BOUNDS ON DIVERGENCE IN NONEXTENSIVE STATISTICAL MECHANICS

Takuya Yamano

**Department of Information Sciences,
Faculty of Science, Kanagawa University,
2946, 6-233 Tsuchiya, Hiratsuka, Kanagawa 259-1293, Japan**
yamano@amy.hi-ho.ne.jp

Abstract. We focus on an important property upon generalization of the Kullback-Leibler divergence used in nonextensive statistical mechanics, i.e., bounds. We explicitly show upper and lower bounds on it in terms of existing familiar divergences based on the finite range of the probability distribution ratio. This provides a link between the observed distribution functions based on histograms of events and parameterized distance measures in physical sciences. The charactering parameter $q < 0$ and $q > 1$ are rejected from the consideration of bounded divergence.

Key words: Information divergences, Relative entropies, Generalized divergences.

Abstrak. Makalah ini berfokus pada sebuah sifat penting dari generalisasi divergensi Kullback-Leibler yang digunakan dalam statistik mekanika nonextensif, yakni, batas. Secara eksplisit diperlihatkan batas atas dan batas bawah dalam terma divergensi yang dikenal berdasarkan pada jelajah terhingga dari rasio distribusi peluang. Hal ini memberikan keterkaitan antara fungsi distribusi pengamatan berdasarkan pada histogram dari kejadian-kejadian dan ukuran jarak terparameter dalam ilmu-ilmu fisik. Parameter $q < 0$ dan $q > 1$ tertolak sebagai divergensi terbatas. consideration of bounded divergence.

Kata kunci: Divergensi informasi, entrofi relatif, generalisasi divergensi.

1. INTRODUCTION

The current upsurge of interest in divergence measures determined by two probability distributions is due to both usefulness and necessity for practical discriminations of different states and also for discovering how much they differ from

each other. Such scenarios appear in many areas which use statistical methods. Especially in statistical physics, the H-theorem is the most relevant notion to divergence measures that probe proximity toward a stationary distribution in the course of a given dynamics. Usually it is specified as the Kulback-Leibler divergence (or relative entropy) [11]. For Markovian processes, however, the validity of the proof of the H-theorem is shown for a wide class of divergences (Csiszár-Morimoto f-divergence) [2, 15]. For the specific forms of the generalized divergence, it is presented in [1, 18, 24]. Historically, an attempt for building a parameterized divergence measure in a statistical mechanics context has presented in [14] without using a notion of averaged information, where instead of the term 'divergence', a word of 'a relative degradation function of nth order' was used.

Numerous properties upon generalization of the conventional relative entropy, on the one hand, is becoming an interesting research topic in its own right, because generalizations of one conventional measure provide an insight into its original ones. Among them, the ranges or bounds of divergences can be regarded as fundamental, since they contain structural property reflecting the geometric manifold governed by the parameter used upon generalization. Also, the availability of bounds for the distance is important in physics and in statistical inference test, where the bound can be used to give an estimate for specific states (usually equilibrium states). Therefore, upper and lower bounds for distance measures in general can be useful information and provide a clue to interpret the meaning of the parameter.

The purpose of the present article is to provide such bounds for the Tsallis relative entropy that are not presented in the literature [23] so far. The approach here is based on the fact that a detector which produces statistical distributions of occured events has a finite dynamic range and consequently has finite probability distribution ratio. Therefore, these bounds obtained must be more relevant in terms of observational point of view. Our presentation proceeds as follows. First, we review the definition of the Tsallis relative entropy. Then we consider the bounds of it by the usual relative divergence. It can provide a degree of change by the parameter q of generalization. We next consider the upper bound by l_1-norm and the lower bound in terms of it (the so-called Pinsker like inequality). Bounds by χ^2 divergence followed by Hellinger's distance are presented as simple applications of inequalities that hold for generic f-divergences. We summarize our consideration and present discussion in the last section.

2. THE NONEXTENSIVE RELATIVE ENTROPY

It was provided in [22] in the context of the consistent testing and some properties were investigated in [1]. This generalization keeps the nonextensive thermostatistics picture [18] and it belongs to the relative information of type s, which has proposed in [17]. Presently, the generalization of the relative entropy which is consistent with the nonextensive entropy of Tsallis [21] is provided by

taking the linear mean (so-called the f-divergence [2]) of the corresponding distance measure f between two probability distributions [22],

$$D_q(p\|r) := \sum_i r_i f(t_i), \quad f(t_i) := \frac{t_i^q - t_i}{q - 1} \quad (q \in \mathbb{R}), \tag{1}$$

where t_i is used to denote the ratio of two probability distributions i.e., p_i/r_i throughout this paper. When supp $p \not\subseteq$ supp r, where supp $p = \{\omega \in \Omega; p(\omega) > 0\}$ in σ-finite measure space Ω, divergences become infinity and this applies for later considerations [25]. Alternatively, $D_q(p\|r)$ is produced from taking a biased average of the quantity $(p_i^{1-q} - r_i^{1-q})/(1-q)$ by p_i^q. This can be expressed as

$$D_q(p\|r) = \sum_i p_i^q \left(\frac{p_i^{1-q} - r_i^{1-q}}{1-q} \right).$$

3. BOUNDS IN TERMS OF USUAL KULLBACK-LEIBLER DIVERGENCE

We first describe our setting of consideration. In discrete cases, it is common to regard a histogram of observed values as a probability distribution associated with the system under study. This means that a measured value in a single measurement falls into one of the finite bins of a detector which has a finite dynamic range. The measuring apparatus consists of a limited number of bins, therefore the probability distribution also has a finite support reflecting the dynamic range. It is therefore highly probable that when we compare the two different probability distributions constructed in that manner, the ratio of them has finite ranges within the identical bin i. Let us call this quantity a *ratio range* hereafter. The ratio range becomes null if there is no detected event for the distribution r of ith bin. Furthermore, we set the minimum and the maximum values u and U, respectively on this ratio range for ith bin: $0 < u \leqslant p_i/r_i \leqslant U < \infty$. Under this setting, we consider bounds for the generalized relative entropy $D_q(p\|r)$.

Under this setting, we can use Theorem 6 provided in [4]. It was proved that when $f \in C^2(u, U)$ and when $tf''(t)$ is bounded from below and from above with constants $m \in \mathbb{R}$ and $M \in \mathbb{R}$, respectively, an inequality holds in terms of the Kullback-Leibler divergence $D_{KL}(p\|r)$ from p to r. More concretely, for the general f-divergences $D^f(p\|r)$ we know an inequality

$$m D_{KL}(p\|r) \leqslant D^f(p\|r) \leqslant M D_{KL}(p\|r). \tag{2}$$

Note that $D_{KL}(p\|r)$ is the most well known measure of the f-divergence class and it is obtained if we choose $f(t) = t \log t$ $(t > 0)$. For the divergence of Tsallis, we have $tf''(t) = qt^{q-1}$, then

$$\sup_{t \in (u, U)} tf''(t) = \begin{cases} qU^{q-1} & (q > 1) \\ qu^{q-1} & (q < 1), \end{cases} \tag{3}$$

$$\inf_{t \in (u, U)} tf''(t) = \begin{cases} qu^{q-1} & (q > 1) \\ qU^{q-1} & (q < 1). \end{cases} \tag{4}$$

Therefore, we obtain the following inequalities,

$$qu^{q-1}D_{KL}(p\|r) \leqslant D_q(p\|r) \leqslant qU^{q-1}D_{KL}(p\|r), \quad (q > 1), \tag{5}$$

$$qU^{q-1}D_{KL}(p\|r) \leqslant D_q(p\|r) \leqslant qu^{q-1}D_{KL}(p\|r), \quad (q < 1). \tag{6}$$

4. UPPER BOUNDS IN TERMS OF VARIATIONAL DISTANCE

The variational distance is also one of the f-divergences, since we can choose $f(t) = |t - 1|$ with $t \in \mathbb{R}_+$. Dragomir provided in [5] that the following inequality holds on $[u, U]$ if f is absolutely continuous and if $f' \in L^\infty[u, U]$,

$$0 \leqslant D^f(p\|r) \leqslant \|f'\|_{L^\infty}V(p, r), \tag{7}$$

where $\|f'\|_{L^\infty} := ess.\sup_{t\in(u,U)}|f'(t)|$ and $V(p, r) = \sum_i |p_i - r_i|$ is the variational distance (l_1-norm). In the range $0 < u \leqslant p_i/r_i \leqslant U < \infty$ for each i, the quantity for the divergence of Tsallis is found to be

$$\|f'\|_{L^\infty} = \begin{cases} \frac{qU^{q-1}-1}{q-1} & (0 < q < u^{1-q} < 1 \text{ or } 2 < q(U^{q-1} + u^{q-1}) \text{ or } 1 < u^{1-q} < q) \\ \frac{qu^{q-1}-1}{1-q} & (U^{1-q} < q < 1 \text{ or } 2 > q(U^{q-1} + u^{q-1}) \text{ or } 1 < q < U^{1-q}) \\ \frac{qu^{q-1}-1}{q-1} & (q < 0). \end{cases} \tag{8}$$

Therefore, we obtain the corresponding inequalities by substituting these into Eq.(7).

5. LOWER BOUNDS BY PINSKER TYPE INEQUALITY

The Pinsker inequality provides a lower bound on the Kullback-Leibler divergence in terms of the variational distance $V(p, r) = \sum_i |p_i - r_i|$ as $D_{KL}(p\|r) \geqslant V^2/2$ [16]. However, for other divergences, the corresponding inequality with the higher order in the variational distances was not known until recently. We present it by using the recent progress on the fourth-order extended Pinsker inequality proved in [9] and this gives a lower bound for the f-divergence measures. Under a certain condition, (for details see Theorem 7 in [9]) the following bound holds

$$D^f(p\|r) \geqslant \frac{f''(1)}{2}V^2 + \frac{1}{72}\left[3f^{(4)}(1) - 4\frac{(f^{(3)}(1))^2}{f''(1)}\right]V^4, \tag{9}$$

where the coefficients must be positive and are best possible in the sense that there exist no larger constants. Applying this bound for the divergence of Tsallis, we obtain

$$D_q(p\|r) \geqslant \frac{1}{2}qV^2 - \frac{1}{72}q(q-2)(q+1)V^4, \quad (q > 0). \tag{10}$$

When $q \to 1$, we recover the 4th order Pinsker's inequality for the Kullback-Leibler divergence, i.e. $D_{KL}(p\|r) \geqslant V^2/2 + V^4/36$, whose proof was provided in [10]. Note that the above bound are valid when $q > 0$, since the coefficient of V^2, viz.

$f''(1)$, must be positive. The refinement of the Pinsker inequality with best possible coefficients up to eighth order with respect to V has been obtained [20, 8] but the connection with physics remains unexplained so far, while the positivity of $D_q(p\|r) \geqslant 0$ (information inequality or Gibbs inequality) has a clear physical interpretation of the second law of thermodynamics.

6. BOUNDS IN TERMS OF HELLINGER'S DISTANCE

The Hellinger's distance also belongs to the f-divergence class and is obtained if we set $f(t) = (\sqrt{t} - 1)^2/2$ with $t \in \mathbb{R}_+$. It is shown in [6] that for $f \in C^2$ on (u, U) with the range $0 < u \leqslant 1 \leqslant U < \infty$, if $\exists m, M$ s.t. $m \leqslant t^{3/2} f''(t) \leqslant M$, then the following inequality holds,

$$4mh^2(p\|r) \leqslant D^f(p\|r) \leqslant 4Mh^2(p\|r). \tag{11}$$

We apply these bounds for our present consideration. For the divergence of Tsallis, we have $t^{\frac{3}{2}} f''(t) = q t^{q - \frac{1}{2}}$. Therefore, M and m are determined as

$$\sup_{t \in (u,U)} t^{\frac{3}{2}} f''(t) = \begin{cases} q U^{q-\frac{1}{2}} & (q > \frac{1}{2}) \\ q u^{q-\frac{1}{2}} & (q < \frac{1}{2}), \end{cases} \tag{12}$$

$$\inf_{t \in (u,U)} t^{\frac{3}{2}} f''(t) = \begin{cases} q u^{q-\frac{1}{2}} & (q > \frac{1}{2}) \\ q U^{q-\frac{1}{2}} & (q < \frac{1}{2}). \end{cases} \tag{13}$$

We then have the inequalities,

$$4q u^{q-\frac{1}{2}} h^2(p\|r) \leqslant D_q(p\|r) \leqslant 4q U^{q-\frac{1}{2}} h^2(p\|r), \quad (q > \frac{1}{2}), \tag{14}$$

$$4q U^{q-\frac{1}{2}} h^2(p\|r) \leqslant D_q(p\|r) \leqslant 4q u^{q-\frac{1}{2}} h^2(p\|r), \quad (q < \frac{1}{2}). \tag{15}$$

7. BOUNDS IN TERMS OF χ^2-DIVERGENCE

It is useful to give bounds of $D_q(p\|r)$ in terms of χ^2-divergence because it provides a bound for the mixing time of Markov chains [3]. When we set $f(t) = (t-1)^2$ on $[0, \infty)$, the χ^2-divergence $D_{\chi^2}(p\|r) = \sum_i (p_i - r_i)^2/r_i$ is also found to be a f-divergence. The Kullback-Leibler divergence is asymmetric about the exchange of any two probability distributions p and r. Their difference quantifies information to what extent the symmetry breaks. It was shown that the absolute value of the difference of that measure from p to r and from r to p is bounded from above in terms of χ^2-divergence [7],

$$|D_{KL}(p\|r) - D_{KL}(r\|p)| \leqslant \frac{U - u}{4uU} D_{\chi^2}(p\|r), \tag{16}$$

where p_i and r_i satisfy the range $0 < u \leqslant p_i/r_i \leqslant U < \infty$ for each i. The derivation of Eq.(16) comes from a trapezoid inequality, which holds for any f-divergences

$(f(1) = 0)$ obtained in [7]

$$\left| D^f(p\|r) - \frac{1}{2} D^{f_\#}(p\|r) \right| \leqslant \frac{1}{8}(\Gamma - \gamma)D_{\chi^2}(p\|r), \tag{17}$$

where the function $f_\#$ is defined for $t \in (0, \infty)$ as $f_\#(t) = (t-1)f'(t)$ and $f''(t)$ is assumed to be bounded by $\gamma = \inf_{t \in [u,U)} f''(t)$ and $\Gamma = \sup_{t \in [u,U)} f''(t)$. This inequality provides bounds for the difference of various f-divergences and enables us to evaluate and compare them in terms of the characteristics (the infimum and the supremum) of the second derivative of the functions f. We now apply it to the Tsallis divergence $D_q(p\|r)$. We have

$$f_\#(t) = \frac{1}{q-1}(t-1)(qt^{q-1} - 1). \tag{18}$$

With this $f_\#(t_i)$, where $t_i = p_i/r_i$, we obtain for the second term of Eq.(17) as,

$$
\begin{aligned}
D_q^{f_\#}(p\|r) &= \sum_i r_i(t_i - 1)\frac{q(t_i)^{q-1} - 1}{q-1} \\
&= q\left\{ \sum_i p_i \frac{(t_i)^{q-1} - 1}{q-1} - \sum_i r_i \frac{(t_i)^{q-1} - 1}{q-1} \right\} \\
&= q\left\{ D_q(p\|r) + D_{2-q}(r\|p) \right\}.
\end{aligned}
\tag{19}
$$

We note that when $q \to 1$, we have the Jeffereys divergence $D_q^{f_\#}(p\|r) \to D_{KL}(p\|r) + D_{KL}(r\|p) = \sum_i (p_i - r_i)\log(p_i/r_i)$, which is also one of the f-divergences. Then, we have for the l.h.s of the inequality Eq.(17),

$$\left| D_q(p\|r) - \frac{1}{2} D_q^{f_\#}(p\|r) \right| = \frac{1}{2}\left| (2-q)D_q(p\|r) - qD_{2-q}(r\|p) \right|. \tag{20}$$

Since $f''(t) = qt^{q-2}$ and $0 < u \leqslant t_i \leqslant U < \infty$, $\forall i$, we have

$$
\begin{cases}
qu^{q-2} \leqslant f''(t) \leqslant qU^{q-2} & (q > 2) \\
qU^{q-2} \leqslant f''(t) \leqslant qu^{q-2} & (0 < q < 2).
\end{cases}
\tag{21}
$$

Noting the relation,

$$D_q(r\|p) = \frac{q}{1-q} D_{1-q}(p\|r), \tag{22}$$

and using Γ and γ obtained from Eq.(21) and applying them to the inequality Eq.(17), we have finally inequalities

$$
\begin{aligned}
&\left| (2-q)\{D_q(p\|r) - \frac{q}{q-1}D_{q-1}(p\|r)\} \right| \\
&\leqslant \begin{cases}
\frac{q}{4}(U^{q-2} - u^{q-2})D_{\chi^2}(p\|r) & (q > 2) \\
\frac{q}{4}(u^{q-2} - U^{q-2})D_{\chi^2}(p\|r) & (0 < q < 2).
\end{cases}
\end{aligned}
\tag{23}
$$

When one needs the bounds for l.h.s. of Eq.(23) in terms of other f-divergences, we need to know the corresponding bounds on χ^2-divergence. Taneja and Kumar [19] provided the upper bounds on the χ^2-divergence in terms of the Kullback-Leibler divergence and of the Hellinger's distance $h(p\|r) := \sum_i (\sqrt{p_i} - \sqrt{r_i})^2/2$ as

$D_{\chi^2}(p\|r) \leqslant 2U^2 D_{KL}(r\|p)$ and as $D_{\chi^2}(p\|r) \leqslant 8\sqrt{U^3}h(p\|r)$, respectively. With these inequalities, the inequality Eq.(23) can be bounded by using $D_{KL}(p\|r)$ and $h(p\|r)$. Anyway, the upper bounds in terms of χ^2-divergence is tighter than those in terms of others.

8. AN UPPER BOUND ON AN OVERLAP BETWEEN DIVERGENCES

We here concern how much the values of divergences differ each other when we measure them with the identical reference distribution r. Namely, we shall want to know an upper bound of an *overlap* between quantities $D_q(p\|r)$ and $D_q(p'\|r)$ for a given q. To this end we define the following quantity

$$\int_0^U \int_0^U f(t)g(u)|t-u|^{-\lambda}dtdu, \tag{24}$$

where the two functions are respectively defined as $f(t) = (t^q - t)/(q-1)$ and $g(u) = (u^q - u)/(q-1)$ with $t = p/r$ and $u = p'/r$. The case $\lambda = 0$ corresponds to the usual overlap of functions f and g (or the inner product of two real funcions). Otherwise it can give a normalized overlap tempered by λ. We are concerned with the bound on it. Recalling that for $\alpha, \beta > 1$ and $0 < \lambda < n$ with a relation $1/\alpha + \lambda/n + 1/\beta = 2$, the Hardy-Littlewood-Sobolev inequality [12] reads

$$\left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} dxdy \frac{f(x)g(y)}{|x-y|^\lambda} \right| \leqslant C(n, \lambda, p)\|f\|_\alpha \|g\|_\beta \tag{25}$$

where a sharp constant $C(n, \lambda, p)$ independent of functions f and g when $\alpha = \beta = 2n/(2n-\lambda)$ is given as,

$$\pi^{\frac{\lambda}{2}} \frac{\Gamma(n/2 - \lambda/2)}{\Gamma(n - \lambda/2)} \left(\frac{\Gamma(n/2)}{\Gamma(n)} \right)^{\frac{\lambda}{n} - 1}. \tag{26}$$

Our problem is in a special case by equating $n = 1$ and $f = g$ by setting its form as $(t^q - t)/(q-1)$. The desired upper bound is the multiple of the value of Eq.(26) and that of $\|f\|^2_{\frac{2}{2-\lambda}}$ given in Appendix.

9. SUMMARY AND CONCLUDING REMARKS

We have presented the fundamental bounds on the generalized KL divergence used in nonextensive statistical mechanics in terms of several known divergences. The bounds for the parameterized divergence is indispensable, since without it most (if not all) of the nonextensive structures in system would not fully be understood. Our starting assumption was the existence of maximum and minimum values for the ratio of probability distributions, which originates from the finite dynamic range of measuring apparatus for events associated with physical system. The bounds depend on the parameter range that bears nonextensivity. Then, a natural question arises concerning the parameter range i.e., which value of q we should use.

Regarding this question, we resort to the criterion given as an inequality, which holds for generic f-divergence [13],

$$0 \leqslant D^f(p\|r) \leqslant \lim_{u \to +0} \left\{ f(u) + uf(u^{-1}) \right\}, \quad u = \frac{p}{r}. \qquad (27)$$

Since we have $f(u) = (u^q - u)/(q-1)$ for Tsallis divergence, the upper bound after taking the limit in the above becomes

$$\begin{cases} \frac{1}{1-q} & (0 < q < 1) \\ +\infty & (q > 1). \end{cases} \qquad (28)$$

When $q < 0$ it does not give a bound. Therefore, the use of limitation is found to be $0 < q < 1$ in order to guarantee the feasibility of measuring distance between two distribution functions.

10. APPENDIX

When $0 < q < 1$ the norm $\|f\|^2_{\frac{2}{2-\lambda}}$ is calculated to be

$$(1-q)^{-2} \left(\frac{(\lambda - 2)U^{1-q}(U - U^q)^{\frac{4-\lambda}{2-\lambda}}\, {}_2F_1 \left[1, \frac{(2-q)(2-\lambda)-2}{(1-q)(2-\lambda)}; \frac{(2-q)(2-\lambda)-2q}{(1-q)(2-\lambda)}; U^{1-q} \right]}{2q + 2 - \lambda} \right)^{2-\lambda} \qquad (29)$$

where $_2F_1(a, b; c; d)$ is the Hypergeometric function.

REFERENCES

[1] Borland, L., Plastino, A.R., and Tsallis, C., "Information gain within generalized thermostatistics", *J. Math. Phys.*, **39** (1998), 6490 - 6501; *Erratum* **40** (1999), 2196.

[2] Csiszár, I., "Information-type measures of difference of probability distributions and indirect observations", *Studia Math. Hungarica*, **2** (1967), 299 - 318; *ibid* **2** (1967), 329 - 339.

[3] Diaconis, P., and Stroock, D., "Geometric bounds for eigenvalues of Markov chains", *Ann. Appl. Prob.*, **1** (1991), 36 - 61; Fill, J.A., "Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process", *ibid* **1** (1991), 62 - 87 .

[4] Dragomir, S.S., *RGMIA Monographs: Inequalities for Csiszár f-Divergences in Information Theory*, (1999) Ch. II 1.

[5] Dragomir, S.S., *RGMIA Monographs: Inequalities for Csiszár f-Divergences in Information Theory*, (1999) Ch. II 3.

[6] Dragomir, S.S., *RGMIA Monographs: Inequalities for Csiszár f-Divergences in Information Theory*, (1999) Ch. II 2, Theorem 6.

[7] Dragomir, S.S., Gluščević, V., and Pearce, C.E.M., "New approximation for f-divergence via trapezoid and midpoint inequalities", *RGMIA Res. Rep. Coll.*, **14** (2002), 1 - 8.

[8] Fedotov, A., Harremoës, P., and Topsøe, F., "Refinements of Pinsker Inequality", *IEEE Tans. Infom. Theory*, **49** (2003), 1491 - 1498.

[9] Gilardoni, G.L., *On Pinsker's type inequalities and Csiszár's f-divergence. Part I: Second and Fourth-order inequalities (Preprint* arXiv:cs/0603097v2).

[10] Kraft, O., and Schmitz, N., "A note on Hoeffding inequality", *J. Amer. Statist. Assoc.*, **64** (1969), 907 - 912.

[11] Kullback, S., and Leibler, R.A., "On Information and Sufficiency", **Ann. Math. Stat.**, **22** (1951), 79 - 86; Kullback, S. *Information Theory and Statistics* (Wiley, New York, 1959).

[12] Lieb, E.H., and Loss, M., *Analysis Graduate Studies in Mathematics Vol.14, Second Edition*, AMS, 2001.

[13] Liese, F., and Vajda, I., *Convex statistical distances, vol. 95 of Teubner-Texte zur Mathematik*, BSB BG Teubner Verlagsgesellschaft, Leipzig, 1987.

[14] Lindhard, J., "On the theory of measurement and its consequences in statistical dynamics", *Det Kongelige Danske Videnskabernes Selskab Matematisk-fysiske Meddelelser*, **39** (1974), 1 - 39.

[15] Morimoto, T., "Markov processes and the H-Theorem", *J. Phys. Soc. Jpn.*, **18** (1963), 328 - 331.

[16] Pinsker, M.S., *Information and Information Stability of Random Variables and Processes, ed. A. Feinstein*, Holden-Day, San Francisco, 1964.

[17] Sharma, B.D., and Autar, R., "Relative-Information Functions and Their Type (α, β) Generalizations", *Metrika*, **21** (1974), 41 - 50.

[18] Shiino, M., "H-Theorem with Generalized Relative Entropies and theTsallis Statistics", *J. Phys. Soc. Jpn.*, **67** (1998), 3658 - 3660.

[19] Taneja, I.J., and Kumar, P., "Relative Information of Type s, Csiszar f-Divergence, and Information Inequalities", *Information Sciences* **166** (2004), 105 - 125.

[20] Topsøe, F., "Bounds for entropy and divergence of distributions over a two-element set", *J. Ineq. Pure Appl. Math.*, **2** (2001), Article 25 .

[21] Tsallis, C., "Possible generalization of Boltzmann-Gibbs statistics", *J. Stat. Phys.*, **52** (1988), 479 - 487.

[22] Tsallis, C., "Generalized entropy-based criterion for consistent testing", *Phys. Rev. E*, **58** (1998), 1442 - 1445.

[23] Tsallis, C., *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, 2009.

[24] Yamano, T., "H-theorems based upon a generalized, power law-like divergence", *Phys. Lett. A.*, **374** (2010), 3116 - 3118.

[25] For mathematical rigor, we mention the following for the continuous case. When the ratio of distribution functions p and r appears, we suppose that P is absolutely continuous with respect to R ($P \ll R$), where P and R are probability measures with σ-finite measure μ from the Radon-Nikodym derivative $p = dP/d\mu$ and $r = dR/d\mu$, respectively. If $P \ll R$ is not satisfied, the divergence is $+\infty$.