On a simple derivation of a family of nonextensive entropies from information content

Takuya Yamano

Department of Physics, Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan E-mail: yamano@cosmos.phys.ocha.ac.jp

Received: 1 June 2004 / Accepted: 31 July 2004 / Published: 1 August 2004

Abstract:

The nonextensive entropy of Tsallis can be seen as a consequence of postulates on a self-information, i.e., the constant ratio of the first derivative of a self-information per unit probability to the curvature (second variation) of it. This constancy holds if we regard the probability distribution as the gradient of a self-information. Considering the form of the nth derivative of a self- information with keeping this constant ratio, we arrive at the general class of nonextensive entropies. Some properties on the series of entropies constructed by this picture are investigated.

Keywords: Information theory; Nonadditive entropy; Information content.

1 Introduction

Tsallis statistics[1] can be seen as a formalism based on a pair of deformed functions of usual exponential and logarithmic ones[2, 3]. Their deformed functions are called the q-exponential and the q-logarithmic functions, respectively. The q-exponential function is defined as $e_q^x \equiv [1 + (1 - q)x]_+^{1/1-q}$, where $[x]_+ = max\{x, 0\}$ and the q-logarithmic function $\ln_q x = (x^{1-q}-1)/(1-q)$, where $x \ge 0$. The derivation of the deformed (or, the generalized) form of the usual logarithmic entropy and its mathematical structure based on it have been attracted much attention. Furthermore, studies on deriving probability distribution functions consistent with these generalized entropies from microscopic dynamics have been done [4, 5] and some statistical mechanical foundations of

the Tsallis distribution are summarized in Ref.[6] (and references therein). One question arises regarding the origin of the nonextensive entropy and the clarification of it is desirable for further study. Although the Tsallis entropy was devised in physics in a heuristic way motivated by multifractals[1], the use of it has been stretched to neighboring scientific fields such as information theory [7], econophysics[8], to name a few. Therefore, it is required to understand how the entropy is connected to the existing quantities. In this paper, we shall provide a partial answer to that query from the information theoretical point of view and present a general form of entropies.

Let us review the elicitation procedure of the Tsallis entropy. It is shown that the Tsallis entropy is derived from the following set of postulates and a notation on the self-information with standard average:

Postulate 1: For $p \in [0, 1]$, h(p) is a continuous function Postulate 2: For two different probabilities p_1 and p_2 , $h(p_1p_2) = h(p_1) + h(p_2) + \lambda h(p_1)h(p_2)$, where $\lambda \neq 0$

Postulate 3 (Notation): h(0) = 1/(q-1), where $q \in \mathbb{R}$.

Since $\lambda \neq 0$, the expression in Postulate 2 can be rewritten as a factorized form $s(p_1p_2) = s(p_1)s(p_2)$ with $s(p) = 1 + \lambda h(p)$. The trivial solution s(p) = 0, which gives constant function for h(p) $(= -1/\lambda)$ is not appropriate due to the fact that it provides constant information irrespective of the associated probabilities. Therefore we may put the general solution as $s(p) = p^{q-1}[9]$, which leads to the form $h(p) = (p^{q-1} - 1)/\lambda$. From Notation, we determine $\lambda = 1 - q$, hence the linear average of h(p) (i.e., $\sum_k p_k h(p_k)$) leads to the Tsallis entropy. Since Notation fixes the value of λ , if we replace Notation with h(1/2) = 1, resulting entropy gives the type β generalization of Shannon entropy by Havrda-Charvat[10] and Daroczy[11]. In this case the factor λ becomes $2^{1-\beta} - 1$, where $\beta \neq 1$. The difference between the Tsallis entropy and the type β entropy stems from the arbitrary factor which fixes the form of the self-information. It should be noted that, in a nonadditive self-information $h(p) = (p^{q-1} - 1)/(1-q)$, we have finite information for $q \neq 1$ even if the least probable event occurs[7], whereas we get infinite information in the ordinary logarithmic definition of the self-information (see Appendix for the definition).

In view of the fact that some alternative entropies and discussions based on them are recently attracting much attention for their potential uses and applications[12, 13, 14], our motivation in this study is set to the investigation for certain types of a hierarchical structure embedded in the generalized entropies starting with a self-information, which characterizes the associated entropy. This paper presents a procedure for obtaining a series of generalized class of (nonextensive) entropies, which include some existing entropies as special cases.

In the next section, we give an elementary prescription in order to obtain a series of nonextensive entropies including Tsallis one. In Section 3, we briefly discuss properties of basic ingredients for the nonextensive entropies. We present our conclusions in the last section.

2 A method to construct a generalized class of entropies

Information content is a basic quantity for measuring information of systems considered in information theory. This notion is used for making entropies in our study. As a starting point, let us consider a probability distribution under α rescaling transformation defined on \mathbb{R} , where α $(\alpha \neq 1)$ is a positive real number. The transformed probability distribution $p_{\alpha}(x)$ is defined as $p_{\alpha}(x) = \alpha p(x/\alpha)$. In general, the value of α determines the broadness and the size of the distribution. When $\alpha < 1$, $p_{\alpha}(x)$ shrinks, whereas if $\alpha > 1$ it enlarges, however, its shape is maintained. In our present consideration, we focus on a special case with respect to this rescaling, where the probability function is invariant under the transformation:

$$\alpha p\left(\frac{x}{\alpha}\right) = p(x). \tag{1}$$

This transformation expresses that the probability distribution is a homogeneous function of degree one. It can be shown that, when the probability distribution is continuous on [0, 1], the ratio p(x)/x is found to be constant over this range. For positive integers n_i $(i = 1, \dots, 4)$ satisfying $n_1/n_2(=r_1), n_3/n_4(=r_2) \in [0, 1]$, repeated use of the equation (1) yields

$$p(n_1/n_2) = n_1 p(1/n_1 \cdot n_1/n_2) = n_1 p(1/n_2)$$

= $n_1 n_4 p(1/n_2 \cdot 1/n_4) = n_1 n_4 p((n_2 n_3)^{-1} \cdot n_3/n_4)$
= $n_1/n_2 \cdot n_4/n_3 p(n_3/n_4),$ (2)

which leads to $p(r_1)/r_1 = p(r_2)/r_2$. Therefore, the ratio p(x)/x results in a constant value. This fact constitutes a crucial basis in the following consideration. It is shown that the constant ratio of the change of the self-information per unit probability to its curvature (second derivative) leads to the Tsallis entropy. For this purpose, we add the following postulate:

Postulate 4: The ratio $\frac{h'(x)/x}{h''(x)}$ is invariant under transformation Eq.(1). This postulate implies (h'(x)/x)/h''(x) = const., in which the general solution can be expressed as $h'(x) = ax^b$ $(a, b \in \mathbb{R})$. Then, if we choose a = -1, b = q - 2, and setting the integration constant as 1/(q-1), we obtain the form $h(x) = \ln_q(x^{-1})$. This form leads to the Tsallis entropy if we use the standard mean value $\int xh(x)dx$ (If we employ the escort average[15], we have the normalized Tsallis entropy[7, 16, 17]). We immediately generalize the above procedure to the *n*th derivative of the self-information. That is,

Postulate 4': The ratio $\frac{h^{(n)}(x)/x}{h^{(n+1)}(x)}$ is invariant under transformation Eq.(1), where h(x) is a C^{n+1} class function.

The general solution of the above expression can be put as $h^{(n)}(x) = ax^{q-(n+1)} + b$. Therefore, we

obtain the solutions:

$$h(x) = \begin{cases} \sum_{k=1}^{n+1} c_{k-1} \frac{x^{k-1}}{(k-1)!}, & c_n = a+b \quad (\text{if } q = n+1) \\ \frac{ax^{q-1}}{C(q;n)} + \sum_{k=1}^{n+1} c_{k-1} \frac{x^{k-1}}{(k-1)!}, & c_n = b \quad (\text{if } q \neq n+1 \text{ and } q \neq n) \\ \frac{ax^{n-1}}{(n-1)!} \ln x + \sum_{k=1}^{n+1} c_{k-1} \frac{x^{k-1}}{(k-1)!}, & c_n = b \quad (\text{if } q = n) \end{cases}$$
(3)

where $C(q;n) = (q-1)(q-2)\cdots(q-n)$ and c_0,\ldots,c_n are integration constants. We find that the Shannon entropy (the self-information by Hartley[18]) is recovered when n = 1, a = -1, b = 1and $c_k = 0, \forall k$ in the third line of the above expression. Since the form of the cases $q \neq n+1$ and $q \neq n$ in the second line can be considered as a generic self-information, we shall consider properties of it in the next section. The values of a and c_k 's are constrained by the feature of h(x)in addition to the sign of C(q; n). Using the standard definition of the mean value, the resulting average of the self-information (i.e., entropy) can be expressed as follows:

$$\frac{a}{C(q;n)} \sum_{i} p_{i}^{q} + \sum_{i} \sum_{k=1}^{n+1} c_{k-1} \frac{p_{i}^{k}}{(k-1)!}, \quad (n \ge 1).$$
(4)

We denote this entropy as $H_q^n(p; \{c_k\})$, where a set of probabilities is denoted simply by p. Some existing entropies can be expressed as a paticular class of $H_q^n(p; \{c_k\})$: when a = -1,

- (1) the Rényi entropy of order α , $\ln \sum_{i} p_i^{\alpha}/(1-\alpha)$ is expressed as $\ln[(1-\alpha)H_{\alpha}^1(p;0,0)]/(1-\alpha)$. (2) $H_q^1(p; \frac{1}{q-1}, 0)$ is equivalent to the Tsallis entropy $\frac{\sum_i p_i^q - 1}{1-q}$. (3) the quadratic information, $1 - \sum_i p_i^2$ (e.g.,[19]) is given by $H_2^1(p; 1, 0)$.
- (4) the type β entropy is $H^1_{\beta}(p; \frac{1}{1-2^{1-\beta}}, 0)$.

If we allow a to take various values, we can express other entropies with $H_a^n(p; \{c_k\})$. For example, the cubic information $1 - \sum_i p_i^3$ [20] can be represented either by $H_3^2(p; 1, 0, 1)$ with a = -1, or $H_3^2(p; 1, 0, -2)$ with a = 0, or $H_3^1(p; 1, 0, 0)$ with a = -2.

In the subsequent discussion, we set a = -1 by taking into account that the conventional entropy is recovered in this case. If we put $a = -k_B$, where k_B is the Boltzmann factor, it can be regarded as an entropy in statistical mechanics, however, recall that we usually put $k_B = 1$ to make the entropy dimensionless.

3 Some general properties of $H_q^n(p; \{c_k\})$

There are many properties characterizing an entropy. Among these, we shall investigate ingredients to clarify the relation to the physical entropy, which is expected to have concavity[21] for instance. Our way of consideration is to determine or to give constraints between parameters included in $H_q^n(p; \{c_k\})$ so that the generalized entropies can possess the fundamental features below.

Certainty and monotonicity w.r.t the number of states:

In order to guarantee the certainty $H_q^n(1, 0, ..., 0; \{c_k\}) = 0$, we need to have the relation $C(q; n)^{-1} = \sum_{k=1}^{n+1} c_{k-1}/(k-1)! \ (a = -1)$. This provides a strong constraint on the value of c_k 's and q for a given order n. Whether or not the entropy for the equiprobability $p_i = w^{-1}$ is an increasing

function of w, in general, depends complicately on the range of value q and the values of c_k 's. As a specific example, let us mention the case n = 1. The behavior of $H^1_q(w^{-1}; c_0, c_1)$ is shown in Figure 1, where the relation that ensures the certainty condition described above has used. It



Figure 1: Entropy for a two-state system $H_q^2(p; c_0, c_1)$ as a function of p for different values of q with $c_1 = 0.7$. Note that the value of c_0 is determined from the values of q and c_1 such that $(c_0 + c_1)^{-1} = (q - 1)(q - 2)$ due to the requirements $H_q^2(0; c_0, c_1) = 0$ and $H_q^2(1; c_0, c_1) = 0$. From the value of the second derivative of the entropy at p = 1/2, it is concave when $c_1 < \frac{q}{q-1}2^{1-q}$.

increases monotonically with respect to w when 1 < q < 2 for all values of c_0 . For other regions of q, $H_q^1(w^{-1}; c_0, c_1)$ monotonically decreases as w grows. In a specific choice of the integration constants, i.e., when $c_k = 1$, $\forall k$, $dH_q^n(w^{-1}; 1, \ldots, 1)/dw$ is written as $-w^{1-q}/C(q; n) + \sum_{i=1}^n w^k/k!$ (a = -1) for $q \neq n$. This means that $H_q^n(w^{-1}; 1, \ldots, 1)$ can be an increasing function of w (when the values of n and q are chosen appropriately), which is one of the desirable behavior for an entropy at the equiprobability.

Two-state system:

One of the simple examinations of a behavior of the generalized entropy is for a two-state system (W = 2), where the probabilities is defined $p_1 = p$ and $p_2 = 1 - p$, respectively. From the definition, straightforward calculations yield a final expression:

$$H_q^1(p;c_0,c_1) = \frac{1-p^q - (1-p)^q}{(q-1)(q-2)} + 2c_1 p(p-1),$$
(5)

where we set a = -1, and used the end-points condition (i.e, the entropy should vanish at p = 0and p = 1, leading to the relation $c_0 + c_1 = [(q - 1)(q - 2)]^{-1})$. Figure 2. shows the shape of $H_q^2(p; c_0, c_1)$ for different values of q, where the end-points condition and the concavity are both satisfied. Note that the concavity of the entropy holds only for the parameter range $c_1 < \frac{q}{q-1}2^{1-q}$. *Expansibility*:



Figure 2: $H_q^2(w^{-1}; c_0, c_1)$ as a function of w for different values of q when $c_0 = 1.3$. Note that the value of c_1 is determined from the values of q and c_0 .

 $H_q^n(p; \{c_k\})$ is expansible for all q and n, i.e.,

$$H_q^n(p_1, \cdots, p_w; \{c_k\}) = H_q^n(p_1, \cdots, p_w, 0; \{c_k\}),$$
(6)

where w denotes the number of states. Regarding the entropy for uniform probabilities and certainty, we shall add comments in the next section.

Transformation from discrete entropy to continuous one:

If we define a discrete probability as $p_i = p(\bar{x}_i)\Delta x$ ($\forall i, i = 1, ..., m$), where $\bar{x}_i \in [x_{i-1}, x_i]$ and the bins of length $\Delta x = x_i - x_{i-1}$ with x_0 and x_m being end-points of its support, and p(x) is continuous within the bins, then the corresponding discrete entropy can be expressed as

$$H_q^n(p; \{c_k\}) = -\sum_{i=1}^m \frac{\Delta x p^q(\bar{x}_i)}{C(q; n)} (\Delta x)^{q-1} + \sum_{i=1}^m \sum_{k=1}^{n+1} \frac{c_{k-1}}{(k-1)!} \Delta x p^k(\bar{x}_i) (\Delta x)^{k-1}.$$
 (7)

If we put $\Delta x = 1$ as explained in Ref.[22], the first and second terms come close to the integral of $-p^q(\bar{x}_i)/C(q;n)$ and $\sum_{k=1}^{n+1} c_{k-1}p^k(\bar{x}_i)/(k-1)!$, respectively by the Riemann integrability.

4 Concluding remarks

It is noteworthy that for the general form of entropy such as $\sum_i f(p_i)$ (in our case $f(p_i) = p_i h(p_i)$), the additivity of the entropy holds only when it satisfies the relation $[p_i f''(p_i)]' = 0$. This relation straightforwardly leads to the Shannon form $f(p_i) = -p_i \ln p_i$ [23]. In this sense, a family of generalized entropies constructed in this paper can be considered to be intrinsically nonadditive.

Since the generalized form of entropy presented in this study contains a number of parameters (integration constants) depending on the order n, some properties satisfied by some known generalized entropies can be used as constraints for the parameters. Among preferable properties that should be possessed by entropies, a behavior for the equiprobability and certainty have been mentioned. As for obtaining the analytical expression of the equilibrium distribution p_i by the Lagrange multiplier method, it seems to be difficult to perform it for the general form. This is mainly because of the fact that the constrained (by energies) entropy functional, which should be solved, becomes a polynomial with respect to p_i .

In this paper, we have presented an elementary procedure to obtain a series of nonextensive entropies from some postulates on the self-information. As a main ingredient, the ratio $(h^{(n)}(x)/x)/h^{(n+1)}(x)$ has been assumed to be constant. Our formulation provides us that the Tsallis nonextensive entropy can be viewed as a special realization of a set of the large class of generalized entropies. In fact, this view is not new perspective. Some authors seem to have been shared this viewpoint in developing the generalized entropies [14, 24, 25]. We believe that the discussion presented in this study serves as understanding the structure of entropies and giving a deeper insight into a number of (generalized) entropies, not from a heuristic definition. We note that applications of these entropies to systems both in statistical mechanics and in information theory are potentially fruitful, but we need to know what aspect of the system relatetes to the nonextensivity parameters for further studies.

Acknowledgements

The author acknowledges the Research Fellowships from JSPS and support in part from Grant-in-Aid for Scientific Research No.06632 from the Ministry of Education, Science, Sports and Culture of Japan.

Appendix

Rudiments of information theory are how to define the self-information. Let Ω be a certain set or phase space: $\Omega = \{\omega_1, \omega_2, \ldots\}$. The phase space consists of an ensemble of elementary events (*microstates* in the language of the statistical mechanics). We are now considering the discrete case, which enables us to suppose to take the probability space (Ω, \mathcal{F}, P) , where \mathcal{F} is a σ -algebra over Ω and P is a probability measure $P: \mathcal{F} \to [0, 1]$ given by $P(\mathcal{F}) = \sum_{\omega_i \in \mathcal{F}} p_i$, where $p_i = p(\{\omega_i\})$.

Definition

A function I which takes real value defined on the \mathcal{F} is a self information or information content on \mathcal{F} if the axioms below are satisfied:

(A₁) The impossible event contains finite information $\theta(q)$ determined by system's intrinsic parameter q, whereas the certain event does not contain any information;

$$I(\phi) = \theta(q), \quad I(\Omega) = 0.$$

(A₂) For any two event ω_i and ω_j belonging to \mathcal{F} , such that $\omega_j \subset \omega_i$, we have

$$I(\omega_i) < I(\omega_j).$$

Proposition

Let $I(\omega_i)$ be an information content on \mathcal{F} . If the function f satisfies conditions:

$$f^{-1}: \mathbf{R}^+ \to [0, 1], \quad f^{-1}(0) = 1, \quad f^{-1}(\theta(q)) = 0$$

then the set function

$$\mu_f^{(q)}(\omega_i) = f^{-1}(I(\omega_i)) \quad (q \in \mathcal{R})$$
(8)

for every event $\omega_i \in \mathcal{R}$ is a probability measure on \mathcal{R} .

From the two axioms we conclude $I(\omega_i) \ge 0$ ($\forall \omega_i \in \mathcal{F}$). Then we immediately have the following statement on the self-information:

If the information independence of events belonging to \mathcal{F} does not coincide with the probability independence such that $I(\omega_i \cap \omega_j) = I(\omega_i) + I(\omega_j) + (q-1)I(\omega_i)I(\omega_j)$ (additivity violation), then we have $f(x) = k \ln_q x(k > 0)$. Hence the probability measure is given as $\mu_f^{(q)}(\omega_i) = e_q^{-\frac{I(\omega_i)}{k}}$ which is equivalent to

$$I(\omega_i) = -k \ln_q \mu_f^{(q)}(\omega_i).$$
(9)

Proof

By the condition of the above statement, for any two events ω_i and ω_j we have

$$\mu_f^{(q)}(\omega_i \cap \omega_j) = \mu_f^{(q)}(\omega_i)\mu_f^{(q)}(\omega_j) \tag{10}$$

only when $I(\omega_i \cap \omega_j) = I(\omega_i) + I(\omega_j) + (q-1)I(\omega_i)I(\omega_j)$. Let us put $x_i = I(\omega_i)$ and $x_j = I(\omega_j)$. Then

$$f^{-1}(x_i + x_j + (q-1)x_ix_j) = f^{-1}(I(\omega_i \cap \omega_j)) = \mu_f^{(q)}(\omega_i \cap \omega_j) = f^{-1}(I(\omega_i))f^{-1}(I(\omega_j))$$

The functional equation obtained is $f^{-1}(x_i + x_j + (q - 1)x_ix_j) = f^{-1}(x_i)f^{-1}(x_j)$. When we put $\phi(x) = ln_q f^{-1}(x)$, the functional equation which should be satisfied is

$$\phi(x_i + x_j + (q-1)x_ix_j) = \phi(x_i) + \phi(x_j) + (q-1)\phi(x_i)\phi(x_j)$$

whose solution is $\phi(x) = kx$. Therefore we obtain $f^{-1}(x) = e_q^{kx}$. When we replace the arbitrary constant k with -1/k, the functional form $f(x) = -k \ln_q x(k > 0)$ is concluded.

References

- [1] C. Tsallis, J. Stat. Phys. 52 (1988) 479.
- [2] J. Naudts, Physica A 316 (2002) 323.
- [3] C. Tsallis, Quimica Nova 17 (1994) 468; T. Yamano, Physica A 305 (2002) 486.
- [4] C. Beck, Phys. Rev. Lett. 87 (2001) 180601.
- [5] F. Sattin, J. Phys. A 36 (2003) 1583.
- [6] A.K. Rajagopal, S. Abe, Statistical mechanical foundations of power-law distributions, eprint, [cond-mat/0303064].

- [7] T.Yamano, Phys. Rev. E 63 (2001) 46105; Physica A 305 (2002) 190.
- [8] C. Tsallis, C. Anteneodo, L. Borlandand and R. Osorio, Physica A 324 (2003) 89.
- [9] G.H. Hardy, J.E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press (1973).
- [10] J. Havrda, F. Charvat, Kybernetika 30 (1967) 30.
- [11] Z. Daroczy, Inf. and Control 16 (1970) 36.
- [12] G. Kaniadakis, Phys. Rev. E 66 (2002) 056125.
- [13] C. Anteneodo, A.R. Plastino, J. Phys. A 32 (1999) 1089.
- [14] A.R.R. Papa, J. Phys. A 31 (1998) 5271.
- [15] C. Beck and F. Schlögl, Thermodynamics of chaotic Systems, Cambridge: Cambridge University Press (1993).
- [16] P.T. Landsberg, V. Vedral, Phys. Lett. A 247 (1998) 211.
- [17] A.K. Rajagopal, S. Abe, Phys. Rev. Lett. 83 (1999) 1711.
- [18] R.V.L. Hartley, Bell Syst. Tech. J. 7 (1928) 535.
- [19] S. Watanabe, Knowing and Guessing, a Quantitative Study of Inference and Information, New York: Wiley (1969).
- [20] C.H. Chen, Inform. Sci. 10 (1976) 159.
- [21] The concavity of the present family of entropies is highly parameter dependent. The function $f(x) = -x^q/C(q;n) + \sum_{k=1}^{n+1} c_{k-1}x^k/(k-1)!$ can be convex for $c_k > 0$, $\forall k$ if n satisfies $C(q;n) < 0 \ (> 0)$ when $q > 1 \ (q < 1)$. Then, for every $p_i, p'_i \in [0,1]$ and $\lambda_1, \lambda_2 \ (\lambda_1 + \lambda_2 = 1, \lambda_1 > 0, \lambda_2 > 0)$, the difference between entropy for the intermediate probability $p''_i \ (p''_i = \lambda_1 p_i + \lambda_2 p'_i)$ and the sum of the weighted entropies for two end-point probabilities p and p', which can be calculated as $\Delta_n^q = H_q^n(p''; \{c_k\}) \lambda_1 H_q^n(p; \{c_k\}) \lambda_2 H_q^n(p'; \{c_k\})$, can be shown to be positive $(\Delta_n^q \ge 0)$ with the Jensen's inequality. The equality holds if $c_k = 0$ for all k except for $k \neq 1$.
- [22] J.C.A. Van der Lubbe, *Information Theory*, Cambridge University Press (1997).
- [23] R. Rossignoli, N. Canosa, Phys. Lett. A 264 (1999) 148.

- [24] B.D. Sharma, D.P. Mittal, J. Math. Sci. 10 (1975) 28.
- [25] J.C.A. Van der Lubbe, Y. Boxma, D.E. Boekee, Inform. Sci. 32 (1984) 187.

©2004 by MDPI (http://www.mdpi.org). Reproduction for noncommercial purposes permitted.