New Sigmoid-like function better than Fisher z transformation

Authors: Sana Afreen, Naomi Cedeño-Manrique, Luis Pinto-Alemán

Statistics for Data Science Final Project

Master of Data Science and Business Informatics

May 29, 2024

Summary

Standard Normal Distribution

Standard Normal Distribution

A continuous random variable has a standard normal distribution with parameters $\mu = 0$ and $\sigma^2 = 1$ if its probability density function f is given by:

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

We denote this distribution by $N \sim (0, 1)$, where $f(x)$ is also written as $\phi(x)$.

CDF of a Standard Normal Distribution

Let $x, t \in \mathbb{R}$. The cumulative distribution function of the standard normal distribution is:

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

Problem Statement

The cumulative distribution function of normal distribution is **not an explicit elementary** function, and it cannot be solved analytically. If its accurate value is required, equation (2) is approximated as:

$$\Phi(x) \approx \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{3}$$

where $\Phi \in [0, 1]$.
On the other hand, $\operatorname{erf} \in [-1, +1]$ is the **Gauss error function** employed inside the approximation, the latter being written as:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{4}$$

The main problem is that the approximation with erf, although analytical, is **still composed by an special function**. This makes it difficult to **describe causalities** between the different variables involved.

The Fisher z Transformation

In 1915, **Fisher** [1] proposed the Fisher z transformation. Let $r \in [-1, 1]$ be the Pearson's correlation coefficient. In this context, the z transformation is proposed as a way to transform the sampling distribution of r so that it **approaches a normal distribution** using the following formula:

$$z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \tag{5}$$

The inverse of equation (5) is:

$$r = \frac{\exp(2z_r) - 1}{\exp(2z_r) + 1} \tag{6}$$

Moreover, $z_r \sim \left(\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right), \frac{1}{n-3} \right)$ **approximately**. Using the inverse given in the equation (6), it is possible to approximate $\operatorname{erf}(x)$, so replacing in $\Phi(x)$:

$$\Phi(x) \approx \frac{1}{2} \left[1 + \frac{\exp(2z_r) - 1}{\exp(2z_r) + 1} \right] \tag{7}$$

Other approximations

- The Fisher z transformation can be employed to approximate erf and, consequently, $\Phi(x)$; and is widely used because it leaves the CDF in terms of elementary functions, which facilitates **comparisons and understanding of causal relationships**.
- There are at least **9 explicit elementary functions** to approximate the theoretical CDF, and their maximal distances to this range from $6.8 \cdot 10^3$ to $3.0 \cdot 10^{10}$ [3].
- For approximating this distributions, not only numerical accuracy is sought, but also interpretability, i.e. that the variables characterizing complex systems can be described by causality. In this context, it is important to find an approach with a **good tradeoff between being sufficiently accurate and robust to explain the inherent complexity** of the data.

The Sigmoid-like Function

Author's proposal for an explicit elementary function that is approximate is the **Sigmoid-like function**. It aims to close approximately to the standard normal distribution by:

$$SY(y) = \frac{1}{1+\exp(-ky)} \tag{8}$$

where $y \in (-\infty, +\infty)$ and $SY \in [0, 1]$. Taking a **change of variable** $y = \frac{x}{\sqrt{2}}$ and replacing it in equation (8), then:

$$SY(x) = \frac{1}{1+\exp\left(-k\frac{x}{\sqrt{2}}\right)} \tag{9}$$

corresponds with equation (3). On the other hand, erf can be approximated as:

$$\begin{aligned} SE(x) &= 2 \cdot SY(x) - 1 \\ &= \frac{2}{1+\exp(-kx)} - 1 \end{aligned} \tag{10}$$

which corresponds directly with equation (4), where $x \in (-\infty, +\infty)$ and $SE \in [-1, 1]$.

The Sigmoid-like Function

Then, the Fisher z transformation is suggested to be replaced by the function:

$$SZ(r) = \frac{1}{k} \ln\left(\frac{1+r}{1-r}\right) \tag{11}$$

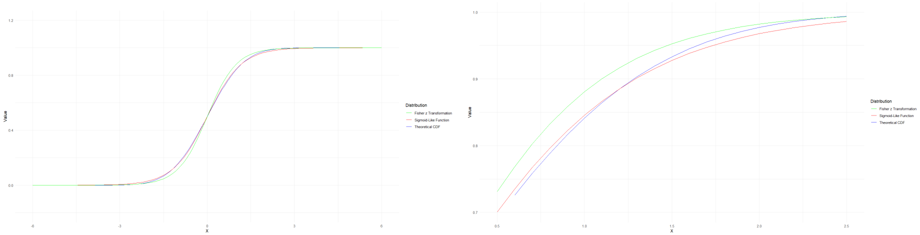
As it can be observed in equations (11) and (5), the Fisher z transformation is a **particular case** of the Sigmoid-like function for $k = 2$.
On the other hand, the inverse function of $SZ(r)$ will be:

$$r = \frac{\exp(k \cdot SZ(r)) - 1}{\exp(k \cdot SZ(r)) + 1} \tag{12}$$

where $r \in (-1, +1)$ and $SZ \in [-\infty, +\infty]$.

Sigmoid-like function vs. Fisher z transformation

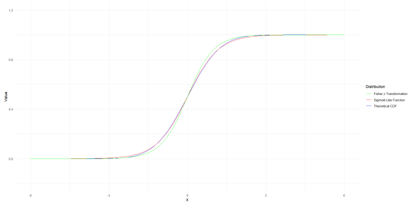
The best k in $SY(x)$ is not easy to get only by analytical deduction. Now, for graphical comparison purposes, the authors provided $k = 1.701$ as a plausible candidate for the optimal value.



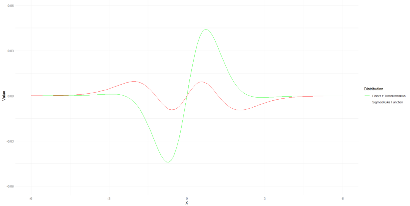
(a) General comparison (b) Detailed view

Figure: CDF of standard normal distribution and its explicit elementary function approximations by the Sigmoid-like function and the Fisher z transformation.

Sigmoid-like function vs Fisher z transformation



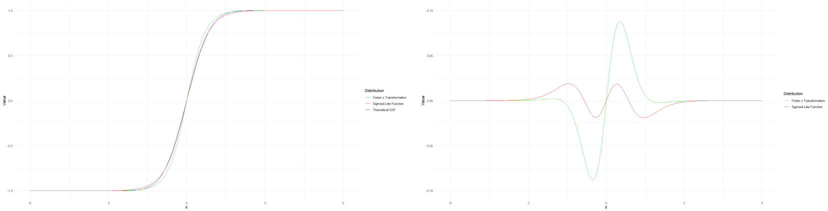
(a) CDF



(b) Direct errors of CDF

Figure: Results with respect to the CDF.

Sigmoid-like function vs Fisher z transformation



(a) ERF

(b) Direct errors of ERF

Figure: Results with respect to the error function.

Optimizing the Coefficient k

Both analytical optimization and numerical deduction will be used to find the optimal k. To compare the best candidates, two methods will be used:

1. **Minimal Total Error:** Aims to minimize the error between $SY(x)$ and $\Phi(x)$ over k by:

$$\int_{-\infty}^{+\infty} (SY(k,x) - \Phi(x))^2 \, \mathrm{d}x \tag{13}$$

   Denote the minimum as K_t.

2. **Minimum Maximal Distance:** Searches for the minimum over k and x of:

$$|SY(k,x) - \Phi(x))| \tag{14}$$

   Let the optimal value be K_d.

Optimal k with Minimal Total Error

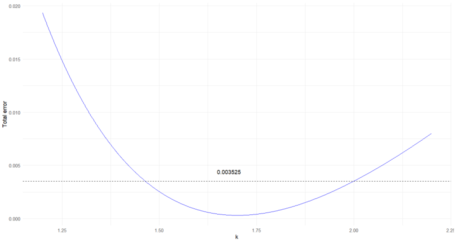
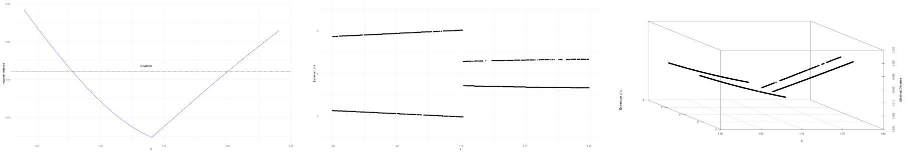


Figure: Minimal total error for different values of k, comparing the performance of Fisher z transformation with the Sigmoid-like function.

The interval where the sigmoid function has an error less than that of Fisher's z transformation is $k \in (1.464, 2.000)$.

Optimal k with Minimal Maximum Distance



(a) Minimal total error for different values of k

(b) k values and its extremum of x

(c) 3D version of 5b

Figure: Maximal distances between SY(x) and $\Phi(x)$, comparing the performance of Fisher z transformation with the Sigmoid-like function.

The interval that shows better performance than Fisher's z transformation is $k \in (1.390, 2.000)$.

Approximate Errors Contrast

	k	Total error to $\Phi(x)$	Maximal distance to $\Phi(x)$
K$_t$	1.7009	0.0003140087	0.009507619
K$_d$	1.7010	0.0003140084	0.0095016
K$_a$	1.7017	0.0003140314	0.0094598

Table: Error contrast of the new Sigmoid-like function and the Fisher z transformation.

Conclusions and Future Research

1. The Sigmoid-like function can enhance the interpretability in the cases where the Fisher Z-transform has been used, as in estimation of confidence interval of Pearson product moment correlation coefficient.
2. The Sigmoid-like function allows to approximate the standard normal distribution function. With a value of $k = 1.701$, this function is 11.225 times more accurate than Fisher's z transformation using the total, while it is 4.655 times better by the minimal maximum distance.
3. Despite the computational limitation involved in the number of steps taken to optimize k, resulting in a variation of the number of digits, the results do not vary considerably from the paper.
4. The search for the best parameter k is still open.

Bibliography

1. Fisher, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* **10,** 507–521. ISSN: 00063444. http://www.jstor.org/stable/2331838 (2024) (1915).
2. Yang, Z.-L., Song, Y.-W., Duan, Z.-F., Wang, T. & Zhang, J. New Sigmoid-like function better than Fisher z transformation. *Communications in Statistics - Theory and Methods* **45,** 2332–2341. https://doi.org/10.1080/03610926.2013.771750 (2016).
3. Yun, B. I. Approximation to the cumulative normal distribution using hyperbolic tangent based functions. *Journal of The Korean Mathematical Society - J KOREAN MATH SOC* **46,** 1267–1276 (Nov. 2009).

Thanks!