Advantages of $q$-logarithm representation over $q$-exponential representation from the sense of scale and shift on nonlinear systems

Hiroki Suyari$^{1,a}$, Hiroshi Matsuzoe$^2$, and Antonio M. Scarfone$^3$

$^1$ Graduate School of Engineering, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
$^2$ Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
$^3$ Istituto dei Sistemi Complessi (ISC-CNR) c/o, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

Received 16 September 2019 / Received in final form 13 December 2019 / Published online 12 March 2020

Abstract. Addition and subtraction of observed values can be computed under the obvious and implicit assumption that the scale unit of measurement should be the same for all arguments, which is valid even for any nonlinear systems. This paper starts with the distinction between exponential and non-exponential family in the sense of the scale unit of measurement. In the simplest nonlinear model $dy/dx = y^q$, it is shown how typical effects such as rescaling and shift emerge in the nonlinear systems and affect observed data. Based on the present results, the two representations, namely the $q$-exponential and the $q$-logarithm ones, are proposed. The former is for rescaling, the latter for unified understanding with a fixed scale unit. As applications of these representations, the corresponding entropy and the general probability expression for unified understanding with a fixed scale unit are presented. For the theoretical study of nonlinear systems, $q$-logarithm representation is shown to have significant advantages over $q$-exponential representation.

1 Introduction

In Boltzmann-Gibbs- Shannon theory, i.e., the standard statistical mechanics [1,2] and information theory [3], most of the important probability distributions such as canonical distribution, Gaussian distribution, and probability for optimal code length belong to the so-called exponential family [4]. The distributions in the exponential family follow the exponential law:

$$\exp (x) \exp (a) = \exp (x + a), \quad \exp (x) / \exp (a) = \exp (x - a)$$

which play significant roles in every computation within this family. This law represents the operation by the shift in each argument, which means that multiplication

$^a$ e-mail: suyari@faculty.chiba-u.jp
and division in the exponential family is just given by plus and minus shift in arguments, respectively:

\[ x \mapsto x + a, \quad x \mapsto x - a. \] (2)

On the other hand, if we consider a power-law distribution out of the exponential family, such shift operations in multiplication and division disappear:

\[ x^{-\gamma}a^{-\gamma} = (xa)^{-\gamma}, \quad x^{-\gamma}/a^{-\gamma} = (x/a)^{-\gamma}. \] (3)

Instead, rescaling is emerging:

\[ x \mapsto xa, \quad x \mapsto x/a. \] (4)

Let us compare shift (2) and rescaling (4) from the sense of the scale unit of measurement in the following example. Consider a situation in which there are two rulers with two different scale units to measure a length on \( \mathbb{R} \) (see Fig. 1).

For a given length, one ruler (ruler 1) indicates 3 meters and the other (ruler 2) 2 meters when we measure it with these two different rulers. In this example, the units of measurement (e.g., meter) are the same, but the scale units of measurement of these two rulers are different from each other. Of course, if we use the correct ruler, we obtain the correct length. However, the correct scale unit of measurement is determined by humans, and nature does not depend on kinds of rulers. Then, in the shift (2) \( x \) and \( a \) must have the same scale units of measurement, so that the computations \( x + a \) and \( x - a \) can be done. Thus the scale unit of measurement must be invariant over addition and subtraction. In the exponential family, any multiplication and division can be done under the obvious invariance of the scale unit of measurement in any argument. But in the rescaling (4) the scale unit of measurement is not invariant if \( a \neq 1 \).

In general, scale units variant observation can be found in nonlinear dynamics with rescaling. Especially, in sequential observations, each observation ideally should have the same scale unit of measurement to deal with data in science or engineering. Thus, the assumption of independence among observations is the most ideal, which does not yield scale change in each observation. The invariance of scale unit of measurement is captured by the functions such as the probability distributions in the exponential family. However, some correlations due to rescaling can often be observed in nonlinear systems, which leads to one of the reasons for the emergence of power-law distributions far from exponential one.

In order to find a unified understanding of these two operations (shift (2) and rescaling (4)) in the simplest way, we go back to the foundation: the simplest nonlinear generalization characterizing the exponential function

\[ \frac{dy}{dx} = y^q. \] (5)
The choice of the starting point (5) in the present work originates from two aspects: statistical physics and mathematics. In statistical physics, especially for generalization of Boltzmann-Gibbs statistics, (5) is the basis for sensitivity to initial conditions, relaxation time, and stationary state (see [5] for details). In the mathematical sense, (5) recovers the famous characterization of \( \exp(x) \) for the shift (2) when \( q \to 1 \). Moreover, (5) is expected to have the rescaling (4) due to the nonlinearity when \( q \neq 1 \).

2 Scale unit of measurement in the nonlinear systems

2.1 Scale unit of measurement, inevitably determined by the initial condition

Obviously, (5) is a nonlinear differential equation with respect to \( y \). But, if the following generalized logarithm, the so-called \( q \)-logarithm defined by

\[
\ln_q y := \int_1^y \frac{1}{v^q} dv = \frac{y^{1-q} - 1}{1-q}
\]

(6)
is employed, (5) is reformed to a linear differential equation with respect to \( \ln_q y \).

\[
\frac{d \ln_q y}{dx} = 1 \quad \text{i.e.,} \quad \ln_q y = x + \ln_q C_0.
\]

(7)

Here \( C_0 \) is a positive real number determined by

\[
\ln_q C_0 = \ln_q y_0 - x_0
\]

(8)

for an initial condition \( (x_0, y_0 (> 0)) \) in (5).

Equation (7) is reformed to

\[
y = \frac{y}{C_0} = \exp_q \left( \frac{x}{C_0^{1-q}} \right),
\]

(9)

where

\[
\exp_q x := [1 + (1-q) x]^{\frac{1}{1-q}}
\]

(10)

for \( 1 + (1-q) x > 0 \), which is the inverse function of \( \ln_q x \) and is called \( q \)-exponential function.

Therefore, for the rescaling:

\[
\tilde{y} := \frac{y}{C_0}, \quad \tilde{x} := \frac{x}{C_0^{1-q}},
\]

(11)

(9) is rewritten as

\[
\tilde{y} = \exp_q (\tilde{x}).
\]

(12)
This means that the nonlinear differential equation (5) is invariant under the rescaling (11), i.e.,

$$\frac{d\tilde{y}}{d\tilde{x}} = \tilde{y}^q.$$  

**Proposition 1 (rescaling).** The nonlinear differential equation (5) is invariant under the rescaling (11).

The rescaling factor $C_0$ is determined by an initial condition $(x_0, y_0 (> 0))$ in (5) with $\ln_q C_0 = \ln_q y_0 - x_0$ (see (8)), which implies that $C_0$ can be taken as any positive real number. In other words, an initial condition $(x_0, y_0 (> 0))$ determines the scale unit of measurement in (5).

Then, in (9) the elementary scale unit “1” of observed value appears as a unit in the argument of the $q$-exponential function $\exp_q(x)$ such that $x/C_0^{1-q} = 1$, i.e., $x = C_0^{1-q}$. For a different initial condition $(x_1, y_1 (> 0))$ with $x_0 \neq x_1$ and $\ln_q C_1 = \ln_q y_1 - x_1$, $x = C_1^{1-q}$ is similarly obtained as its elementary scale unit “1” of observed value. When $q = 1$, the elementary scale unit “1” of observed value always appears as $x = 1$ (of course!) which does not depend on the initial condition of the corresponding differential equation. However, as shown above, in the nonlinear dynamics governed by (5), the scale unit of observed value inevitably depends on the initial condition. Therefore, when $q \neq 1$, the usual normalization for probability depends on the scaling effect on observed value ($x$-axis), so that the normalization in the case $q \neq 1$ should be very careful, as discussed in detail in the last section.

For example, graphs of $\frac{y}{C} = \exp_q \left( -\frac{x}{C^{1-q}} \right)$ for $C = 1, 10, 20$ are described in the left three figures of Figure 2. The shapes of the left three graphs in Figure 2 are completely the same, but each scale unit of measurement is different with each other due to the rescaling $C$ (see both $x$-axis and $y$-axis in the left three graphs of Fig. 2). This means under the rescaling (11) the graph of $y = \exp_q (-x)$ is invariant.
The invariance under the rescaling (11) is confirmed by the same slope of $\ln_q y = -x + \ln_q C$ (see the rightmost graph in Fig. 2).

2.2 Scale unit of measurement, inevitably changed by shift

In the previous subsection, for a given nonlinear differential equation (5) a rescaling (11) in both $x$ and $y$ arguments inevitably appears. More precisely, the nonlinear differential equation (5) is invariant under the rescaling (11) (see (5), (11), and (13)). Such a rescaling can appear without using the nonlinear differential equation (5), that is, shift in argument. For a given $y = \exp_q (x)$, if we apply a shift $x \mapsto x + c$ to this equation, we obtain

$$y = \exp_q (x + c) = \exp_q (c) \exp_q \left( \frac{x}{\left( \exp_q (c) \right)^{1-q}} \right),$$

that is,

$$\frac{y}{\exp_q (c)} = \exp_q \left( \frac{x}{\left( \exp_q (c) \right)^{1-q}} \right).$$

(15)

Thus, by the rescaling:

$$y' := \frac{y}{\exp_q (c)}, \quad x' := \frac{x}{\left( \exp_q (c) \right)^{1-q}},$$

(16)

we obtain

$$y' = \exp_q (x').$$

(17)

This means that $y = \exp_q (x)$ is invariant under a shift $x \mapsto x + c$ in argument $x$, which yields the same rescaling as (11).

These two operations rescaling and shift in $y = \exp_q (x)$ are equivalent to each other. In fact, for a given rescaling such as (16) we obtain $y = \exp_q (x + c)$ which is a shift $x \mapsto x + c$ in $x$-argument of $y = \exp_q (x)$. On the other hand, for a given shift such as (14) we can get a rescaling (16).

Proposition 2 (shift and rescaling). A shift $x \mapsto x + c$ to $y = \exp_q (x)$ for any $c \in \mathbb{R}$ satisfying $1 + (1 - q) c > 0$ is equivalent to a rescaling in both $x$-axis and $y$-axis.

Shift in the argument of the $q$-exponential function results in various scale units of measurement in sequential observations. According to the property of the $q$-exponential function:

$$\exp_q (x_1 + \cdots + x_n) = \exp_q (x_1) \cdots \exp_q \left( \frac{x_n}{1 + (1 - q) \sum_{i=1}^{n-1} x_i} \right),$$

(18)

$x_1, \ldots, x_n$ on the left side must have a same scale unit of measurement, so that the sum $x_1 + \cdots + x_n$ can be computed. On the other hand, we get the observed values
\[ x_1', \ldots, x_n' \text{ on } \mathbb{R} \text{ (i.e., } (x_1', \ldots, x_n') \in \mathbb{R}^n) \text{ with different scale units of measurement such as} \]
\[
x_1' = x_1, \quad x_2' = \frac{x_2}{1 + (1 - q) x_1}, \ldots, \quad x_n' = \frac{x_n}{1 + (1 - q) \sum_{i=1}^{n-1} x_i}. \tag{19}
\]

Recall that \( x_1, \ldots, x_n \) have the same scale unit, so that observed values \( x_1', \ldots, x_n' \) have different scale units if \( q \neq 1 \). This representation is due to the property of the \( q \)-exponential \((10)\).

As shown in the study of the dynamics determined by \((5)\), there exist two representations, namely the \( q \)-exponential representation and the \( q \)-logarithm representation. The choice of these two representations depends on what we want to express. \( q \)-Exponential representation is useful for rescaling, while \( q \)-logarithm representation for unified studies with a fixed scale unit of measurement.

Note that “unified studies” in the \( q \)-logarithm representations mean that it is possible to study the dynamics with a fixed scale unit of measurement. On the other hand, in the \( q \)-exponential representations such as \((18)\), observed values \( x_1', \ldots, x_n' \) in \((19)\) have different scale unit of measurement, which makes the unified studies difficult in general.

3 Two representations in the systems determined by the fundamental nonlinear differential equation

3.1 \( q \)-Exponential representation for rescaling

If we want to represent a rescaling effect in our formulations, \( q \)-exponential representation such as \((9)\) and \((18)\) is more useful than the corresponding \( q \)-logarithm representation given in \((7)\). In fact, \( q \)-exponential representation reveals how each variable in the formulation is rescaled by other variables or constants (e.g., \((18)\)).

But there are some disadvantages to using \( q \)-exponential representation. One of these is the appearance of complicated rescaling in sequential observations. For a given \( q \)-exponential representation \( y = \exp_q (x) \), a shift in \( x \) such that \( x \mapsto x + c_1 \) is applied to this \( q \)-exponential representation. Then, in the same way as \((14)\) we obtain

\[
\frac{y}{\exp_q (c_1)} = \exp_q \left( \frac{x}{(\exp_q (c_1))^{1-q}} \right). \tag{20}
\]

Again, one more shift in the argument of \( q \)-exponential function is applied to this expression \((20)\), then we can get

\[
\frac{y}{\exp_q (c_1) \cdot \exp_q (c_2)} = \exp_q \left( \frac{x}{(\exp_q (c_1))^{1-q} (\exp_q (c_2))^{1-q}} \right). \tag{21}
\]

Note that a shift by \( c_1 \) is different from that by \( c_2 \) in the sense of scale unit. More concretely, a shift by \( c_1 \) is given by \( x \mapsto x + c_1 \), but a shift by \( c_2 \) is given by \( x/(\exp_q (c_1))^{1-q} \mapsto x/(\exp_q (c_1))^{1-q} + c_2 \). Then, scale unit of shift \( c_1 \) is different from that of \( c_2 \).

Here we need to make some comments on the \( q \)-product \([6,7]\). As discussed in the previous section, \( x_1, \ldots, x_n \) on the left side of \((18)\) must have the same scale unit of the measurement, but for the observed values \( x_1', \ldots, x_n' \) appeared on the right
side does not so. In particular, each \( x'_t \) has different scale unit of the measurement by rescaling with past internal values (often called “state variables” in control theory) \( x_{t-1}, x_{t-2}, \ldots \), which makes theoretical analysis difficult. In order to avoid these difficulties, the \( q \)-product is useful in many applications [8–10]. The \( q \)-product \( \otimes_q \) is introduced to satisfy

\[
\exp_q (x_1 + x_2) = \exp_q (x_1) \otimes_q \exp_q (x_2)
\]

as a generalization of the exponential law [6,7]. Then, the property (18) can be rewritten by means of the \( q \)-product.

\[
\exp_q (x_1 + x_2 + \cdots + x_n) = \exp_q (x_1) \otimes_q \exp_q (x_2) \otimes_q \cdots \otimes_q \exp_q (x_n).
\]

Therefore, the \( q \)-product preserves scale unit of measurement among \( x_1, \ldots, x_n \), so that there are a lot of successful applications in this field [8]. But at the same time, there are some disadvantages to use the \( q \)-product as shown below.

One of some disadvantages using the \( q \)-product is as follows: From the requirement (22), the definition of the \( q \)-product \( \otimes_q \) is given by

\[
x \otimes_q y := [x^{1-q} + y^{1-q} - 1]^{\frac{1}{1-q}},
\]

which is valid only under the constraints \( x, y > 0 \) and \( x^{1-q} + y^{1-q} - 1 > 0 \). In each computation by means of \( q \)-product or \( q \)-ratio (inverse operation of the \( q \)-product), it should be confirmed if these constraints are satisfied or not. Another disadvantage is that there is no room to employ a scaling effect \( C \) in the formulations using the \( q \)-product. Of course, a scaling effect \( C \) can be added in ad hoc such that \( y = \exp_q (x) \otimes_q \exp_q (C) \), but this expression does not show a rescaling effect in arguments.

### 3.2 \( q \)-Logarithm representation for unified studies with a fixed scale unit of measurement

As shown in (9) and (11), a scaling factor \( C_0 \) (i.e., initial condition) affects significantly on observed data in the nonlinear dynamics. In the dynamics governed by the fundamental nonlinear differential equation (5) the scaling factor \( C_0 \) is determined by the initial condition (8) and inevitably appears in (7) or (9). If \( q \)-exponential representation is used in formulations such as (9) and (18), a scaling factor \( C_0 \) appears in every argument (e.g., both sides in (9) and \( x_1 (= \ln_q C_0) \) on the right side of (18)). This strong dependence of \( C_0 \) on each argument yields serious difficulties in analysis and understanding. However, in \( q \)-logarithm representation such as (7) (the origin of (9)), a scaling factor \( C_0 \) appears only one time in one formula which has a lot of advantages over \( q \)-exponential representation. For example, in (7), a shift in \( x \) like \( x \rightarrow x + c \) is described by just a shift of a graph on a \( x-q \)-log plot.

Moreover, in \( q \)-logarithm representation such as (7), all arguments have the same scale unit of measurement. On the other hand, in the \( q \)-exponential representation (9), scale units of \( x \) and \( x/C_0^{1-q} \) are obviously different with each other. Thus, \( q \)-logarithm representation has an important advantage over \( q \)-exponential representation in the sense of scale unit.
4 Application of q-logarithm representation

4.1 Rederivation of Tsallis entropy via q-logarithm representation

In [8], q-product (24) is applied to the derivation of Tsallis entropy as the unique entropy corresponding to the fundamental nonlinear differential equation (5). For the following discussion, let us briefly review how several formulations such as q-Stirling’s formula and Tsallis entropy can be uniquely obtained from the fundamental nonlinear differential equation (5) with some modifications of the original version [8]. The distinction from the original derivation is that the q-product is not explicitly used to avoid some difficulties stated in the previous section.

For any natural number $n \in \mathbb{N}$, the q-logarithm of the q-factorial is introduced:

$$\ln_q n!_q := \sum_{k=1}^{n} \ln_q k.$$  \hspace{1cm} (25)

Then, for large $n \in \mathbb{N}$ we can get the q-Stirling’s formula:

$$\ln_q n!_q \simeq \begin{cases} \frac{n}{2-q} \ln_q n - \frac{n}{2-q} + \frac{1}{2} \ln_q n + \frac{1}{2} \frac{1}{2-q}, & (q \neq 2) \\ n - \ln n - \frac{1}{2n} - \frac{1}{2}, & (q = 2). \end{cases} \hspace{1cm} (26)$$

By means of (25), the q-logarithm of the q-multinomial coefficient is defined by

$$\ln_q \left[ \begin{array}{c} n_1 \times \cdots \times n_k \\ n_1 \cdots n_k \end{array} \right]_q := \ln_q n!_q - \ln_q n!_q - \cdots - \ln_q n!_q,$$  \hspace{1cm} (27)

where

$$n = \sum_{i=1}^{k} n_i, \hspace{0.5cm} n_i \in \mathbb{N} \hspace{0.1cm} (i = 1, \ldots, k). \hspace{1cm} (28)$$

Note these definitions (25) and (27) hold for any natural number $n \in \mathbb{N}$. Then, we apply the q-Stirling’s formula in (27) which uniquely leads to

$$\ln_q \left[ \begin{array}{c} n_1 \times \cdots \times n_k \\ n_1 \cdots n_k \end{array} \right]_q \simeq \begin{cases} \frac{n^{2-q}}{2-q} \cdot S_{2-q}^{\text{Tsallis}} \left( \frac{n_1}{n}, \ldots, \frac{n_k}{n} \right), & (q \neq 2) \\ -S_1^{\text{Tsallis}} (n) + \sum_{i=1}^{k} S_1^{\text{Tsallis}} (n_i), & (q = 2). \end{cases} \hspace{1cm} (29)$$

where $S_q^{\text{Tsallis}}$ is Tsallis entropy [11] defined by $S_q^{\text{Tsallis}} := \left( 1 - \sum_{i=1}^{k} p_i^q \right) / (q - 1)$ and $S_1^{\text{Tsallis}} (n) := \ln n$. This is a straightforward derivation of Tsallis entropy from the fundamental nonlinear differential equation (5).

4.2 Reformulation of q-Gaussian distribution with scale invariance

There are several important probability distributions associated with Tsallis entropy such as a q-canonical distribution and a q-Gaussian distribution. In this section, we derive the q-logarithm representation of the q-Gaussian distribution for unified
studies with a fixed scale unit of measurement. There are several ways to derive a $q$-Gaussian distribution [9]. The simplest way is the Maximum Likelihood Principle (MLP for short) [13]. In the course of the derivation of $q$-Gaussian distribution in the MLP, $q$-logarithm representation including a scaling factor $C$ is naturally appears.

Here $n$ observed values $x'_1, x'_2, \ldots, x'_n \in \mathbb{R}$ are given, but these values do not have the same scale unit. Instead, there exist

$$x_1, x_2, \ldots, x_n \in \mathbb{R}$$

with a same scale unit. Each $x_i \in \mathbb{R}$ corresponds to each $x'_i \in \mathbb{R}$ ($i = 1, \ldots, n$), respectively (e.g., (19)). The $q$-logarithm likelihood function $\log_q L_q (\theta)$ is defined by

$$\log_q L_q (\theta) := \sum_{i=1}^{n} \log_q f (x_i - \theta),$$

where $\theta$ is a variable for this function $L_q$ and $f$ is a probability density function with $x_i - \theta$ as a value of its corresponding random variable.

If the function $\log_q L_q (\theta)$ of $\theta$ for any fixed $x_1, x_2, \ldots, x_n$ attains the maximum value at

$$\theta = \theta^* := \frac{x_1 + x_2 + \cdots + x_n}{n},$$

the probability density function $f$ must be a $q$-Gaussian:

$$f (e) = \frac{\exp_q (-\beta_q e^2)}{\int \exp_q (-\beta_q e^2) \, de},$$

where $\beta_q$ is a $q$-dependent positive constant.

See [13] for the detailed proof. Note that the requirement (32) means that the scale units of $x_i$ ($i = 1, \ldots, n$) should be the same among them so that this addition can be computed.

In the course of the proof [13], the following differential equation is derived from the requirement of the theorem

$$\frac{f' (e)}{(f (e))^q} = a_q e,$$

where $a_q \in \mathbb{R}$. Equation (34) can be integrated with respect to $e$:

$$\ln_q f (e) = \frac{a_q e^2}{2} + C_q,$$

where $C_q$ is a $q$-dependent integration constant. This expression (35) is obviously $q$-logarithm representation. If $1 + (1 - q) \left( a_q e^2 / 2 + C_q \right) > 0$ and $1 + (1 - q) C_q > 0$, then we obtain a $q$-Gaussian probability density function (33) with $\beta_q := -a_q / (2 (1 + (1 - q) C_q)) > 0$. Within constraints on $C_q$, the arbitrariness of an integration constant $C_q$ still remains.

Note that the final expression (33) is clearly $q$-exponential representation and in this expression, $C_q$ is included in both denominator and numerator of (33).

In order to see a rescaling effect in the final expression (33), the corresponding frequency distribution can be obtained as follows.
Let $\gamma_q$ be defined by $\gamma_q := -a_q/2$. Then (35) is rewritten as

$$\ln_q f(e) = -\gamma_q e^2 + C_q.$$ (36)

Hence, we obtain

$$\frac{f(e)}{c} = \exp_q \left(-\gamma_q \left(\frac{e}{c^{1/2}}\right)^2\right),$$ (37)

where $c := \exp_q (C_q) > 0$. $f(e)$ is the probability density function, so the left side $f(e)/c$ is no longer a probability density function. But $(f(e)/c) \Delta e$ represents frequency distribution which has scale invariance due to arbitrariness of $c$. Obviously, under the rescaling:

$$\tilde{f}(\tilde{e}) := \frac{f(e)}{c}, \quad \tilde{e} := \frac{e}{c^{1/2}},$$

(38)

(37) is rewritten as

$$\tilde{f}(\tilde{e}) = \exp_q (-\gamma_q \tilde{e}^2).$$ (39)

This also represents invariance of the frequency distribution (37) under the rescaling (38) on both $e$-axis and $f(e)$-axis.

Here, for simplicity and easy understanding, we set $y := f(e), x := e$. The graphs of $\frac{y}{c} = \exp_q \left(-\left(\frac{x}{c^{1/2}}\right)^2\right)$ for $c = 1, 10, 100$ and $q = 1.7$ are described in Figure 3.
The shapes of the graphs for $c = 1, 10, 100$ in Figure 3 are completely the same, but each scale unit of measurement is different with each other due to the rescaling $c$. This means under the rescaling (38) the graph of $y = \exp_q(-x^2)$ is invariant. The distribution (37) can be easily transformed into a probability distribution by imposing a normalization depending on each scale.

Note that when $q = 1$ the scale unit on $x$-axis is fixed for any cases (see also (38)) and a scaling in (38) is appeared on $y$-axis only, which is applied to normalization in probability distributions. In other words, when $q \neq 1$, normalization should be very careful, because usual normalization depends on scale unit on $x$-axis [14].

5 Advantages of $q$-logarithm representation over $q$-exponential representation through a concrete example

In the previous two sections, $q$-exponential representation and $q$-logarithm representation have different purposes of expressing. The former is for rescaling, the latter for unified understanding with a fixed scale unit. However, for the theoretical studies including computer simulations, $q$-logarithm representation has some crucial advantages over $q$-exponential representation. In particular, the non-uniqueness problems in $q$-exponential representation is always appeared in a formulation of a probability distribution. Through the following general example, the non-uniqueness is concretely shown. For simplicity and ease of understanding, we present the case of a discrete distribution. The case of a continuous distribution is similarly discussed. After this example, the solution for this non-uniqueness problem is given by the $q$-logarithm representation as a unique expression.

Consider the following situation such that a frequency $n_i \in \mathbb{N}$ of data $x_i$ is given by

$$n_i = \exp_q(-x_i + c), \quad (i = 1, \cdots, k), \quad (40)$$

where $c$ is a constant. Let the total frequency $n$ be defined by $n := \sum_{i=1}^{k} n_i$. Then,

$$n = \sum_{i=1}^{k} n_i = \sum_{i=1}^{k} \exp_q(-x_i + c). \quad (41)$$

We want to find a probability distribution $\{p_i\}$ for these data, so we can compute

$$p_i := \frac{n_i}{n} = \frac{\exp_q(-x_i + c)}{\sum_{i=1}^{k} \exp_q(-x_i + c)}. \quad (42)$$

When $q = 1$,

$$p_i = \frac{\exp(-x_i)}{\sum_{i=1}^{k} \exp(-x_i)} \quad \text{which does not depend on } c \text{ and is the unique expression using the only observed value } x_i. \text{ However, when } q \neq 1, \text{ innumerably many equivalent representations for probability distribution (42) can be acceptable. For example, for the case } c = c_1 + c_2$$
\( (c_1 \neq c_2) \) we have
\[
p_i = \frac{\exp_q (-x_i + c_1 + c_2)}{\sum_{i=1}^{k} \exp_q (-x_i + c_1 + c_2)}. \tag{44}
\]

We rewrite \( \exp_q (-x_i + c_1 + c_2) \) in the two kinds of representations
\[
\exp_q (-x_i + c_1 + c_2) = \exp_q (c_1) \exp_q (-x_i + c_2) = \exp_q (c_2) \exp_q (-x_i + c_1).
\tag{45}
\]

Therefore, \( p_i \) in (42) is given in the two possible ways:
\[
p_i = \frac{\exp_q \left( \frac{-x_i + c_2}{\exp_q (c_1)^{1-q}} \right)}{\sum_{i=1}^{k} \exp_q \left( \frac{-x_i + c_2}{\exp_q (c_1)^{1-q}} \right)} = \frac{\exp_q \left( \frac{-x_i + c_1}{\exp_q (c_2)^{1-q}} \right)}{\sum_{i=1}^{k} \exp_q \left( \frac{-x_i + c_1}{\exp_q (c_2)^{1-q}} \right)}. \tag{46}
\]

Of course, innumerably many choices of \( c_1 \) to satisfy \( c = c_1 + c_2 \) are available. Even for the simple representation (42), there exist very many equivalent representations of a probability distribution. This is due to arbitrary selection of rescaling and shift for the observed values (see (46)). These non-unique representations such as (46) comes from the fact that the nonlinear system (5) is invariant for any rescaling and shift of observed values \( x_i \).

Therefore, \( q \)-exponential representation as probability distribution is not unique, in general. In order to avoid the non-uniqueness of \( q \)-exponential representation, \( q \)-logarithm representation should be used for probability distribution. From (40),
\[
\ln_q n_i = -x_i + c. \tag{47}
\]

Hence, after some computations, we obtain
\[
\ln_q p_i = -n^{q-1}x_i + \left( n^{q-1}c - \ln_{2-q} n \right), \tag{48}
\]
where we used
\[
\ln_q \frac{y}{x} = x^{q-1} (\ln_q y - \ln_q x). \tag{49}
\]

The \( q \)-logarithm representation (48) is obviously unique except for \( c \). For example, in case \( c = c_1 + c_2 \) as stated above, the expression (48) is invariant.

Therefore, \( q \)-logarithm representation should be used for probability distribution instead of \( q \)-exponential representation in order to avoid non-uniqueness. Recently, this non-uniqueness problem is also discussed in [14] from the information geometrical points of view.

6 Conclusion

Long range correlations and past- or history- dependence have been studied for many years in both linear and nonlinear systems [15]. In this paper, from the sense of
the scale unit of measurement, we analytically discuss how each observed data in a nonlinear system has received influence on scale from other data on the simplest model determined by the fundamental nonlinear differential equation (5). Any correlation among observed data on the dynamics (5) is purely due to rescaling by the previous data, which yields different scale unit of measurement. This rescaling is found to be equivalent to shift in the argument of the dynamics (5). These effects such as rescaling and shift result in long range correlations among the data. In order to avoid different scale units on data, a corresponding logarithm (e.g., $q$-logarithm) representation is shown to have some crucial advantages such as uniqueness over a corresponding exponential representation. These results can be applied to many studies in nonlinear systems.

The first author is grateful to his son for asking the author to solve the arithmetic problem about the two rulers with different scale units, which inspired the first author to find the present idea. The first author is also grateful to Jan Korbel and Atsumi Ohara for their careful reading and comments on the first draft. This work was supported by JSPS KAKENHI Grant Number 17K19957 in Japan.

References

1. R.C. Tolman, *The Principles of Statistical Mechanics* (Dover, 2010)
2. W. Greiner, *Thermodynamics and Statistical Mechanics* (Springer, 2010)
3. T.M. Cover, J.A. Thomas, *Elements of information theory*, 2nd edn (Wiley, 2006).
4. F. Nielsen, V. Garcia, arXiv:0911.4863 (2009)
5. C. Tsallis, Physica D 193, 3 (2004)
6. L. Nivanen, A. Le Mehautem Q.A. Wang, Rep. Math. Phys. 52, 437 (2003)
7. E.P. Borges, Physica A 340, 95 (2004)
8. H. Suyari, Physica A, 368, 63 (2006)
9. C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer, 2009)
10. A.M. Scarfone, Entropy 15, 624 (2013)
11. C. Tsallis, J. Stat. Phys. 52, 479 (1988)
12. A. Einstein, Ann. Phys. 33, 1275 (1910)
13. H. Suyari, M. Tsukada, IEEE Trans. Inform. Theory 51, 753 (2005)
14. H. Matsuzoe, A.M. Scarfone, T. Wada, LNCS 11712, 279 (2019)
15. S. Thurner, R. Hanel, P. Klimek, *Introduction to the Theory of Complex Systems* (Oxford Univ. Press, 2018)