

# On HPM Approximations for the Cumulative Normal Distribution Function

V. K. Shchigolev

Department of Theoretical Physics, Ulyanovsk State University, Ulyanovsk, 432000, Russian Federation

Copyright ©2019 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** In this paper, some new approximations to the cumulative distribution function of the standard normal distribution via the He's homotopy perturbation method are proposed. There are several methods which provide an approximation of the integral in the formula for the cumulative distribution function by different numerical methods. For the same purpose, we first establish a differential equation of the second order that the cumulative distribution function satisfied subjected with the certain initial conditions. Then we apply the Homotopy Perturbation Method to solve the Cauchy problem for the governing equation. As well known, the result of solving an equation by this method and the convergence rate greatly depend on the choice of homotopy applied. Therefore, we consider two cases in this work. In one case, we construct the homotopy from the idea of simplicity. In the next case, we just follow the procedure of the general approach proposed early. As a result, we obtain several approximations which can be easily calculated and are better than some other approximations. Numerical comparison shows that our approximations are very accurate.

**Keywords** Normal Distribution, Cumulative Distribution Function, Approximations, Homotopy Perturbation Method

## 1 Introduction

The importance of the normal distribution in many areas of science (for example, in mathematical statistics and statistical physics) follows from the central limit theorem of probability theory. If the observation result is the sum of many random weakly interdependent quantities, each of which makes a small contribution relative to the total sum, then with an increase in the number of terms, the distribution of the centered and normalized result tends to normal. This law of probability theory has the consequence of a wide distribution of the normal distribution, which was one of the reasons for its name. The normal distribution [1], also called the Gauss or Gauss Laplace distribution, is the probability distribution, which in the one-dimensional case is given by a probability density function that

coincides with the Gauss function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (1)$$

where the parameter  $\mu$  is the mean or mathematical expectation of the distribution, the median and the mode of distribution, and the parameter  $\sigma$  is the standard deviation ( $\sigma^2$  is the dispersion or the variance) of the distribution. Thus, the one-dimensional normal distribution is a two-parameter family of distributions. The standard normal distribution is called the normal distribution with the expectation  $\mu = 0$  and the standard deviation  $\sigma = 1$ .

If a random variable  $x$  is normally distributed with zero mean and the standard deviation, then its probability density function (1) is,

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

A wide usage of this distribution is due to the fact that it is an infinitely divisible continuous distribution with finite variance. Therefore, some others distributions, for example, the binomial and Poisson ones, approach it in the limit. Moreover, many non-deterministic physical processes are modeled by this distribution. The cumulative distribution function (CDF) of the standard normal distribution is the integral

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (3)$$

Unfortunately,  $\Phi(x)$  cannot be expressed in a closed form in terms of elementary functions for all values of  $x$  which would be useful for the practical needs, many numerical approximations for  $\Phi(x)$  are known. A number of approximate functions for CDF have been proposed (see, for example, [2, 3], and references therein). Let us represent CDF (3) in the following form

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \quad (4)$$

It can be seen that the problem of approximating  $\Phi(x)$  is equivalent to the problem of approximating the integral in (4) which

can be expressed by means of the error function  $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$  as

$$\Phi(x) = \frac{1}{2} \left[ 1 + erf\left(\frac{x}{\sqrt{2}}\right) \right]. \tag{5}$$

Therefore, the problem of approximating  $\Phi(x)$  is equivalent to the problem of approximating the error function. There are several methods (see, for example, [4]) which provide an approximation of the integral by different numerical methods: Taylor series, asymptotic series, continual fractions, and some others. For more, several other papers should be mentioned where different approximate formulae were obtained for  $\Phi(x)$ , such as [5]-[10].

It could be mentioned that in [4] the application of the homotopy perturbation method (HPM) to calculate an approximate analytical solution of the normal distribution integral was proposed. Besides, after solving the Gaussian integral by HPM, the result could serve as the base to solve other integrals like error function and the cumulative distribution function. The basic idea of [4] is that the integral similar to (4) can be reformulated as a differential equation subjected to a certain initial condition. Being applied to equation (4), this means that it is necessary to solve

$$\Phi'(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 0 \tag{6}$$

with the initial condition  $\Phi(0) = \frac{1}{2}$  that follows directly from (4). As noted in [4], the solution for (6) is similar, qualitatively, to a hyperbolic tangent because when  $x$  tends to  $\pm\infty$ , the derivative  $\Phi'(x)$  tends to zero, hence by symmetry,  $\Phi(x)$  tends to the same constant on both directions. It is why the authors of [4] made the conclusion that the first approach of the HPM method contains a hyperbolic term. Then they established a differential equation that may be solved using hyperbolic tangent. It seems that the cited article provides one of the possibilities of applying HPM in the problem under consideration. The author of this article has already applied this method to a similar problem earlier.

Recently, for the analytical calculation of the cosmological luminosity distance, we offered to proceed from the solution of differential equation with certain initial conditions instead of calculating the corresponding integral. For this purpose, we have obtained the differential equation which the luminosity distance should satisfy to, and define the appropriate initial conditions for this equation. We have showed that by using the HPM [11], the explicit dependency of luminosity distance on red-shift in arbitrary accuracy can be easily obtained by implementing a simple procedure for the governing equation.

In the present paper, we apply the He's HPM in order to obtain the approximations for the cumulative normal distribution function. For this purpose, we consider  $\Phi(x)$  as the unknown function to be determine from solving a certain differential equation. Indeed, taking into account equations (4) and (6), one can get the following Cauchy problem for  $\Phi(x)$ :

$$\Phi''(x) + x \Phi'(x) = 0; \quad \Phi|_{x=0} = \frac{1}{2}, \quad \Phi'|_{x=0} = \frac{1}{\sqrt{2\pi}}, \tag{7}$$

where the prime stands for the derivative with respect to  $x$ .

## 2 Approximation by HPM

The main equation (7) is a linear differential equation of the second order. It can be solved exactly in quadratures, but the result again leads to the formula (4). Therefore, we will solve this equation analytically, but with a certain approximation. Among all kinds of approximate methods we now use the HPM. In this method, it is not required to introduce a small parameter, because it is naturally contained in the method itself.

Since the HPM has now become standard and for brevity, the reader is referred to [12]-[15] for the basic ideas of HPM. In this section, we shall apply the HPM to solve equation (7). Let us assume that the solution of this equation can be represented by a series in  $p$  as follows

$$\Phi = \Phi_0 + p \Phi_1 + p^2 \Phi_2 + p^3 \Phi_3 + \dots \tag{8}$$

where  $p \in [0, 1]$  is an imbedding parameter. When we put  $p \rightarrow 1$ , then equation (3) corresponds to (2), and (5) becomes the approximate solution of (7), that is

$$\Phi(x) = \lim_{p \rightarrow 1} \Phi = \Phi_0 + \Phi_1 + \Phi_2 + \Phi_3 + \dots \tag{9}$$

It is useful to note that the result of solving an equation by this method and the convergence rate greatly depend on the choice of the homotopy. Therefore, we consider two cases in what follows. In one case, we construct the homotopy from the idea of simplicity. In the next case, we just follow the procedure of the general approach proposed in [12].

### 2.1 The case of trivial homotopy

From the purely pedagogical purposes, first we consider an extremely simple case of constructing homotopy in order to show once more that the simple is not always the best. Applying the HPM method to equation (7) in this case, we build the following simplest homotopy:

$$\Phi''(x) + p x \Phi'(x) = 0, \quad p \in [0, 1], \tag{10}$$

and assume that this equation can be solved by means of the series in  $p$  as (8).

Substituting (8) into equation (10), and equating coefficients of like powers of  $p$ , one obtains the following equations:

$$\begin{aligned} p^0 &: \Phi''_0 = 0, \\ p^1 &: \Phi''_1 + x \Phi'_0 = 0, \\ &\dots \dots \dots \\ p^n &: \Phi''_n + x \Phi'_{n-1} = 0, \\ &\dots \dots \dots \end{aligned} \tag{11}$$

According to (7), the initial conditions for  $\Phi_i(x)$  can be chosen as follows

$$\begin{aligned} \Phi_0|_{x=0} &= \frac{1}{2}, \quad \Phi'_0|_{x=0} = \frac{1}{\sqrt{2\pi}}; \\ \Phi_j|_{x=0} &= 0, \quad \Phi'_j|_{x=0} = 0; \end{aligned} \tag{12}$$

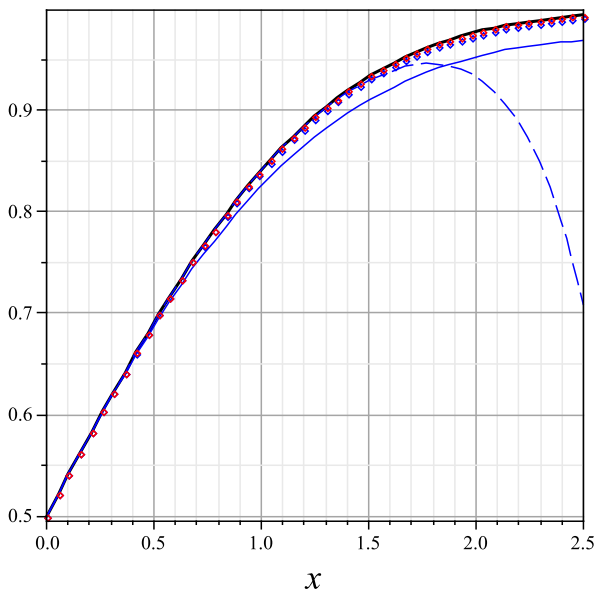


Figure 1: Comparison of the approximate solutions, given by (14) (blue dashed line), (23) (blue solid line), (24) (blue line of points) and (25) (red line), with the exact numerical solution to Eq. (4) (black line).

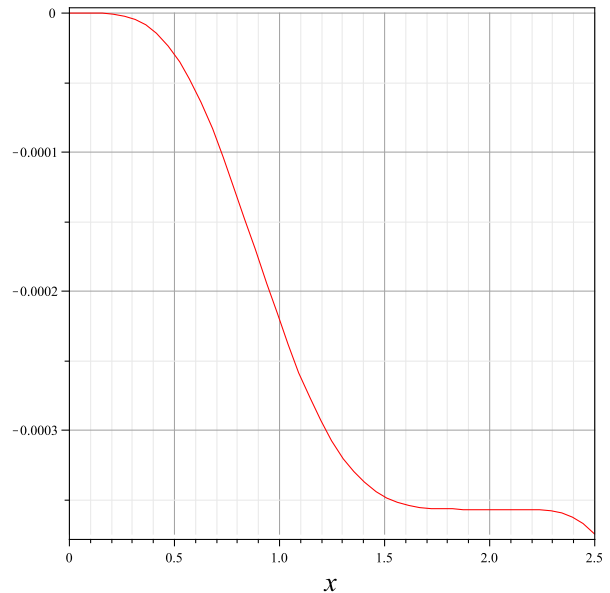


Figure 2: The absolute errors of the approximate solution given by equation (25), compared to the exact numerical solution to equation (4).

where  $j \geq 1$ . The exact solutions for the set of equations (11) with initial conditions (12) can be readily found as

$$\begin{aligned} \Phi_0(x) &= \frac{1}{2} + \frac{x}{\sqrt{2\pi}}, \\ \Phi_1(x) &= -\frac{1}{\sqrt{2\pi}} \frac{x^3}{2 \cdot 3}, \\ \Phi_2(x) &= \frac{1}{\sqrt{2\pi}} \frac{x^5}{2! \cdot 4 \cdot 5} \\ \Phi_3(x) &= -\frac{1}{\sqrt{2\pi}} \frac{x^7}{3! \cdot 8 \cdot 7} \dots \end{aligned} \tag{13}$$

Substituting all solutions (13) into equation (9), we obtain

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left( x - \frac{x^3}{2 \cdot 3} + \frac{x^5}{2! \cdot 4 \cdot 5} - \frac{x^7}{3! \cdot 8 \cdot 7} + \dots \right) \tag{14}$$

The convergence of this solution is rather obvious, because formula (14) could be obtained merely from the decomposition

$$e^{-\frac{t^2}{2}} = 1 - \frac{t^2}{2} + \frac{t^4}{4 \cdot 2!} - \frac{t^6}{8 \cdot 3!} + \dots$$

in equation (4). Thus, such a simple homotopy yielded a rather expected result just coinciding with the approximation obtained by decomposition of the exponent into a Taylor series. Acceptable accuracy of such an approximation can be achieved only by keeping a sufficiently large number of terms in this decomposition that is not always convenient. Obviously, we must somehow improve the homotopy in order to obtain a more accurate approximation for CDF.

## 2.2 The case of improved homotopy

In this case, we build the homotopy according to the general procedure of the method, namely

$$\Phi''(x) + \Phi'(x) + p(x-1)\Phi'(x) = 0, \tag{15}$$

where  $p \in [0, 1]$ . Then, substituting (8) into equation (15) and equating coefficients of like powers of  $p$ , we obtain the following set of the linear differential equations:

$$p^0 : \Phi_0'' + \Phi_0' = 0, \tag{16}$$

$$p^1 : \Phi_1'' + \Phi_1' + (x-1)\Phi_0' = 0, \tag{17}$$

.....

$$p^n : \Phi_n'' + \Phi_n' + (x-1)\Phi_{n-1}' = 0, \tag{18}$$

.....

Obviously, all successive approximations can be obtained rather easily. Solving equation (16) with the initial conditions (12), one can get

$$\Phi_0(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} (1 - e^{-x}). \tag{19}$$

Substituting this function into equation (17), and taking into account (12), we obtain after integration that

$$\Phi_1(x) = \frac{x^2}{2\sqrt{2\pi}} e^{-x}. \tag{20}$$

Then we can solve equation (18) for  $n = 2$  and  $n = 3$  with initial conditions (12). The result is as follows

$$\Phi_2(x) = \frac{1}{\sqrt{2\pi}} - \frac{1}{2\sqrt{2\pi}} \left( 2 + 2x + x^2 + \frac{x^4}{4} \right) e^{-x}, \tag{21}$$

and

$$\Phi_3(x) = \frac{1}{\sqrt{2\pi}} \left[ \left( 5 + 5x + \frac{5x^2}{2} + \frac{5x^3}{6} + \frac{x^4}{4} + \frac{x^6}{48} \right) e^{-x} - 5 \right]. \tag{22}$$

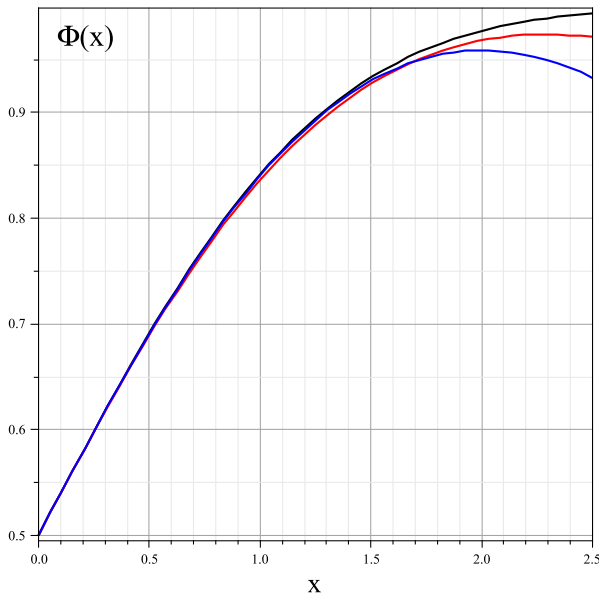


Figure 3: Comparison of the approximate solutions given by formula (32) for  $\lambda = 0.5$  (in blue) and  $\lambda = 1/\sqrt{2}$  (in red) with the exact numerical solution to Eq. (4) (in black).

According to equation (9) and the accuracy achieved in the case  $\Phi(x) \approx \Phi_0(x) + \Phi_1(x)$ , we obtain the following approximate solution

$$\Phi(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left[ 1 - \left( 1 - \frac{x^2}{2} \right) e^{-x} \right], \tag{23}$$

by summing (19) and (20). Taking into account equations (19)-(21), we can obtain the next approximation  $\Phi(x) \approx \Phi_0(x) + \Phi_1(x) + \Phi_2(x)$  as follows

$$\Phi(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left[ 2 - \left( 2 + x + \frac{x^4}{8} \right) e^{-x} \right]. \tag{24}$$

In the framework of the present study, we constrain ourselves to the next approximation, namely,  $\Phi(x) \approx \Phi_0(x) + \Phi_1(x) + \Phi_2(x) + \Phi_3(x)$ . Due to equations (22) and (24), this approximation is as follows

$$\Phi(x) \approx \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \left[ 3 - \left( 3 + 4x + \frac{5x^2}{2} + \frac{5x^3}{6} + \frac{x^4}{8} + \frac{x^6}{48} \right) e^{-x} \right]. \tag{25}$$

In order to demonstrate the accuracy of the method applied, the graphs of the approximate solutions  $\Phi(x)$  according (16) for different values of  $\lambda$  and the numerical solution to equation (4) via the Maple package are given in Fig. 1. The absolute error of the approximate solution (25), compared to the exact numerical solution to equation (4), is plotted in Fig. 2, and demonstrates a rather high level of accuracy for this approximation.

### 2.3 A one-parameter tuning of improved homotopy

In this case, we are going to introduce a real tuning parameter  $\lambda$  into the homotopy (15) considered in the previous section, for example, as follows

$$\Phi''(x) + \lambda \Phi'(x) + p(x - \lambda) \Phi'(x) = 0. \tag{26}$$

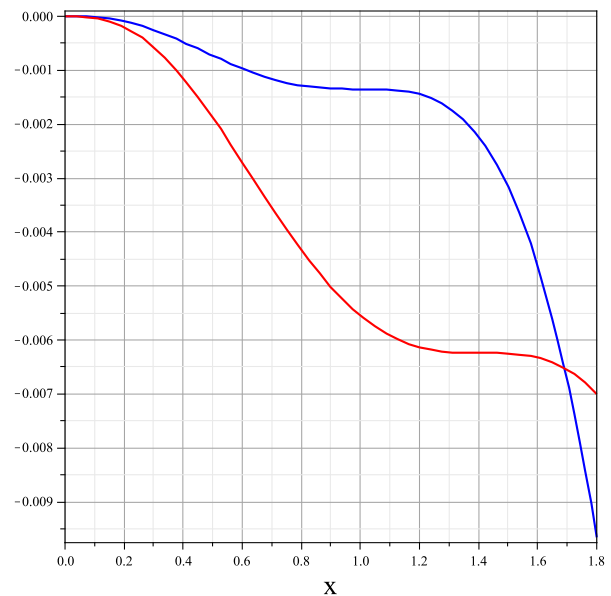


Figure 4: The absolute errors of the approximate solutions, given by equation (33) (in blue) and equation (34) (in red), compared to the exact numerical solution to equation (4).

As a result, the system of equations (16)-(18) transforms into the following set of linear equations

$$p^0 : \Phi_0'' + \lambda \Phi_0' = 0, \tag{27}$$

$$p^1 : \Phi_1'' + \lambda \Phi_1' + (x - \lambda) \Phi_0' = 0, \tag{28}$$

.....

$$p^n : \Phi_n'' + \lambda \Phi_n' + (x - \lambda) \Phi_{n-1}' = 0, \tag{29}$$

.....

Solving equation (27) with the same initial conditions (12), we obtain

$$\Phi_0(x, \lambda) = \frac{1}{2} + \frac{1}{\lambda\sqrt{2\pi}} (1 - e^{-\lambda x}). \tag{30}$$

Using this function in equation (28), and again taking into account (12), we can obtain that

$$\Phi_1(x, \lambda) = \frac{1}{\lambda\sqrt{2\pi}} \left[ \frac{\lambda^2 - 1}{\lambda^2} (1 - (1 + \lambda x)e^{-\lambda x}) + \frac{x^2}{2} e^{-\lambda x} \right]. \tag{31}$$

Taking into account equations (30) and (31), we can obtain the simplest one-step approximate solution  $\Phi(x, \lambda) \approx \Phi_0(x, \lambda) + \Phi_1(x, \lambda)$  as follows

$$\Phi(x, \lambda) \approx \frac{1}{2} + \frac{1}{\lambda\sqrt{2\pi}} \left[ \frac{2\lambda^2 - 1}{\lambda^2} + \left( \frac{x^2}{2} - \frac{\lambda^2 - 1}{\lambda} x - \frac{2\lambda^2 - 1}{\lambda^2} \right) e^{-\lambda x} \right]. \tag{32}$$

A single parameter that can be adjusted in (32) (for example, by using the NonlinearFit command from Maple Release 15) in order to obtain a good approximation is  $\lambda$ . As a consequence, this adjustment allows us to ignore a large number of successive terms in a good approximation. The graphs of the approximate solutions for  $\Phi(x)$  given by equation (32) for  $\lambda = 0.5$  and  $\lambda = 1/\sqrt{2}$  compared to the numerical solution to equation (4) are plotted in Fig.3.

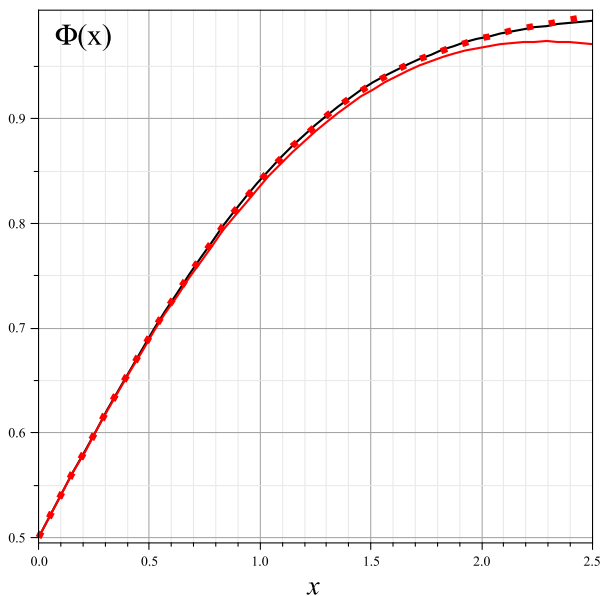


Figure 5: Comparison of the approximate solutions given by equations (34) (red solid line) and (35) (red line of points) with the exact numerical solution to Eq. (4) (black line).

We present expression (32) in these two cases as the illustrative examples, when the formulae take the most simple form. Indeed, from (32) we have

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left[ (x^2 + 3x + 4) e^{-\frac{x}{2}} - 4 \right], \quad (33)$$

when  $\lambda = 0.5$ , and

$$\Phi(x) = \frac{1}{2} + \frac{1}{2\sqrt{\pi}} (x^2 + \sqrt{2}x) e^{-\frac{x}{\sqrt{2}}}, \quad (34)$$

when  $\lambda = 1/\sqrt{2}$ . The absolute errors of these approximations compared to the numerical solution of equation (4) are plotted in Fig. 4.

It should be noted that approximation (32), and, consequently, particular formulas (33) and (34), were obtained in a single iteration. It is clear that subsequent iterations will significantly improve the accuracy of the approximation, while at the same time complicating the expression for CDF. In view of the complexity of the general expression for the second iteration, we give here the following approximation for the case of the  $\lambda = 1/\sqrt{2}$ , which demonstrates the best result with relative simplicity of the expression in the above result. One can easily solve the linear equation (29) for  $n = 2$  taking into account

$\Phi_1(x, 1/\sqrt{2})$  for  $\lambda = 1/\sqrt{2}$  from (31), and then obtain the approximate solution  $\Phi(x) \approx \Phi_0(x) + \Phi_1(x) + \Phi_2(x)$  according to (34) as follows

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \left[ 7 - \left( 7 + 3\sqrt{2}x + \frac{5}{4}x^2 + \frac{\sqrt{2}}{4}x^3 + \frac{x^4}{8} \right) e^{-\frac{x}{\sqrt{2}}} \right]. \quad (35)$$

The graphs of the approximate solutions for  $\Phi(x)$  represented by equations (34) and (35) compared to the numerical solution to equation (4) are plotted in Fig. 5. The absolute error of the

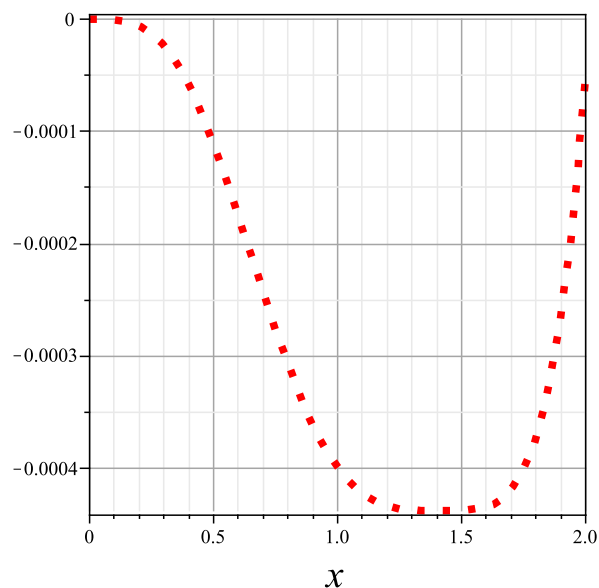


Figure 6: The absolute error of approximate solution, given by equation (35), compared to the exact numerical solution to equation (4).

best approximation (35) compared to the numerical solution of equation (4) is plotted in Fig. 6. It can be seen from this figure that formula (35) provides a rather accurate result along with relative simplicity of its expression.

### 3 Conclusions

In this paper, we presented some approximate analytic solutions for the cumulative normal distribution obtained by using HPM. Formulae (25) and (35) provide the best approximations with low order relative errors. For example, one of the best approximations given by formula (35) for  $0 < x < 2$  has an absolute error of less than  $4.5 \times 10^{-4}$ . Our approximations to CDF are comparable to other approximations found in the literature and has the advantage of being the simple expressions, that may have potential applications in several areas of applied sciences. It is worth to note that the approximate CDFs obtained in the present paper can be employed for the solution of different engineering and scientific problems with rather good approximation.

## REFERENCES

- [1] N. Johnson, S. Kotz, and N. Balakrishnan, Continuous Univariate Distributions, Vol. 1, *John Wiley & Sons*, New York, 1994.
- [2] M. Edous, O. Eidous. A Simple Approximation for Normal Distribution Function, *Mathematics and Statistics*, 6(4): 47-49, 2018.
- [3] O. M. Eidous, and S. A. Al-Salman. One-Term Approximation for Normal Distribution Function, *Mathematics and Statistics*, 4(1): 15-18, 2016.
- [4] H. Vazquez-Leal, R. Castaneda-Sheissa, et al. High Accurate Simple Approximation of Normal Distribution Integral, *Mathematical Problems in Engineering*, Vol. 2012, Article ID 124029, 2012. doi:10.1155/2012/124029.
- [5] K. M. Aludaat, and M. T. Alodat. A Note on Approximating the normal distribution function, *Applied Mathematical science*, 2(9): 425-429, 2008.
- [6] B. J. Bailey. Alternatives to hasting's approximation to the inverse of the normal cumulative distribution function, *Applied statistics*, 30(3): 275-276, 1981.
- [7] G. Polya. Remarks on computing the probability integral in one and two dimensions, Proceeding of the first Berkeley symposium on mathematical statistics and probability, 63-78, 1945.
- [8] S. R. Bowling, M. T. Khasawneh, S. Kaewkuekool, and B. R. Cho. A Logistic approximation to the cumulative normal distribution, *Journal of Industrial Engineering and Management*, 2(1): 114-127, 2009.
- [9] J. T. Lin. A simple Logistic approximation to the normal tail probability and its inverse, *Applied Statistics*, 39: 255-257, 1990.
- [10] R. Yerukala, and N.K. Boiroju. Approximations to Standard Normal Distribution Function, *International Journal of Scientific & Engineering Research*, 6(4): 515-518, 2015.
- [11] V. K. Shchigolev. Calculating Luminosity Distance versus Redshift in FLRW Cosmology via Homotopy Perturbation Method, *Gravitation and Cosmology*, 23(2): 142-148, 2017.
- [12] J.-H. He. Homotopy perturbation technique, *Comput. Meth. Appl. Mech. Eng.*, Vol. 178, 257-62, 1999.
- [13] J.-H. He. A coupling method of homotopy technique and perturbation technique for nonlinear problems, *Int. J. Nonlinear Mech.*, 35(1): 37-43, 2000.
- [14] J.-H. He. Homotopy perturbation method: a new nonlinear analytical technique, *Appl. Math. Comput.*, 135: 73-79, 2003.
- [15] L. Cveticanin. Homotopy-perturbation method for pure nonlinear differential equation, *Chaos, Solitons & Fractals*, 30(5): 1221-1230, 2006.