A Logistic Model for the Approximation of The Gaussian Probability Integral Modified by A Skewed Sinusoid

by
James R Warren BSc MSc PhD

The Logistic Model

Let:-

$$p = \frac{1}{1 + e^{a-bz}} \qquad \textbf{Eqn.1}$$

be a Logistic Function which approximates the Integral of The Gaussian Distribution Function:-

$$\Phi(z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}.dt \qquad \textbf{Eqn.2}$$

Solution of Eqn.1 for the Linear Exponent x = a-bz yields:-

$$1 + e^{a-bz} = \frac{1}{p}$$

$$\therefore e^{a-bz} = \frac{1}{p} - 1 \qquad \textbf{Eqn.3}$$

$$\therefore x = a - bz = \log_n(\frac{1}{p} - 1)$$

The solid line in Figure 1 "The Plot of Linear Exponent x = a - bz" depicts the behavior of x on the interval 4≤z≤+4.

It is manifest that the Function x itself approximates a straight-line function modified by a shallow-amplitude slightly-distorted monomodal sine curve. But this implies that b is itself an inconstant function of z. Accordingly we may write that:-

$$x_z = l_z + s_z \qquad \textbf{Eqn.4}$$

or in words that the value of Function x for any Standard Score z is approximated by the sum of Linear and Sinusoidal Components, and that approximation is x_z.

The Analysis of the Linear Exponent Function, x

The Linear Component, l_z

The Intercept of line l is given by:-

$$l_0 = \log_n(\frac{1}{p_0} - 1) \qquad \textbf{Eqn.5}$$

whilst the Grade is given by:-

$$g_i = -\frac{l_0}{4} = -0.25\,l_0 \qquad \textbf{Eqn.6}$$

Accordingly:-

$$l_z = l_0 - 0.25\,l_0(z+4)$$
$$\therefore l_z = l_0 - 0.25\,l_0\,z - l_0 \qquad \textbf{Eqn.7}$$

$$\therefore l_z = -0.25\,l_0\,z \qquad \textbf{Eqn.8}$$

The Sinusoidal Component, s_z

The greatest depth of the superimposed skewed sinusoidal trace may be defined to occur at z_m, the Standard Score for which The Sinusoidal Component is a maximum. Hence z_m marks the amplitude of the skewed sinusoid normal to the abscissa.

Computer tests infer that the Sinusoidal Component, s_z, is quite accurately modelled by:-

$$s_z = d\,\text{Sin}[0.25\pi(k_1 z + k_3 z^3)] \qquad \textbf{Eqn.9}$$

normal to the abscissa along which z plots.

The Line Dip Component, ψ

The Line Dip Component is the angle Ψ illustrated in Figure 1. It may be quantified using:-

$$\Psi = \mathrm{Tan}^{-1}\left[\frac{4}{\log_n(\frac{1}{p_0}-1)}\right] = \mathrm{Tan}^{-1}\left[\frac{4}{l_0}\right] \qquad \textbf{Eqn.10}$$

Composition of the Function, x

By combining Eqn.8 and 9 in Eqn.4 we obtain:-

$$x_z = -0.25\, l_0\, z + d.\mathrm{Sin}\left[0.25\pi(k_1 z + k_3 z^3)\right] \qquad \textbf{Eqn.11}$$

<u>The Numerical Form of the Linear Exponent</u>

The Determination of the Line Dip Component, Ψ

Since l_0 = 10.36006962 we may write:-

$$\Psi = \mathrm{Tan}^{-1}(4/l_0) = 0.368464567$$

The Determination of the Ordinal Amplitude, d

It is apparent that the maximum of the sinusoid normal to the linear component lies somewhere in the range $2.0 \leq z \leq 2.5$ and is labelled by z_m. The Sinusoidal Component, s, has in the region of interest the following correspondences with z:-

z	s
2.0	1.4198
2.25	1.4355
2.5	1.3996

Fitment of a three-point Lagrangian Interpolation to this data yields the following polynomial equation:-

$$s = -0.4128\, z^2 + 1.8172z - 0.5634 \qquad \textbf{Eqn.12}$$

Differentiation of this expression gives:-

$$\frac{ds}{dz} = -0.8256z + 1.8172 = 0 \qquad \textbf{Eqn.13}$$

at the maximum.

Solution of Eqn.13 for z gives z_m = 2.201065891 as an approximate value of z at the maximum. Back substitution of z_m in Eqn.12 allows us to estimate:-

$$d = 1.436488469 \qquad \textbf{Eqn.14}$$

The Estimation of $\Phi(z_m)$

If you require to estimate a value of the Probability Integral at the location of the greatest s displacement (perhaps as a source of x_m for checking purposes) then you can exploit such methods of $\Phi(z)$ approximation as are quoted on Page 932 of "Handbook of Mathematical Functions" edited by Abramowitz and Stegun (Published by Dover of New York).

I disrecommend the Standard Power Series Eqn. 26.2.10 because after nine iterations of the summative series this estimates $\Phi(z)$ as 0.986433456 which shows a 0.034% departure from the (four-figure) tabulated value of 0.9861.

On the other hand the Polynomial Approximation Eqn. 26.2.19 yields $\Phi(z)$=0.986134236 with a claimed error of $\varepsilon(z)<1.5*10^{-7}$. It is not known if my Romberg-based computer program would improve this estimate.

The Determination of the Skew Coefficients, k_1 and k_3

The First Order Skew Coefficient, k_1, and the Third Order Skew Coefficient, k_3, can be determined by the solution of simultaneous equations based upon Eqn.9 which may be transposed to give:-

$$\mathrm{Sin}^{-1}(s_z/d) = 0.25\pi(k_1 z + k_3 z^3) \qquad \textbf{Eqn.15}$$

It is probably best to erect equations for values of z at which the skewed deviation from the probability residual is strongest, i.e. z=±1 and z=±2.
If we can assume that $x \approx x_z$ as we do then introduction of a transposed Eqn.4 enables us to re-write Eqn.15 as:-

$$\mathrm{Sin}^{-1}\left(\frac{x - l_z}{d}\right) = 0.25\pi(k_1 z + k_3 z^3) \qquad \textbf{Eqn.16}$$

or:-

$$\frac{4}{\pi}\mathrm{Sin}^{-1}\left(\frac{x + 0.25 l_0 z}{d}\right) = k_1 z + K_3 z^3$$ **Eqn.17**

Now:-

$z=1$ $x=-1.668267927$

$z=2$ $x=-3.760172043$

Hence:-

$0.887030047 = k_1 + k_3$ **Eqn.18a**

$1.806097325 = 2k_1 + 8k_3$ **Eqn.18b**

Or:-

$7.096240383 = 8k_1 + 8k_3$ **Eqn.19a**

$\underline{1.806097325 = 2k_1 + 8k_3}$ **Eqn.19b**

$5.290143058 = 6k_1$

Therefore:-

$k_1 = 0.881690509$

Substitution of k_1 into Eqn.19a yields:-

$k_3 = 0.005339538285$

These values agree closely with those resulting from heuristic trials with a computer spreadsheet which indicated that $k_1 = 0.9$ and $k_3 = 0.006$.

Suppression of Residual Error (3 August 1991)

The Damped Sine Curve Corrector, c_1

Study of the Residual Error of Romberg minus Warren estimates shows that the error reflects diadically about the origin z=0.

In particular, error in the range $-2 \leq z \leq +2$ is well-modelled by the heuristically-fitted Damped Sine Curve:-

$$c_1 = k_6 e^{-k_7 z^2} \mathrm{Sin}(\pi z)$$ **Eqn.20**

where $k_6 = 0.0000397$, $k_7 = 0.7$ and other notation is as indicated in the notational supplement.

If c_1 is added to the Eqn.1 estimate of the integral the RMS Deviation of the Residual Error is decreased by 65.91% as shown by the worksheet GAUSNEW.CAL.

The Cubed Sine Corrector, c_2

Unlike the other functions so far discussed the Cubed Sine Corrector, c_2, is asymmetric and suited only to regions of positive z. (Negative z integrals may of course be estimated using the symmetry $\Phi(-x) = 1-\Phi(z)$).

c_2 is only useful for modelling high z integrals in regions where z>+2 and indeed its behavior below the abscissa in the region $0 \le z \le +2$ makes it necessary to clip computed c_2 in that region using perhaps a logistic Clipping Function such as:-

$$S = 1 - \left(\frac{1}{1 + e^{u(z^2 - v^2)}} \right) \qquad \textbf{Eqn.21}$$

where u = 10 is most practicable numerically and gives a sufficiently abrupt switching. v is the Transition Score above which S = 1 and below which S = 0.

The Raw Cubed Sine Corrector, c_r, is itself given by:-

$$c_r = k_8 e^{k_9 z} \mathrm{Sin}(k_{10} + 2\pi\, k_{11} z) \qquad \textbf{Eqn.22}$$

where $k_8 = 0.0000092$, $k_9 = 0.0001$, $k_{10} = 4.37$ and $k_{11} = 0.165$. In computations the bracketed term is simplified to (4.37 + 1.036725576z).

Because k_9 is so small it was found that a scarcely inferior model was:-

$$c_r = k_{12}(1 + z)\mathrm{Sin}^3(k_{13} + 2\pi\, k_{11} z) \qquad \textbf{Eqn.23}$$

where $k_{12} = 0.0000021$ and $k_{13} = 4.41$. This latter model does, however, remain untested for Alacrity and Accuracy. I guess it is faster but a bit less precise.

We may now write the Composite Cubed Sine Corrector as:-

$$c_2 = S\, c_r \qquad \textbf{Eqn.24}$$

or:-

$$c_2 = \left[1 - \left(\frac{1}{1 + e^{10(z^2 - 4)}} \right) \right] k_8 e^{k_9 z} \mathrm{Sin}^3(k_{10} + 2\pi\, k_{11} z) \qquad \textbf{Eqn.25}$$

When added to the sum of the Eqn.1 logistic and the Damped Sine Curve Corrector, c_1, this reduces the RMS Deviation by 93.94%.

The Composite Warren Estimator of The Gaussian Distribution Integral

We may now write:-

$$W = \frac{1}{1 + e^{-0.25 l_0 z + d.\mathrm{Sin}[0.25\pi(k_1 z + k_3 z^3)]}} + k_6 e^{-k_7 z^2} \mathrm{Sin}(\pi z) + \left[1 - \left(\frac{1}{1 - e^{u(z^2 - v^2)}}\right)\right] k_8 e^{k_9 z} \mathrm{Sin}^3(k_{10} + 2\pi k_{11} z) \qquad \textbf{Eqn.26}$$

for:-

$k_1 = 0.881690509$
$k_3 = 0.005339538285$
$k_6 = 0.0000397$
$k_7 = 0.7$
$k_8 = 0.0000092$
$k_9 = 0.0001$
$k_{10} = 4.37$
$k_{11} = 0.165$
$k_{12} = 0.0000021$
$k_{13} = 4.41$

Comparisons of the Warren Estimator With Alternative Approximations

The MicroSoft QBasic program GAUSCOMP.BAS and its revision GAUSCOMA.BAS generated Mill Time and Root Mean Square Method Deviation from Romberg Estimate statistics as a by-product of series computation.

Mill Time (in seconds) was employed as a metric of relative Method Alacrity whilst RMS Deviation was used as a precision measure, Romberg Integration being fiducial.

Series and Statistics were computed for 16, 32, 64, 128, 256 and 512 intervals in the range $0 \le z \le +4$. The Alacrity and Accuracy statistics are summarised and displayed by the SuperCalc5 worksheet GAUSTIME.CAL.

These Methods were compared:-

A. Romberg Integration
B. Power Series Summation (20 Iterations)[1]
C. Hasting's Low Precision Polynomial[2]
D. Hasting's High Precision Polynomial[3]
E. The Warren Estimator with Correctors

It was discovered that for each method the solution time rose in virtual linearity with the Number of Intervals and accordingly Mill Time in Seconds per Interval suggested itself as a composite metric. The lower the better.

RMS Deviation was independent of the Number of Intervals but differed for each method. (RMS Deviation was higher for low-interval-number Power Series estimates due to the preponderance of highly-inaccurate summations at extreme z scores). Accordingly the absolute Napierian logarithm of the 512-interval RMS Deviation was adopted as characteristic. The higher the better.

These statistics are tabulated below:-

	Mill Time	RMS Deviation
A	0.662323	
B	0.0687485	7.4934034
C	0.0052719	8.7144399
D	0.0062180	16.1776143
E	0.0215607	15.1488375

It is clear that both Hastings methods are much cheaper than Warren's Method in time expenditure but the Warren Method greatly exceeds the precision of any economic estimator except Hasting's High Precision Polynomial.

References

1. "Handbook of Mathematical Functions"
   Milton Abramowitz and Irene A Stegun (Editors)
   Dover of New York 1970
   SBN 486-61272-4
   p. 932
   Section 26.2.10
2. SBN 486-61272-4
   Section 26.2.18
3. SBN 486-61272-4
   Section 26.2.19

Supplements (not available on the InterNet)

1. Notation for The Approximation of a
   Gaussian Probability Integral as a
   Logistic Function
   GAUSNOTA.DOC
2. Figure 1
   "The Plot of Linear Exponent x = a - bz"
3. The Warren Estimator Component Analysis Worksheet
   GAUSNEW.CAL
4. The Methods' Alacrity and Accuracy Summary
   Worksheet GAUSTIME.CAL
5. The Sixty-Four Interval Estimators' Comparison Worksheet
   GAUS64.CAL