Journal of Uncertain Systems
Vol.9, No.2, pp.103-112, 2015
Online at: www.jus.org.uk

Standard Statistical Transformations (Logarithm and Logit) are Uniquely Determined by the Corresponding Symmetries

Misha Koshelev*
1330 Old Spanish Trail #5303, Houston, TX 77054-1835, USA

Received 1 January 2014; Revised 1 February 2014

Abstract

Logarithm and logit transformations are well-established in statistics: logarithm transforms an all-positive quantity into a quantity that can take arbitrary real value, and logit does the same for a random variables whose values are limited to the interval $(0, 1)$. In this paper, we analyze possible symmetries of such transformations, and we also show that, in effect, these two transformations are the only ones which are invariant with respect to the corresponding symmetries.

Keywords: standard statistical transformations, logit, symmetry

1 Introduction: Formulation of the Problem

Standard statistical transformations: logarithm and logit. For many statistical distributions (e.g., for a normal distribution), the probability density is everywhere positive. This means that for such distributions, in principle, all real numbers are possible with non-zero probability.

In some practical situations, the quantity of interest x cannot take all real values. For example:

- some quantities (such as mass) are always positive;
- some quantities (such as fraction) are always located between 0 and 1.

One of the well-established statistical practices for handling such situations is to apply an appropriate continuous 1-1 transformation $x \to y = f(x)$ that transforms the range of x into the whole real line.

Usually (see, e.g., [2, 3]),

- logarithm $y = \ln(x)$ is used for all-positive quantities, and
- logit (logistic transformation)
$$y = \ln\left(\frac{x}{1-x}\right)$$
is used for quantities that lie between 0 and 1.

Often, it is assumed that the transformed variable $y = f(x)$ (whose range is the whole real line) is normally distributed. For example, random variables x whose logarithm $\ln(x)$ is normally distributed are known as *lognormally distributed*. Lognormal distribution indeed occurs in many real-life situations.

Similar transformations are used for different ranges of x. In some cases, the range of possible values of the variable x is the semi-line (x_0, ∞). For example, in the Kelvin scale, possible temperatures take all possible non-negative values, but in the Celsius scale, in which the absolute zero is $x_0 = -273.15°$C, possible temperatures form the range (x_0, ∞). In this case, a natural idea is to first shift the values into the scale $\widetilde{x} = x - x_0$ in which the lower bound is 0 (e.g., into the Kelvin scale for temperatures), and then apply the logarithm transformation $y = \ln(\widetilde{x})$ to the quantity expressed in the new scale. This is equivalent to applying the transformation $y = \ln(x - x_0)$ to the original scale.

*Corresponding author.
Email: misha680@gmail.com (M. Koshelev).

Similarly, if some quantity x is known to be always located between $\underline{x}$ and $\overline{x}$, a natural idea is to first apply a linear transformation that would transform this interval $(\underline{x}, \overline{x})$ into the interval $(0, 1)$, and then apply the logit transformation. The corresponding linear transformation has the form

$$\widetilde{x} = \frac{x - \underline{x}}{\overline{x} - \underline{x}}$$

and thus, this two-stage procedure is equivalent to applying the transformation

$$y = \ln\left(\frac{\widetilde{x}}{1 - \widetilde{x}}\right) = \ln\left(\frac{x - \underline{x}}{\overline{x} - x}\right).$$

What we do in this paper. In this paper, we analyze possible symmetries of these transformations, and we also show that logarithm and logit are, in effect, the only transformations which are invariant with respect to the corresponding symmetries.

Comment. This result may provide one more reason why logarithm and logit are so well-established in the statistical practice.

2 How to Describe Symmetries of a Transformation: Analysis of the Problem

General idea. In general, how can we describe symmetries of a transformation $f : X \to Y$ between the two sets X and Y? To provide a meaningful definition of this notion, let us start with a simple physical example.

It is known that the dependence $F = C \cdot r^{-2}$ of the gravitational force F on the distance r between the interacting bodies is *scale-invariant* – in the sense that this dependence does not allow us to fix any single value of the distance r_0. Specifically, when we change the unit for measuring distance, the numerical value change from r to $\widetilde{r} = \lambda \cdot r$ for some constant λ. The numerical value $\widetilde{r}^{-2}$ is, of course, different from the numerical value of r^{-2}, but we can retain the same formula $\widetilde{F} = C \cdot \widetilde{r}^{-2}$, with the exact same constant C, if we accordingly change the measuring unit for the force F, to $\widetilde{F} = \lambda^{-2} \cdot F$.

The situation is completely different if we have a dependence of the type $F = \exp(-k \cdot r)$ for some constant k. In this case, if we change r to $\widetilde{r} = \lambda \cdot r$, the dependence between $\widetilde{F}$ and $\widetilde{r}$ will be different $\widetilde{F} \neq \exp(-k \cdot \widetilde{r})$ no matter how we re-scale F; see, e.g., [1].

This physical example prompts the following idea:

- we select a reasonable group G_X of mappings ("symmetries") $g : X \to X$;
- we select a reasonable group G_Y of mappings ("symmetries") $h : Y \to Y$; and
- we say that a transformation $f : X \to Y$ is *invariant* if for every mapping $g \in G_X$, there exists an appropriate mapping $h \in G_Y$ that preserves the same dependence between x and y, i.e., if $y = f(x)$, then $\widetilde{y} = f(\widetilde{x})$, where $\widetilde{x} \stackrel{\text{def}}{=} g(x)$ and $\widetilde{y} \stackrel{\text{def}}{=} h(y)$.

In our case, the set Y is the set of all real numbers, and the set X is a semi-line or an interval. So, in order to describe our idea in precise terms, we must find our which mappings (symmetries) are reasonable for such sets.

Natural symmetries of the real line Y. Let us start with the real line Y. As we have already mentioned, one natural symmetry comes from the fact that in order to get a numerical value of a quantity, we must select a unit for its measurement. For example, when we measure the same distance in meters and in feet, we get different numerical results. In general, when we replace the original measuring unit with a new unit which is λ times smaller, then instead of the original numerical value x we get a new numerical value $x' = \lambda \cdot x$. For example, if instead of meters, we use centimeters, then the distance $x = 10$ meters becomes $x' = 1000$ centimeters.

Another natural symmetry comes from the fact that for many physical quantities, their numerical value also depends on the selection of a starting point. In general, if we replace the original starting point with a new point which precedes the original one by x_0 units, then the original numerical value x is replaced by a new value $x' = x + x_0$. For example, if instead of the usual calendar, we use an alternative calendar that starts at year 1000 B.C.E. (before common era), then instead of 2010, we will count year 3010.

In general, by changing the measuring unit *and* the starting point, we replace the original numerical value x with a new value $x' = a \cdot x + b$ for some $a > 0$ and b. An example of such a re-scaling is provided by a known relation between Fahrenheit (F) and Celsius (C) temperature scales, where $t_F = 32 + 1.8 \cdot t_C$.

Thus, for the set Y, as natural symmetries, we take mappings (re-scalings) of the type $g(y) = a \cdot x + b$ for $a > 0$ and arbitrary b.

These mappings form a *transformation group* in the sense that

- First, if $\widetilde{y} = g(y)$ is a reasonable re-scaling, then the inverse mapping $y = g^{-1}(\widetilde{y})$ should also be a reasonable re-scaling.
- Second, if $y \to y_1 = g_1(x)$ is a reasonable re-scaling, and $y_1 \to y_2 = g_2(y_1)$ is a reasonable re-scaling, then the corresponding mapping $y \to y_2 = g_2(y_1) = g_2(g_1(y))$ should also be a reasonable re-scaling.

In other words, this class of functions is a transformation group because it is closed under inverse function and under composition.

Natural symmetries of the semi-line $(0, \infty)$**.** When the set X of all possible values of x is a semi-line $(0, \infty)$, then we cannot arbitrarily change the starting point – because 0 is a natural starting point, but we can still change the measuring unit. In this case, we arrive at the re-scalings $x \to a \cdot x$ – from the example with which this section started.

Natural symmetries of the interval $(0, 1)$**.** The situation is more complex when the set X of possible values of the variable x is the interval $(0, 1)$. In this case, we cannot change the starting point – it is 0, and we cannot change the measuring unit – because the largest possible value is 1. Thus, we cannot apply any *linear* mappings between different scales.

In some practical applications, however, there are also reasonable *non-linear* mappings between scales. For example, the energy of an earthquake can be described both in the usual energy units and in the logarithmic (Richter) scale. Similarly, the power of a signal and/or of a sound can be measured in watts and it can also be measured in the logarithmic scale, in decibels.

It is therefore reasonable to consider, instead of the original class of all linear re-scalings, a more general class of mappings that would include some non-linear ones as well. What non-linear re-scalings should we consider?

As before, if $\widetilde{x} = g(x)$ is a reasonable re-scaling, then the inverse mapping $x = g^{-1}(\widetilde{x})$ (corresponding to going back to the original scale) should also be a reasonable re-scaling. Second, if $x \to x_1 = g_1(x)$ is a reasonable re-scaling, and $x_1 \to x_2 = g_2(x_1)$ is a reasonable re-scaling, then the corresponding mapping $x \to x_2 = g_2(x_1) = g_2(g_1(x))$ should also be a reasonable re-scaling. Thus, the desired class of re-scalings should be closed under inverse function and under composition, i.e., this class should form a transformation group.

We talk about meaningful re-scalings, re-scalings that correspond to real-life situations and whose results can be stored in a computer. In a computer, we can only store finitely many parameters. Thus, our transformation group has to be determined by finitely many parameters. In mathematical terms, this means that the corresponding group should be *finite-dimensional*. It is also reasonable to require that the dependence on these parameters should be *smooth* (differentiable), i.e., crudely speaking, that small changes in these parameters should lead to proportionally small changes in the result. In mathematical terms, this means that our transformation group is a *Lie group*.

Thus, we are looking for a (finite-dimensional) Lie group G of mappings of the real line that contains all linear mappings. For an Euclidean space of arbitrary dimension n, such Lie groups have been classified in [4] (following a hypothesis formulated by Norbert Wiener in [5]). In particular, for our case $n = 1$, the only such groups are the group of linear mappings and the group of all fractionally-linear mappings

$$g(x) = \frac{a \cdot x + b}{c \cdot x + d}. \qquad (1)$$

Since we are interested in possibly non-linear re-scalings, we should therefore consider arbitrary re-scalings of the type (1).

Such transformations form a group, and linear mappings are a particular case of this group – corresponding to $c = 0$.

We want re-scalings $g : X \to X$ that map the interval $X = (0, 1)$ onto itself. Thus, 0 should map into either 0 or 1, and 1 into either 1 or 0. For mappings that continuously change from the original value x to the new value $F(x)$, we cannot jump from 0 to 1, thus, 0 should map into 0 and 1 into 1. In other words, we should have $F(0) = 0$ and $F(1) = 1$.

For $x = 0$, the equation $F(0) = 0$ means that $b/d = 0$ hence $b = 0$. Thus, the general expression (1) takes the form

$$F(x) = \frac{a \cdot x}{c \cdot x + d}. \tag{2}$$

The value $F(x)$ is not identically 0, so $a \neq 0$. Thus, we can divide both the numerator and the denominator by a. As a result, we conclude that

$$F(x) = \frac{x}{C \cdot x + D}, \tag{3}$$

where we denoted $C \stackrel{\text{def}}{=} c/a$ and $D \stackrel{\text{def}}{=} d/a$.

The condition $F(1) = 1$ takes the form $C + D = 1$, thus $D = 1 - C$, and the re-scaling (3) takes the form

$$F(x) = \frac{x}{C \cdot x + (1 - C)}. \tag{4}$$

These are the symmetries that we will consider in this paper.

Example. As an example of such a re-scaling, let us consider the Bayesian mapping of the prior probability $P_0(H_0)$ of the hull hypothesis H_0 into the posterior probability $P(H_0|E)$ based on the evidence E:

$$P(H_0|E) = \frac{P(E|H_0) \cdot P_0(H_0)}{P(E|H_0) \cdot P_0(H_0) + P(E|H_1) \cdot P_0(H_1)}, \tag{5}$$

where H_1 is the alternative hypothesis. Here, $P_0(H_1) = 1 - P_0(H_0)$, hence

$$P(H_0|E) = \frac{P(E|H_0) \cdot P_0(H_0)}{P(E|H_0) \cdot P_0(H_0) + P(E|H_1) \cdot (1 - P_0(H_0))}. \tag{6}$$

As one can see from the formula (6), the mapping that maps the prior probability $P_0(H_0)$ into the posterior probability $P(H_0|E)$ is fractionally linear, and it maps $P_0(H_0) = 0$ into $P(H_0|E) = 0$ and $P_0(H_0) = 1$ into $P(H_0|E) = 1$. Thus, this re-scaling has the form (4).

What if we apply fractionally-linear re-scalings to the semi-line $(0, \infty)$? Our discussion of the variables x whose values are in the interval $(0, 1)$ made us conclude that in looking for reasonable symmetries, we should go beyond linear mappings and consider fractionally linear mappings as well. A natural question is: what is we consider these (more general) fractionally-linear mappings (1) also in the case when the set X of all possible values of the variable x is the semi-line $(0, \infty)$?

We want each mapping $g : (0, \infty) \to (0, \infty)$ to be continuous and 1-1 (otherwise, it would not have an inverse). Thus, $g(x)$ must be monotonic – and hence, in the limit, it must map 0 into 0, and ∞ into ∞.

Substituting $x = 0$ into the formula (1), we conclude that $b = 0$, i.e., that

$$g(x) = \frac{a \cdot x}{c \cdot x + d}. \tag{7}$$

If $c \neq 0$, then in the limit $x \to \infty$, we will have $g(x) \to a/c \neq \infty$. Thus, the only possibility for this mapping to map ∞ into ∞ is to have $c = 0$ – in which case we have a mapping $g(x) = \lambda \cdot x$, with $\lambda = a/d$.

Thus, for the set $X = (0, \infty)$, even if we start with the group of all fractionally-linear mappings, we still end up with the same group of mappings $x \to \lambda \cdot x$ as when we started with the group of all linear mappings.

3 Invariant Transformations: Definitions and the Main Results

Now, we are ready for the formal definitions. Let us first describe, in general terms, what we mean by transformations that have symmetries.

Definition 1. *Let X be a topological space with a group G_X of mappings $X \to X$, and let Y be a topological space with a group G_Y of mappings $Y \to Y$.*

- *By a* transformation $f : X \to Y$*, we mean a continuous 1-1 mapping of the set* X *onto the entire set* Y*.*
- *We say that a transformation* $f : X \to Y$ *is* invariant *if for every mapping* $g \in G_X$*, there exists a mapping* $h \in G_Y$ *for which* $f(x) = y$ *implies* $f(\widetilde{x}) = \widetilde{y}$*, where* $\widetilde{x} \stackrel{\text{def}}{=} g(x)$ *and* $\widetilde{y} \stackrel{\text{def}}{=} h(y)$*.*

One can easily check that for every invariant transformation $f : X \to Y$ and for every mapping $h \in G_y$, the transformation $\widetilde{f}(x) \stackrel{\text{def}}{=} h(f(x))$ is also invariant. Thus, we arrive at the following definition:

Definition 2. *We say that transformations* $f : X \to Y$ *and* $\widetilde{f} : X \to Y$ *are* equivalent *if there exists a transformation* $h \in G_y$ *for which, for all* x*, we have* $\widetilde{f}(x) = h(f(x))$*.*

In these terms, the definition of invariance can be reformulated as follows:

Definition 3. *We say that a transformation* $f : X \to Y$ *is* invariant *if for every mapping* $g \in G_X$*, the transformation* $\widetilde{f}(x) \stackrel{\text{def}}{=} f(g(x))$ *is equivalent to the original transformation* $f(x)$*.*

Our first result is about the case when the set X of all possible values is a semi-line $(0, \infty)$. In this case, as we have mentioned, as the class of all natural mappings G_X on the semi-line X, it is reasonable to take the class all mappings of the type $x \to \lambda \cdot x$ with $\lambda > 0$.

The set Y is the real line. In this case, as the set G_Y of the corresponding mappings, it is reasonable to take the class G_Y of all linear mappings $y \to a \cdot y + b$ of the real line Y for which $a > 0$. In this case, the above definitions take the following form:

Definition 1′. *By a* $(0, \infty)$-transformation*, we mean a continuous 1-1 mapping* $f : (0, \infty) \to (-\infty, \infty)$ *that transforms the set of all positive numbers into the whole real line.*

Definition 2′. *We say that two* $(0, \infty)$*-transformations* $f(x)$ *and* $\widetilde{f}(x)$ *are* equivalent *if there exist real numbers* $a > 0$ *and* b *for which, for every* x*, we have* $\widetilde{f}(x) = a \cdot f(x) + b$*.*

Definition 3′. *We say that a* $(0, \infty)$*-transformation* $f(x)$ *is* invariant *if for every* $\lambda > 0$*, the transformation* $\widetilde{f}(x) \stackrel{\text{def}}{=} f(\lambda \cdot x)$ *is equivalent to the original transformation* $f(x)$*.*

All invariant $(0, \infty)$-transformations are described by the following result:

Theorem 1. *Every invariant* $(0, \infty)$*-transformation is equivalent to* $f(x) = \ln(x)$*.*

(For readers' convenience, the proofs of all the results are placed in a special (last) Proofs section of this paper.)

Comment. That the logarithm $f(x) = \ln(x)$ is invariant is clear: when we multiply x by λ, the new function $\widetilde{f}(x) = f(\lambda \cdot x)$ takes the form $\ln(\lambda \cdot x) = \ln(\lambda) + \ln(x)$. This new function is clearly equivalent to the original function $f(x)$, with $a = 1$ and $b = \ln(\lambda)$.

Conclusion. Thus, in effect, the only invariant $(0, \infty)$-transformation is the logarithm.

When the set X of all possible values is an interval $(0, 1)$, the class G_X consists of mappings of the type

$$x \to \frac{x}{C \cdot x + (1 - C)}.$$

The set Y is the same – the real line, so, as G_Y, we take the class of all linear transformations $y \to a \cdot y + b$ of the real line Y for which $a > 0$. In this case, the above definitions take the following form:

Definition 1″. *By a* $(0, 1)$-transformation*, we mean a continuous 1-1 mapping* $f : (0, 1) \to (-\infty, \infty)$ *that transforms the interval* $(0, 1)$ *into the whole real line.*

Comment. In other words, we require that every real number y can be represented as $f(x)$ for some $x \in (0, 1)$.

Definition 2″. *We say that two* $(0, 1)$*-transformations* $f(x)$ *and* $\widetilde{f}(x)$ *are* equivalent *if there exist real numbers* $a > 0$ *and* b *for which, for every* x*, we have* $\widetilde{f}(x) = a \cdot f(x) + b$*.*

Definition 3″. *We say that a* $(0,1)$*-transformation* $f(x)$ *is* invariant *if for every* C*, the transformation*

$$\widetilde{f}(x) \stackrel{\rm def}{=} f\left(\frac{x}{C\cdot x+(1-C)}\right)$$

is equivalent to the original transformation $f(x)$*.*

All invariant $(0,1)$-transformations are described by the following result:

Theorem 2. *Every invariant* $(0,1)$*-transformation is equivalent to the logistic transformation*

$$f(x)=\ln\left(\frac{x}{1-x}\right).$$

Conclusion. Thus, in effect, the only invariant transformation is logit.

4 Proofs

4.1 Proof of Theorem 1

1°. Let us assume that $f(x)$ is an invariant $(0,\infty)$-transformation. By definition of invariance, this means that for every $\lambda>0$, the transformation $f(\lambda\cdot x)$ is equivalent to $f(x)$. By definition of the equivalence, this means that for every λ, there exist values $a(\lambda)$ and $b(\lambda)$ for which, for every x, we have

$$f(\lambda\cdot x)=a(\lambda)\cdot f(x)+b(\lambda). \tag{8}$$

2°. Since the function $f(x)$ is continuous and 1-1, it is monotonic. It is known that every monotonic function from real number to real numbers is almost everywhere differentiable. Let $x_0>0$ be a point at which the function $f(x)$ is differentiable. Then, for every $x_1>0$, by using the formula (8) with $\lambda=x_1/x_0$, we can conclude that the function f is differentiable at this point x_1 as well. Thus, the function $f(x)$ is differentiable for all x.

Let us prove that the functions $a(\lambda)$ and $b(\lambda)$ are also differentiable.

2.1°. Let us start by proving that the function $a(\lambda)$ is differentiable.

Indeed, let us take any two values x_1 and x_2 for which $f(x_1)\ne f(x_2)$. Substituting $x=x_1$ and $x=x_2$ into the formula (8), we conclude that

$$f(\lambda\cdot x_1)=a(\lambda)\cdot f(x_1)+b(\lambda); \tag{9}$$

$$f(\lambda\cdot x_2)=a(\lambda)\cdot f(x_2)+b(\lambda). \tag{10}$$

Subtracting (10) from (9), we conclude that

$$f(\lambda\cdot x_1)-f(\lambda\cdot x_2)=a(\lambda)\cdot(f(x_1)-f(x_2)), \tag{11}$$

hence

$$a(\lambda)=\frac{f(\lambda\cdot x_1)-f(\lambda\cdot x_2)}{f(x_1)-f(x_2)}. \tag{12}$$

Since the function $f(x)$ is differentiable, the right-hand side of (12) is differentiable as well, hence the function $a(\lambda)$ is differentiable.

2.2°. Let us now prove that the function $b(\lambda)$ is also differentiable.

Indeed, from the formula (9), we conclude that

$$b(\lambda)=f(\lambda\cdot x_1)-a(\lambda)\cdot f(x_1). \tag{13}$$

Since the function $f(x)$ is differentiable, and the function $a(\lambda)$ has also been proven to be differentiable, we thus conclude that the right-hand side of the equality (13) is differentiable as well, hence the function $b(\lambda)$ is also differentiable.

3°. Now that we know that all three functions $f(x)$, $a(\lambda)$, and $b(\lambda)$ are differentiable, we can differentiate both side of the equality (8) by λ. To differentiate the left-hand side of (8), we use the chain rule and the fact that $\frac{d(\lambda\cdot x)}{d\lambda}=x$. As a result, we get the following formula

$$x\cdot f'(\lambda\cdot x)=\frac{da}{d\lambda}(\lambda)\cdot f(x)+\frac{db}{d\lambda}(\lambda), \tag{14}$$

where $f'\stackrel{\text{def}}{=}\frac{df}{dx}$.

In particular, for $\lambda=1$, we get the following equality:

$$x\cdot\frac{df}{dx}=a\cdot f+b, \tag{15}$$

where we denoted $a\stackrel{\text{def}}{=}\frac{da}{d\lambda}_{|\lambda=1}$ and $b\stackrel{\text{def}}{=}\frac{db}{d\lambda}_{|\lambda=1}$.

By moving all the terms related to x to one side of this equation and all the terms related to f to the other side, we get the following equation:

$$\frac{df}{a\cdot f+b}=\frac{dx}{x}. \tag{16}$$

To solve this equation, let us consider two possible case: when $a\neq 0$ and when $a=0$.

3.1°. When $a\neq 0$, then $a\cdot f+b$ can be represented as

$$a\cdot\left(f+\frac{b}{a}\right),$$

thus (16) takes the form

$$\frac{1}{a}\cdot\frac{df_1}{f_1}=\frac{dx}{x}, \tag{17}$$

where we denoted

$$f_1(x)\stackrel{\text{def}}{=}f(x)+\frac{b}{a}.$$

Integrating both sides of the equation (17), we conclude that

$$\frac{1}{a}\cdot\ln(f_1)=\ln(x)+C \tag{18}$$

for some constant C, hence

$$\ln(f_1)=a\cdot\ln(x)+C_1, \tag{19}$$

where $C_1\stackrel{\text{def}}{=}a\cdot C$.

Applying $\exp(z)$ to both sides of this equality, we conclude that

$$f_1(x)=\exp(a\cdot\ln(x)+C_1)=C_2\cdot x^a, \tag{20}$$

where $C_2\stackrel{\text{def}}{=}\exp(C_1)$. Thus,

$$f(x)=f_1(x)-\frac{b}{a}=C_2\cdot x^a-\frac{b}{a}. \tag{21}$$

Since $x>0$, the value $f(x)$ are always greater than or equal to $-b/a$, and thus, cannot take any real value smaller than $-b/a$.

This contradicts to our definition of an $(0,\infty)$-transformation. Therefore, the case $a\neq 0$ is impossible.

3.2°. Let us now consider the only remaining case $a=0$. In this case, the equation (16) takes the form

$$\frac{1}{b}\cdot df=\frac{dx}{x}. \tag{22}$$

Integrating both sides of the equation (22), we conclude that

$$\frac{1}{b}\cdot f=\ln(x)+C \tag{23}$$

for some constant C, hence

$$f(x)=b\cdot\ln(x)+b\cdot C. \tag{24}$$

Thus, the function $f(x)$ is indeed equivalent to $\ln(x)$. The theorem is proven.

4.2 Proof of Theorem 2

1°. Let us assume that $f(x)$ is an invariant $(0,1)$-transformation. By definition of invariance, this means that for every C, the transformation

$$f\left(\frac{x}{C\cdot x+(1-C)}\right)$$

is equivalent to $f(x)$. By definition of the equivalence, this means that for every C, there exist values $a(C)$ and $b(C)$ for which, for every x, we have

$$f\left(\frac{x}{C\cdot x+(1-C)}\right)=a(C)\cdot f(x)+b(C). \tag{25}$$

2°. Since the function $f(x)$ is continuous and 1-1, it is monotonic. It is known that every monotonic function from real number to real numbers is almost everywhere differentiable. Let $x_0\in(0,1)$ be a point at which the function $f(x)$ is differentiable. Then, for every $x_1\in(0,1)$, by using the formula (25) with the value C for which

$$\frac{x_0}{C\cdot x_0+(1-C)}=x_1,$$

we can conclude that the function f is differentiable at this point as well. Thus, the function $f(x)$ is differentiable for all x.

Let us prove that the functions $a(C)$ and $b(C)$ are also differentiable.

2.1°. Let us start by proving that the function $a(C)$ is differentiable.

Indeed, let us take any two values x_1 and x_2 for which $f(x_1)\neq f(x_2)$. Substituting $x=x_1$ and $x=x_2$ into the formula (8), we conclude that

$$f\left(\frac{x_1}{C\cdot x_1+(1-C)}\right)=a(C)\cdot f(x_1)+b(C); \tag{26}$$

$$f\left(\frac{x_2}{C\cdot x_1+(1-C)}\right)=a(C)\cdot f(x_2)+b(C). \tag{27}$$

Subtracting (27) from (26), we conclude that

$$f\left(\frac{x_1}{C\cdot x_1+(1-C)}\right)-f\left(\frac{x_1}{C\cdot x_2+(1-C)}\right)=a(C)\cdot(f(x_1)-f(x_2)), \tag{28}$$

hence

$$a(C)=\frac{f\left(\frac{x_1}{C\cdot x_1+(1-C)}\right)-f\left(\frac{x_1}{C\cdot x_2+(1-C)}\right)}{f(x_1)-f(x_2)}. \tag{29}$$

Since the function $f(x)$ is differentiable, the right-hand side of (29) is differentiable as well, hence the function $a(C)$ is differentiable.

2.2°. Let us now prove that the function $b(C)$ is also differentiable.

Indeed, from the formula (26), we conclude that

$$b(C)=f\left(\frac{x_1}{C\cdot x_1+(1-C)}\right)-a(C)\cdot f(x_1). \tag{30}$$

Since the function $f(x)$ is differentiable, and the function $a(C)$ has also been proven to be differentiable, we thus conclude that the right-hand side of the equality (30) is differentiable as well, hence the function $b(C)$ is also differentiable.

3°. Now that we know that all three functions $f(x)$, $a(C)$, and $b(C)$ are differentiable, we can differentiate both side of the equality (25) by C and take $C=0$. As a result, we get the following equation

$$x\cdot(1-x)\cdot\frac{df}{dx}=a\cdot f+b, \tag{31}$$

where we denoted $a \stackrel{\text{def}}{=} \dfrac{da}{dC}_{|C=0}$ and $b \stackrel{\text{def}}{=} \dfrac{db}{dC}_{|C=0}$.

By moving all the terms related to x to one side of this equation and all the terms related to f to the other side, we get the following equation:

$$\frac{df}{a \cdot f + b} = \frac{dx}{x \cdot (1 - x)} = \frac{dx}{x} + \frac{dx}{1 - x}. \tag{32}$$

To solve this equation, let us consider two possible case: when $a \neq 0$ and when $a = 0$.

3.1°. When $a \neq 0$, then $a \cdot f + b$ can be represented as

$$a \cdot \left(f + \frac{b}{a}\right),$$

thus (32) takes the form

$$\frac{1}{a} \cdot \frac{df_1}{f_1} = \frac{dx}{x} + \frac{dx}{1 - x}, \tag{33}$$

where we denoted

$$f_1(x) \stackrel{\text{def}}{=} f(x) + \frac{b}{a}.$$

Integrating both sides of the equation (33), we conclude that

$$\frac{1}{a} \cdot \ln(f_1) = \ln\left(\frac{x}{1 - x}\right) + C \tag{34}$$

for some constant C, hence

$$\ln(f_1) = a \cdot \ln\left(\frac{x}{1 - x}\right) + C_1, \tag{35}$$

where $C_1 \stackrel{\text{def}}{=} a \cdot C$.

Applying $\exp(z)$ to both sides of this equality, we conclude that

$$f_1(x) = \exp\left(a \cdot \ln\left(\frac{x}{1 - x}\right) + C_1\right) = C_2 \cdot \left(\frac{x}{1 - x}\right)^a, \tag{36}$$

where $C_2 \stackrel{\text{def}}{=} \exp(C_1)$. Thus,

$$f(x) = f_1(x) - \frac{b}{a} = C_2 \cdot \left(\frac{x}{1 - x}\right)^a - \frac{b}{a}. \tag{37}$$

Since $x > 0$, the value $f(x)$ are always greater than or equal to $-b/a$, and thus, cannot take any real value smaller than $-b/a$.

This contradicts to our definition of a $(0, 1)$-transformation. Therefore, the case $a \neq 0$ is impossible.

3.2°. Let us now consider the only remaining case $a = 0$. In this case, the equation (32) takes the form

$$\frac{1}{b} \cdot df = \frac{dx}{x} + \frac{x}{1 - x}. \tag{38}$$

Integrating both sides of the equation (38), we conclude that

$$\frac{1}{b} \cdot f = \ln\left(\frac{x}{1 - x}\right) + C \tag{39}$$

for some constant C, hence

$$f(x) = b \cdot \ln\left(\frac{x}{1 - x}\right) + b \cdot C. \tag{40}$$

Thus, the function $f(x)$ is indeed equivalent to the logistic function

$$\ln\left(\frac{x}{1 - x}\right).$$

The theorem is proven.

Acknowledgments

This work was supported by a training fellowship from the National Library of Medicine to the Keck Center for Interdisciplinary Bioscience Training of the Gulf Coast Consortia (NLM Grant No. T15 LM007093).

References

[1] Feynman, R., Leighton, R., and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[2] Gelman, A., Carlin, J.B., Stern, H.S., and D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC: Boca Raton, Florida, 2004.

[3] Gelman, A., and D.B. Rubin, Inference from iterative simulations using multiple sequences, *Statistical Science*, vol.7, pp.457–472, 1992.

[4] Singer, I.M., and S. Sternberg, Infinite groups of Lie and Cartan, Part 1, *Journal d'Analyse Mathematique*, vol.XV, pp.1–113, 1965.

[5] Wiener, N., *Cybernetics, or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, Massachesetts, 1962.