On a Logistic Approximation to the Normal and Other Cumulative Distribution Functions

[1]Roberto N. Padua, [2]Mark S. Borres and [3]Efren O. Barabat

Abstract

This paper proposes an approach that maintains the simplicity of various approximations as well as having a closed- form algebraic inverse yet easily generalizable to the approximation of other cumulative distribution functions. The proposed logistic approximation to the normal cumulative distribution function has a maximum error of 0.000560, lower than the maximum error reported by Aludaat et al (2007). Furthermore, the logistic approximation is more flexible in that it can be used to approximate the cumulative distribution functions of other distribution. A Chi-square approximation was also investigated and reported a maximum error of 0.00494.

Keywords: logistic approximation, cumulative distribution functions, normal distribution, chi-square distribution, estimation of parameters

1.0 Introduction

The cumulative distribution function of the standard normal distribution is given by:

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

and is not expressible in a closed-form expression. Consequently, several attempts have been made to approximate (1) in simple forms (Johnson et al. (1994)). A one-term- to calculate approximation was proposed by Polya (1945), namely:

$$(2)\quad \Phi(x) \simeq 0.5\left(1 + \sqrt{1 - e^{\frac{-2}{\pi}x^2}}\right), \quad x \geq 0$$

which was subsequently improved by Aludaat et al. (2007) to:

$$(3)\quad \Phi(x) \simeq 0.5\left(1 + \sqrt{1 - e^{-\sqrt{\frac{\pi}{8}}x^2}}\right), \quad x \geq 0$$

The latter approximation was shown to have a maximum error of 0.00197323 and had the added advantage of having a closed-form algebraic inverse.

Approximations (2) and (3), while indeed simple, are not intuitively appealing and, thus, not easily generalizable to the approximation of other cumulative distribution functions e.g. Gamma distribution. We propose an approach that maintains the simplicity of (2) and (3) as well as having a closed- form algebraic inverse yet easily generalizable to the approximation of other cumulative distribution functions.

2.0 The Logistic Model

The logistic function given by:

$$(4)\quad S(x) = \frac{e^{h(x)}}{1+e^{h(x)}}, \quad -\infty < x < \infty$$

is an S-shaped curve and can be used to fit the shape of a cumulative distribution function depending on the choice of the function h(x). We require that:

$$(5)\quad h(x) \to 0 \quad \text{as } x \to -\infty$$

$$h(x) \to \infty \quad \text{as } x \to \infty$$

h(x) is an increasing function of x.

Theorem 1. Under assumption (5), S(x) is a cumulative distribution function.

[1]Consultant
[2,3] University of San Jose- Recoletos

Proof: We need to check the following conditions:

(a) $\lim_{x\to\infty} S(x) = 1$, $\lim_{x\to-\infty} S(x) = 0$

(b) $S(x)$ is a monotonically increasing function.

Note that:

$\lim_{x\to\infty} S(x) = \lim_{x\to\infty}\left(\frac{1}{1+e^{-h(x)}}\right)$=1 by (5) and also,

$\lim_{x\to\infty} S(x) = \lim_{x\to\infty}\left(\frac{e^{h(x)}}{1+e^{h(x)}}\right)$=0

Likewise, the first derivative of S(x) is:

$S'(x) = \frac{h'(x)}{(1+e^{-h(x)})^2}$ >0 since $h'(x) > 0$ *by* (5).

The simplest choices of h(x) are:

$$h(x) = a + bx, \qquad b > 0$$

and the more general polynomial

$$h(x) = a + bx + cx^2 + dx^3 + \cdots + ex^n$$

Estimation of Parameters

From (4), we deduce that:

(6) $h(x) = \ln\frac{S(x)}{1-S(x)} = y$

In particular, if then we obtain the linear relation:

(7) $y = a + bx$

Let $x_1, x_2, x_3, \ldots, x_i$ be iid N(0,1), we find estimates of the parameters by fitting

$S_n(x)$ to $\Phi(x_i)$ for large n.

3.0 Maximum Error and Inverse

We generated n=10, 000 standard normal random observations and obtained:

(8) $h(x)=-0.000675+1.60513x-0.0000570x^2+0.0671498x^3$

with R^2= 100 %

Neglecting the constant and quadratic terms, we find:

(9) $h(x) = 1.60513x + 0.0671498x^3$

Which is very similar to Aludaat's (2007)

(10) $\Phi_3(x) \simeq \frac{\exp(2y)}{1+\exp(2y)}$, $y = 0.7988x(1 + 0.04417x^2)$

The approximation is given by:

(11) $S(x) = \frac{\exp(-0.000675+1.60513x-0.0000570x^2+0.0671498x^3}{1+\exp(-0.000675+1.60513x-0.0000570x^2+0.0671498x^3}$

yielded a maximum error of 0.000560, compared to Aludaat et al. (2007) which was 0.00197323.

By Cardan's (1545) formula, the inverse is given by:

$$\left(q + \sqrt{q^2 + (r-p^2)^3}\right)^{\frac{1}{3}} + \left(q - \sqrt{q^2 + (r-p^2)^3}\right)^{\frac{1}{3}} + p$$

where

$$p = \frac{-b}{3a}, q = p^3 + \frac{bc - 3ad}{6a^2}, r = \frac{c}{3a}$$

and

a=0.0671498 b=-0.0000570

c=1.6015 d=0.0000675- $\ln\frac{\Phi}{1-\Phi}$

4.0 Chi-square Distribution

We tried to find a corresponding logistic approximation to the Chi-square distribution. The cumulative distribution function of the Chi-square distribution with k=5 degrees of freedom, is given by:

(12) $$F_5(x) = \frac{1}{\sqrt{2^5}\Gamma(\frac{5}{2})}\int_0^x t^{\frac{3}{2}} e^{-\frac{t}{2}}\,dt$$

We generated n=10, 000 chi-square random observations with k = 5and obtained:

(13) $h(x) = 0.0215490 + 0.850538x - 0.0281537x^2 + 0.0007539x^3$

where the maximum error was computed to be 0.00494.

5.0 Conclusion

The proposed logistic approximation to the normal cumulative distribution function has a maximum error of 0.000560, lower than the maximum error reported by Aludaat et al (2007). Furthermore, the logistic approximation is more flexible since it can be used to approximate the cumulative distribution functions of other distribution. A Chi-square approximation was also investigated and reported a maximum error of 0.00494.The inverse can also be calculated using Cardana's (1545) cubic formula.

References

Van der Vaart, A. W. (1998), Asymptotic Statistics. Cambridge Series in Probabilistic Mathematics.

Aludaat, K. M & Alodat, M. T (2008). A Note on Approximating the Normal Distribution Function. Applied Mathematical Sciences, Vol.2, 2008.

Cook, R. D. & Weisberg, S. (1999). Applied Regression Including Computing and Graphics, New York: Wiley.

Cardano, G. (1663). Opera Omnia Hieronymi Cardani, Mediolanensis: Ars Magna (The Great Art), Lyons, Europe.

Craig, W. & Hogg (2000), An Introduction to Mathematical Statistics, (Wiley and Sons, New York)

Dudley, R. M. (1999). "Uniform Central Limit Theorems", Cambridge University Press. ISBN 0 521 46102.

Durrett, R. (1991). Probability: Theory and Examples. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Esseen, C. G. (1956). "A moment inequality with an application to the central limit theorem". Skand. Aktuarietidskr. 39: 160–170.

Feller, W. (1972). An Introduction to Probability Theory and Its Applications, Volume II (2nd ed.). New York: John Wiley & Sons.

Graybill, J. (1987) An Introductory Course in Mathematical Statistics (Wiley Series, New York)

Huber, P. (1985). Projection pursuit. The annals of Statistics, 13(2):435 – 475.

Johnson, R and Wichern (2000) Applied Multivariate Statistical Analysis (Wiley and Sons, New York)

Manoukian, E. B. (1986). Modern Concepts and Theorems of Mathematical Statistics. New York: Springer-Verlag.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. New York: John Wiley & Sons.

Shevtsova, I. G. (2007). "Sharpening of the upper bound of the absolute constant in the Berry–Esseen inequality". Theory of Probability and its Applications 51 (3): 549–553.

Shevtsova, I. G. (2008). "On the absolute constant in the Berry-Esseen inequality". The Collection of Papers of Young Scientists of the Faculty of Computational Mathematics and CyberneticsTheory of Probability and its Applications (5): 101-110.

Shiganov, I.S. (1986). "Refinement of the upper bound of a constant in the remainder term of the central limit theorem". Journal of Soviet mathematics 35: 109–115.

Shorack, G.R., Wellner J.A. (1986) Empirical Processes with Applications to Statistics, Wiley.

Tyurin, I.S. (2009). "On the accuracy of the Gaussian approximation". Doklady Mathematics 80 (3): 840-843.

Vapnik, V.N. and Chervonenkis, A. Ya (1971). On uniform convergence of the frequencies of events to their probabilities. Theor. Prob. Appl. 16, 264-280

Yeo, In-Kwon and Johnson, Richard (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87, 954-959.