Approximations of the Standard Normal Distribution

B. Zogheib[1] and M. Hlynka[2]

[1]*Division of Mathematics, Science and Technology, Nova Southeastern University*
3301 College Ave., Fort Lauderdale-Davie, FL, 33314-7796, USA

[2]*Department of Mathematics and Statistics, University of Windsor*
401 Sunset Ave., Windsor, ON, N9B 3P4, Canada

Email: zogheib@nova.edu, hlynka@uwindsor.ca

Abstract: This paper presents five new formulas for approximation of cumulative standard normal probabilities. Two of these approximations are polynomial based and are only accurate for $0 \le z \le 1$; the other three formulas are accurate on the interval $-3.4 \le z \le 3.4$ which is the domain often used in normal tables. We recommend the last of these new formulas.

1. Introduction

Much research has been done to find approximations for the area under the standard normal curve. 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.

In this paper, the goal is to develop simple functions that give good approximations of the cumulative standard normal probabilities.

Good approximations should have various properties. They should

(a) be simple in form

(b) have simple coefficients

(c) be invertible

(d) have very good accuracy

(e) not require different forms over different intervals

(f) have a logical asymptotic form.

Accuracy can be measured by the maximum error or by the maximum relative error or by some other performance measure.

The standard normal probability density function (pdf) and the cumulative distribution function (cdf) are $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$, respectivly. Let $\overline{\Phi}(x) = 1 - \Phi(x)$.

2. Overview of the Approximations of the Normal Standard Cumulative Function

Work found in the literature to approximate the area under the normal curve is divided into two groups. One group offers simple formulas with reasonable accuracy, the other offers less simple formulas but high accuracy. Discussion of some methods appears in Johnson, Kotz, and Balakrishnan [1]. Shah[2] found a very simple formula, but it does not approximate the area well in many regions.

$$\Phi_S(z) - .5 = \begin{cases} z(4.4 - z)/10, & 0 \le z \le 2.2, \\ 0.49, & 2.2 \le z \le 2.6, \\ 0.5, & z \ge 2.6. \end{cases}$$

Norton [3] obtained two approximations and his results were more accurate than Shah's[2], but also more complicated.

$$\overline{\Phi}_{N1}(t) = \frac{1}{2}\exp\left[-\left(t^2 + t\right)/2\right], \qquad 0 \le t \le 2.6$$

$$\overline{\Phi}_{N2}(t) = \frac{1}{2}\exp\left[-\left(t^2 + 1.2t^8\right)/2\right], \quad 0 \le t \le 2.7.$$

Norton uses $\overline{\Phi}_N(t) = \phi(t)/t$ for $t > 2.6$ or $t > 2.7$ respectively, $\phi(t)$ is the standard normal p.d.f.

Bryc[4] found two nice approximations with an error less than 0.5% . One of these formulas was found to offer a more accurate alternative to a formula that was found by Hart[5, 6]. Bryc's formula

$$\overline{\Phi}_B(z) = \frac{z + 3.333}{\sqrt{2\pi}z^2 + 7.32z + 2 \times 3.333} e^{-z^2/2}$$

has a largest error 0f 0.00071 occurs in the range $1.07 \le z \le 1.13$. Hoyt[7], Cody[8], and others found some additional approximations.

Waissi and Rossin[9] found an approximation that has the form of sigmoid function. The approximation function for, $-8 \le z \le +8$, has the form

$$\Phi_W(z) = \frac{1}{1 + e^{-\sqrt{\pi}\left(\beta_1 z^5 + \beta_2 z^3 + \beta_3 z\right)}}$$

where $\beta_1 = -0.0004406$, $\beta_2 = 0.0418198$, $\beta_3 = 0.9$. The error of this approximation varies between ± 0.0000431 for $-8 \le z \le +8$.

3. Five New Approximations

The cumulative distribution function of a standard normal distribution, given by

$$\Phi(z) = \frac{1}{2} + \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \text{ for } z > 0,$$

can be approximated by a polynomial over a specific domain such as [0,3], which is the domain often used in normal tables. One way to get such a polynomial would be to use a Taylor's series expansion of $e^{-\frac{t^2}{2}}$. So for $z \ge 0$, the cumulative distribution function is given by

$$\Phi(z) = 0.5 - 0.398942z - 0.066490z^3 + 0.09974z^5 - 0.01187z^7 + \ldots .$$

One approximation would be

$$\Phi_1(z) = 0.5 - 0.398942z - 0.066490z^3 + 0.09974z^5 .$$

Figure 1 illustrates the difference between $\Phi_1(z)$ and $\Phi(z)$. The error of approximation is very small for $0 \le z \le 1$. Another polynomial can be obtained by using regression and regressing $\Phi(z)$ on $z,\ z^3,\ z^5$ where

$$z = 0,\ 0.1,\ 0.2,\ \ldots,\ 3.0$$

with the values of $\Phi(z)$ taken from a normal table. This approximation is given by

$$\Phi_2(z) = 0.50 - .368929z - 0.037758z^3 + 0.01645z^5 .$$

Figure 2 illustrates the difference between $\Phi_2(z)$ and $\Phi(z)$. Again this polynomial approximation is very accurate for the region $0 \le z \le 1.2$. The reason why these approximations are not good for large z is that for any polynomial $f(z)$, $\lim_{z\to\infty} f(z)$ is $\pm\infty$, whereas $\lim_{z\to\infty} \Phi(z) = 1$.

A very good approximation, based on the structure found by Waissi [8], but with different values, is as follows.

The model is

$$\Phi(z) = \left(1 + e^{p(z)}\right)^{-1}$$

where

$$p(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 .$$

After using regression analysis with

$$z = -3.4,\ -3.3,\ \ldots 0,\ 0.1,\ 0.2,\ \ldots,\ 3.3,\ 3.4$$

our new approximation will be given by

$$\Phi_3(z) = \left(1 + e^{0.000345\ z - 0.039547\ z^3 - 1.604326\ z^5}\right)^{-1} .$$

Figure 3 shows the difference between $\Phi_3(z)$ and $\Phi(z)$.

Another approximation can be obtained by using same formula, but regressing on $z,\ z^3$ only. The formula obtained is

$$\Phi_4(z) = \left(1 + e^{0.0054\ -1.6101\ z - 0.0674\ z^3}\right)^{-1}$$

Figure 4 shows the difference between $\Phi_4(z)$ and $\Phi(z)$. This approximation has a simple form yet is very accurate.

A function of the form

$$\Phi(z) = 1 - 0.5 e^{-Az^b}$$

can be used as an approximation to the standard normal cumulative function. By using regression analysis and after rounding the coefficient to one decimal place, the approximation obtained is

$$\Phi_5(z) = 1 - 0.5 e^{-1.2 z^{1.3}} .$$

Figure 5 shows the difference between $\Phi_5(z)$ and $\Phi(z)$. Figure 6 shows $\Phi_5(z)$ and $\Phi(z)$ overlaid, for $z > 0$. This last approximation has a simpler form than $\Phi_4(z)$ and $\Phi_3(z)$, but is not as accurate.

4. Conclusion

In this paper, some approximations to the standard normal cumulative distribution function are found. Some of these approximations have simple form but do not achieve accuracy, others are more complicated in form but achieve accuracy. Two formulas which are simple in form and accurate are found. The final formula $\Phi_5(z)$ satisfies ALL of the desirable properties given in the introduction and we recommend it.

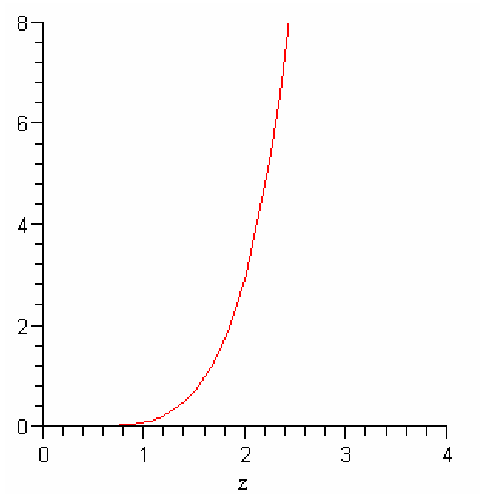

Figure 1

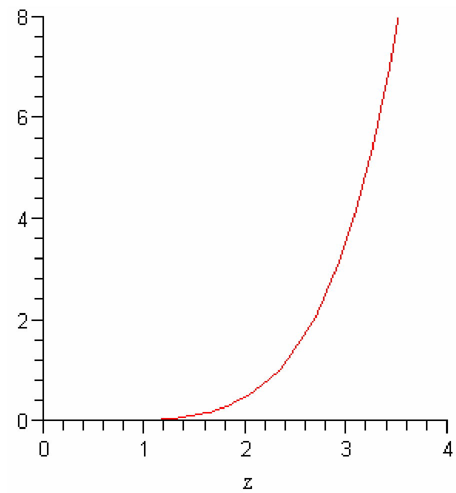

Figure 2

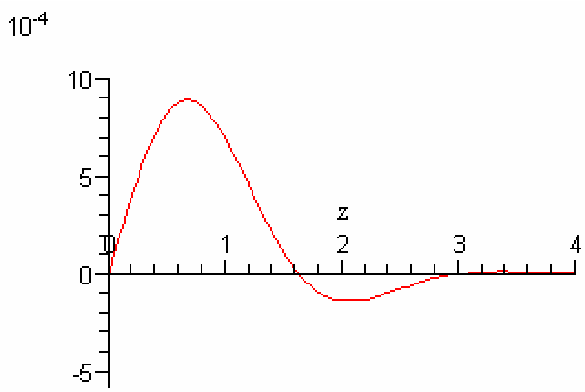

Figure 3

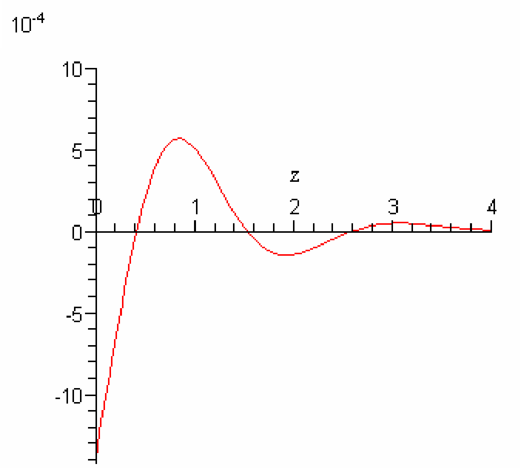

Figure 4

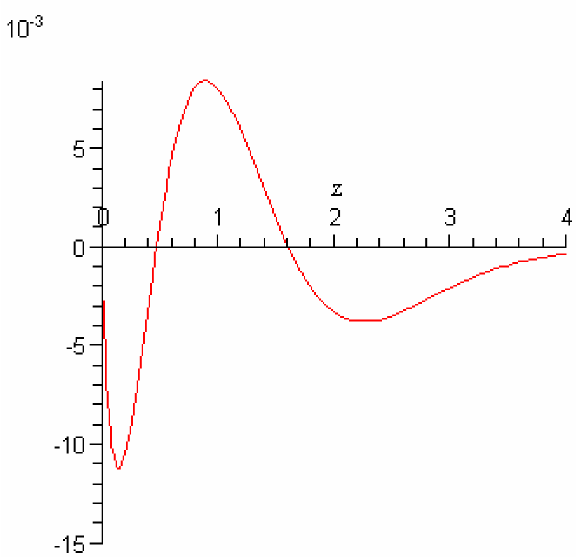

Figure 5

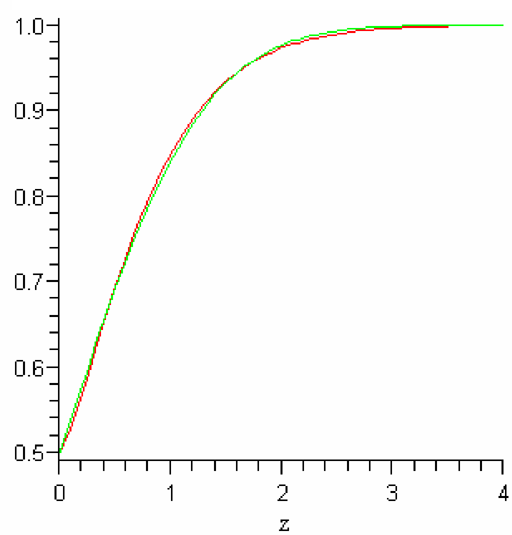

Figure 6

References

[1] Johnson, N., Kotz, S. and Balakrishnan, N., 1994, 1995. Continuous Univariate Distributions, Vols. 1 and 2, John Wiley, New York,.

[2] Shah, A. K., 1985. "A Simpler Approximation for Area Under the Standard Normal Curve," The American Statistician, Vol. 39, No. 1, 80-80.

[3] Norton, R. M., 1989. "Pocket Calculator approximation for areas under the Standard Normal Curve," The American Statistician, Vol. 43, No. 1, 24-26.

[4] Bryc, W., 2002. "A Uniform Approximation to the Right Normal Tail Integral," Applied Mathematics and Computation, Vol. 127, 365-374.

[5] Hart, R. G., 1957. "A Formula for the Approximation of the Normal Distribution Function," Mathematical Tables and Other Aids to Computation, Vol. 11, 60.

[6] Hart, R. G., 1966. "A Close Approximation Related to the Error Function," Mathematics of Computation, Vol. 20, 96.

[7] Hoyt, J., 1968. "A Simple Approximation to the Standard Normal Probability Density Function, American Statistician, Vol. 22, No. 2.

[8] Cody, W., 1969. "Rational Chebyshev Approximations for the Error Function ," Mathematics of Computation, Vol. 23, 631-638.

[9] Waissi, G. R., Rossin, D. F., 1996. "A Sigmoid Approximation of the Standard Normal Integral", Applied Mathematics and Computation, Vol. 77, 91-95.