TEACHER'S CORNER

Origin of the Scaling Constant $d = 1.7$ in Item Response Theory

Gregory Camilli
Rutgers, The State University of New Jersey

Key words: *item response theory, normal ogive, logistic ogive*

The scaling constant d $= 1.702$ *used in item response theory minimizes the maximum difference between the normal and logistic distribution functions. The theoretical and numerical derivation of* d *given by Haley (1952) is briefly recapitulated to provide access to curious researchers, instructors, and students.*

The two most common probability models used in item response theory are the normal and logistic distribution functions. As indicated by Birnbaum (1968), the logistic ogive very nearly coincides with the normal ogive and has advantages, both computationally and theoretically. The claim of near equivalence is based on the fact that, with the use of a scaling constant $d = 1.702$,

$$|\Phi(x) - \Psi(dx)| < .01 \tag{1}$$

for all real values of x where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2}, \tag{2}$$

$$\Psi(x) = (1 + e^{-dx})^{-1}.$$

Cox (1970) observed that the most apparent method of scaling ψ to coincide with ϕ is to standardize the logistic variable, which is done by multiplying x by $\pi/\sqrt{3} = 1.81380$. Johnson and Kotz (1970) graphically showed that the scaling could be improved by the factor $(\pi/\sqrt{3})\,(15/16) = 1.70044$. However, Haley (1952) outlined the theoretical derivation of $d = 1.702$. Because the use of d is widespread and Haley's (1952) unpublished work is not easily accessible, the derivation of d re-presented in this brief note provides a resource to users and instructors of item response theory. I emphasize that the value of d is not intrinsically important; linear scaling does not affect the fit of logistic models to item response data. Rather, this note serves the purposes of satisfying the curious, and historical completeness.

Theoretical Derivation

The derivation procedure presented below can be conceptualized graphically as choosing d so that the maximum difference between $\psi(dx)$ and $\phi(x)$ is as small as possible. Thus, d is a *minimax* estimator; other types of estimators might minimize the average squared or absolute difference between ψ and ϕ. In the approach taken by Haley (1952), the difference between ψ and ϕ was defined as

$$F(x, d) = \Phi(x) - \Psi(dx). \tag{3}$$

To find the largest values of the difference, the first derivative of $F(x,d)$ is taken with respect to x, and the result is set equal to zero:

$$\begin{aligned} \frac{\partial}{\partial x} F(x, d) &= f_x(x, d) \\ &= \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} - \frac{de^{-dx}}{(1 + e^{-dx})^2} \\ &= 0. \end{aligned} \tag{4}$$

Haley (1952) proved that this equation has four roots of which two are positive and two are negative (the negative roots are symmetric around zero with the positive roots). Furthermore, for values of d near $\pi/\sqrt{3}$, it is visually evident that, for $x > 0$, $F(x,d)$ is positive at one root (say x_1) and negative at the other (say x_2). Thus, the minimax estimator can be defined as the value of d at which $F(x_1,d)$ is equal to $-F(x_2,d)$, or

$$S(x_1, x_2, d) = F(x_1, d) + F(x_2, d) = 0. \tag{5}$$

Thus, while x_1 and x_2 are the positive roots to the equation $f_x(x,d) = 0$, d is the root to the equation $S(x_1,x_2,d) = 0$.

Numerical Derivation

One simple method of solving the two nonlinear equations given above can be described in three steps. Using an iterative algorithm to find the roots of a nonlinear equation:

1. Solve $f_x\ (x,d) = 0$ using a trial value of d (say $\pi/\sqrt{3}$) and obtain the positive roots x_1 and x_2.

2. Fixing x_1 and x_2, and using the trial d (from Step 1) as a starting value, solve the equation $S(x_1, x_2, d) = 0$ for an updated value of d.

3. Repeat Steps 1 and 2 until convergence, using the most current roots x_1 and x_2 as starting values in Step 1 and the most current root d as the starting value in Step 2.

This procedure was employed, and the following values of x_1, x_2 and d were obtained in five iterations:

Iteration	d	x_1	x_2
0	1.8	.5	1.5
1	1.699205	.674591	2.334629
2	1.701731	.566725	2.035436
3	1.701744	.571034	2.043031
4	1.701745	.571055	2.043070
5	1.70174439	.571057	2.043073

At Iteration 5, the largest absolute value of $F\ (x_1, d)$ was found to be .009457, which occurs at the roots $\pm$ .571057 and $\pm$ 2.043073. These results are in accord with those given by Haley (1952).

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Cox, D. R. (1970). *The analysis of binary data.* London: Methuen.

Haley, D. C. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error, Technical Report No. 15* (Office of Naval Research Contract No. 25140, NR-342-022). Stanford University: Applied Mathematics and Statistics Laboratory.

Johnson, N. J., & Kotz, S. (1970). *Continuous univariate distributions—2.* Boston: Houghton Mifflin.

Author

GREGORY CAMILLI is Professor, Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08903. He specializes in applied measurement and statistics.

Received May 26, 1993
Accepted October 10, 1993