\Phi(z)$ have been proposed, where $A(z)$ is a polynomial function in z consisting of only two terms.

Some Existing Approximations and Discussion

There are many approximations for $\Phi(z)$ in the literature that take the following form,

$$0.5\left(1+\sqrt{1-e^{-A(z)}}\right)$$

where $A(z)$ is a polynomial function in z. Some of these approximations focused on considering $A(z)$ as a polynomial with only one term. Among these approximations, we mention those presented by Polya, 1949, Aludaat and Alodat, 2008, Eidous and Al-Salman, 2016 and Hanandeh and Eidous, 2021, who took the values of $A(z)$ equal $2z^2/\pi$, $\sqrt{\pi/8}z^2$, $5z^2/8$ and $81z^2/130$ respectively (see also, Eidous and Bani-Salameh, 2024). However, we focused in this paper on approximations with two terms for a polynomial $A(z)$. We listed the following three approximations of $\Phi(z)$, which take the above form with two terms for $A(z)$,

- **Cadwell (1951)**

$$\Phi_1(z) = 0.5\left(1+\sqrt{1-e^{-A_1(z)}}\right), \qquad z \geq 0,$$

where $A_1(z) = \frac{2}{\pi}z^2 + \frac{2(\pi-3)}{3\pi^2}z^4$. The maximum absolute error of $\Phi_1(z)$ equals 0.0065 occurs at $z = 1.655$.

- **Malki (2017)**

$$\Phi_2(z) = 0.5\left(1 + \sqrt{1 - e^{-A_2(z)}}\right),$$

where $A_2(z) = 0.6349114z^2 - 0.0073962z^4$. The approximation $\Phi_2(z)$ improves Caldwell's (1951) approximation, $\Phi_1(z)$. The maximum absolute error of $\Phi_2(z)$ equals 0.00017 occurs at $z = 0.489$.

- **Edous and Eidous (2018)**

$$\Phi_3(z) = 0.5\left(1 + \sqrt{1 - e^{-A_3(z)}}\right), \qquad 0 \le z \le 30.8.$$

where $A_3(z) = 0.647z^2 - 0.021z^3$. Again, the approximation $\Phi_3(z)$ improves the approximation, $\Phi_1(z)$. The maximum absolute error of $\Phi_3(z)$ equals 0.00044, which occurs at $z = 0.2976$.

Excluding the approximation $\Phi_1(z)$, the flaw of the other two approximations $\Phi_2(z)$ and $\Phi_3(z)$ is that these two approximations do not work for all positive values of z. The approximation, $\Phi_2(z)$ working only for the values of z, such that, $0.6349114z^2 - 0.0073962z^4 \ge 0$, which implies, $z \le \sqrt{\frac{0.6349114}{0.0073962}} = 9.26514$. [28] did not pointed out to this deficiency of his suggested approximation.

As Edous and Eidous (2018) have been pointed out in their paper that the approximation $\Phi_3(z)$ working only for $0 \le z \le 31.38$. The approximation $\Phi_3(z)$ is defined only for the values of z, so that, $0.64459z^2 - 0.02054\,z^3 \ge 0$, which gives, $z \le 31.38$.

Proposed Approximations and Discussion

As a new approximation with the same form as $0.5\left(1 + \sqrt{1 - e^{-A(z)}}\right)$ with two terms polynomial for $A(z)$ and which defined on the entire positive values of z, we proposed the following approximation,

$$\Phi_{Prop1}(z) = 0.5\left(1 + \sqrt{1 - e^{-A_{Prop1}(z)}}\right), \qquad z \ge 0,$$

where $A_{Prop1}(z) = 0.623z^2 + 0.00001\,z^4$. The maximum absolute error of $\Phi_{Prop1}(z)$ equals to 0.00163216, which occurs at $z = 0.60194$.

Figure (1) gives the error curve between $\Phi_{Prop1}(z)$ and the exact $\Phi(z)$ (i.e. $h_{Prop1}(z) = \Phi(z) - \Phi_{Prop1}(z)$) for $0 \le z \le 5$. It is found that $|h_{Prop1}(z)| < 3.0 \times 10^{-7}$ for $z > 5$. It is obvious that the above approximation improves the accuracy of approximation, $\Phi_1(z)$. While the maximum absolute error of $\Phi_1(z)$ is 0.00646, it is 0.00163216 (25.27%) for $\Phi_{Prop1}(z)$. That is,

$$\max_z |h_{Prop1}(z)| = 0.2527 \max_z |h_1(z)|$$

or

$$\max_z |h_1(z)| = 3.9579 \max_z |h_{Prop1}(z)|.$$

As another approximation and based on numerical analysis for the absolute value of the error functions $h_1(z)$, $h_2(z)$, $h_3(z)$ and $h_{Prop1}(z)$, which were calculated (not presented here) for $z = 0(0.01)15$. It was found that the absolute values of $h_2(z)$ are less than the absolute values of the corresponding error functions of the other approximations for $0 \le z \le 2.2$, the absolute values of $h_3(z)$ are less than the absolute values of the corresponding error functions of the other approximations for $2.2 < z \le 9.26$, and the absolute values of $h_{Prop1}(z)$ are less that the absolute values corresponding to the other approximations for $z > 9.26$. Accordingly, we propose the following new piecewise approximation,

$$\Phi_{Prop2}(z) = \begin{cases} \Phi_2(z), & if\ 0 \le z \le 2.2 \\ \Phi_3(z), & if\ 2.2 < z \le 9.26 \\ \Phi_{Prop1}(z), & if\ z > 9.26 \end{cases}$$

$$= \Phi_2(z) I_{[0,\ 2.2]}(z) + \Phi_3(z) I_{(2.2,\ 9.26]}(z) + \Phi_{Prop1}(z) I_{(9.26\ ,\infty)}(z)$$

where $I_B(z)$ is the indicator function of z with respect to a set B, which is given by,

$$I_B(z) = \begin{cases} 1 & if\ z \in B \\ 0 & if\ z \notin B \end{cases}.$$

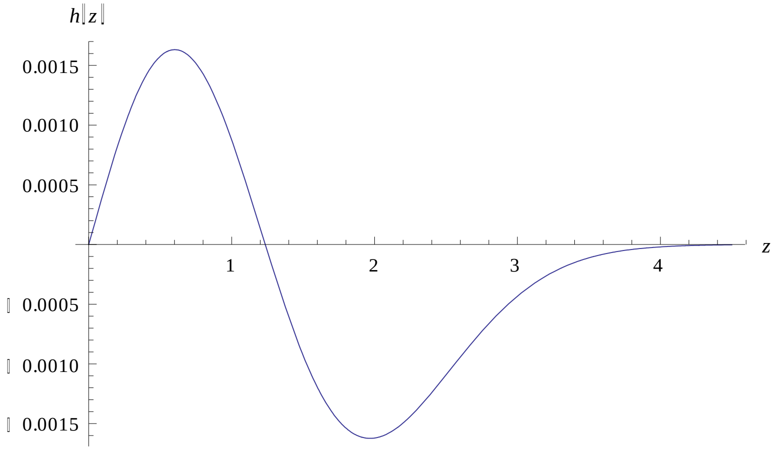

Figure (1) The curve represents the difference between $\Phi_{Prop1}(z)$ and $\Phi(z)$. That is, the curve represents $h(z) = \Phi(z) - \Phi_{Prop1}(z)$.

The maximum absolute error of $\Phi_{Prop2}(z)$ is 0.00017, which occurs at $z = 0.489$ (See Figure 2). The approximations $\Phi_2(z)$ and $\Phi_{Prop2}(z)$ have the same maximum absolute error. However, the mean absolute error of $\Phi_{Prop2}(z)$ is lower than that of $\Phi_2(z)$ as shown in Table (1). In addition, $\Phi_{Prop2}(z)$ is defined for all nonnegative values of the argument z, while $\Phi_2(z)$ is only defined for $0 \le z \le 9.26514$. These two features make the use of $\Phi_{Prop2}(z)$ more reliable than using $\Phi_2(z)$.

For a simple comparison, the $MAAE$ for the various approximations mentioned above have been summarized in Table (1). For a deeper insight, the $MEAE$ for each approximation is also calculated and the results are included in Table (1). The measure $MEAE$ for each approximation was calculated for $z = 0(0.001)5$ (i.e. $n = 5001$). To study the sensitivity of choosing n on the values of $MEAE$ for each approximation, we consider choosing $z = 0(0.01)5$ (i.e. $n = 501$) and $z = 0(0.1)5$ (i.e. $n = 51$). The values of $MEAE$ were very close to each other for the different selections of n as shown in Table (1).

From Table (1), it is clear that the proposed estimator $\Phi_{Prop2}(z)$ is more accurate than the others approximations on the basis of the two measures, $MAAE$ and $MEAE$. It is worthwhile to mention here that all numerical calculations were performed using Mathematica, Version 7.

Additionally, the value of $\Phi(z)$ obtained from Mathematica was considered to be the exact value of $\Phi(z)$.

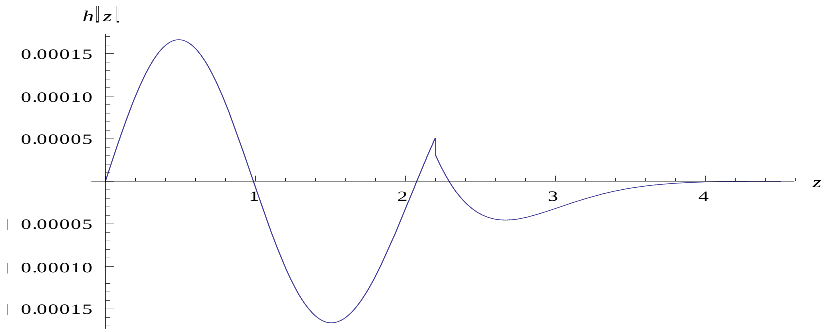

Figure (2) The curve represents the difference between $\Phi_{Prop2}(z)$ and $\Phi(z)$. That is, $h(z) = \Phi(z) - \Phi_{Prop2}(z)$.

Table (1) The maximum absolute error ($MAAE$) and the mean absolute error ($MEAE$) for the approximations $\Phi_1(z)$ to $\Phi_3(z)$ and the proposed approximations $\Phi_{Prop1}(z)$ and $\Phi_{Prop2}(z)$. The values of $MEAE$ are calculated for $n = 5001,\ 501$ and 51.

	$\Phi_1(z)$	$\Phi_2(z)$	$\Phi_3(z)$	$\Phi_{Prop1}(z)$	$\Phi_{Prop2}(z)$
$MAAE$	0.00646	0.00017	0.00044	0.00163	0.00017
$MEAE\ (n = 5001)$	0.002035	0.000087	0.00011	0.000667	0.000052
$MEAE\ (n = 501)$	0.002038	0.000087	0.00011	0.000668	0.000052
$MEAE\ (n = 51)$	0.001999	0.000085	0.00011	0.000655	0.000051

Conclusion:

In this article, two new approximations of the standard normal cumulative distribution function, $\Phi(z)$ are developed and proposed. Based on the two well-known criteria, the maximum absolute error (MAAE) and the mean absolute error (MEAE), It is found that the performance of the proposed approximation $\Phi_{Prop2}(z)$ is dominant when compared to other approximations considered in this study.

References:

Aludaat, K. M. and Alodat, M. T. (2008). A note on approximating the normal distribution function. Applied Mathematical Sciences, 2(9), 425-429.

Ananbeh, E. A. and, Eidous, O. M. (2024). New simple bounds for standard normal distribution function. Communications in Statistics-Simulation and Computation, 1-8. https://doi.org/10.1080/03610918.2024.2326596.

Bercu, G. (2020). New refinements for the error function with applications in diffusion theory. Symmetry, 12. doi:10.3390/sym12122017.

Boiroju, N. K., & Rao, K. R. (2014). Logistic approximations to standard normal distribution function. Assam Statistical Review, 28(1), 27-40.

Cadwell J. H. (1951). The Bivariate normal integral. Biometrika, 38, 475-479.

Edous, M. and Eidous, O. (2018). A simple approximation for normal distribution function. Mathematics and Statistics, 6 (4), 47-49.

Eidous, O. (2022). Improvements of Polya upper bound for cumulative standard normal distribution and related functions. arXiv:2205.03485.

Eidous, O. (2023). New tightness lower and upper bounds for the standard normal distribution function and related functions. Mathematical Methods in the Applied Sciences, 46 (14), 15011-15019.

Eidous, O., and Abu-Hawwas, J. (2021). An accurate approximation for the standard normal distribution function. Journal of Information and Optimization Sciences, 42 (1), 17–27.

Eidous, O. and Abu-Shareefa M. (2020). New approximations for standard normal distribution function. Communications in Statistics - Theory and Methods, 49 (137), 1-18.

Eidous, O. M. and Al-Rawash, M. (2022). Approximations for standard normal distribution function and its invertible. arXiv:2206.12601.

Eidous, O. and Al-Salman, S. (2016). One-term approximation for normal distribution function. Mathematics and Statistics, 4(1), 15-18.

Eidous, O. M. and Ananbeh, E. A. (2022). Approximations for cumulative distribution function of standard normal. Journal of Statistics and Management Systems, 25 (3), 541-547.

Eidous, O. and Bani-Salameh, A. (2024). Accurate approximations to the cumulative normal distribution function. JP Journal of Fundamental and Applied Statistics, 16 (1 & 2), 1-10.

Hanandeh, A. and Eidous, O. (2021). A New one-term approximation to the standard normal distribution. Pakistan Journal of Statistics and Operation Researches, 17(2), 381-385.

Hanandeh, A., and Eidous, O. M. (2022). Some improvements for existing simple approximations of the normal distribution function. Pakistan Journal of Statistics and Operation Research, 18(3), 555-559.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Continuous univariate distributions, volume 2 (289). John wiley & sons.

Lipoth, J., Tereda, Y., Papalexiou, S. and Spiteri, R. (2022). A new very simply explicitly invertible approximation for the standard normal cumulative distribution function. AIMS Mathematics, 7(7): 11635–11646. DOI:10.3934/math.2022648.

Malki, A. (2017). Improvement Bryc's approximation to the cumulative distribution function. Int. J. Swarm Intel. Evol. Comput. Dio: 10.4172/2090-4908.1000151.

Pólya, G. (1949). Remarks on computing the probability integral in one and two dimensions. In Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability (No. 1, p. 63). Berkeley: University of California Press.

Soranzo, A., and Epure, E. (2014). Very simply explicitly invertible approximations of normal cumulative and normal quantile function. Applied Mathematical Sciences, 8(87), 4323-4341.

Soranzo, A., Vatta, F., Comisso, M., Buttazzoni, G. and Babich, F. (2023). Explicitly invertible approximations of the Gaussian Q-function: A survey. IEEE Open Journal of the Communications Society, 4, 3051-3101.