

Article

A Direct Link between Rényi–Tsallis Entropy and Hölder’s Inequality—Yet Another Proof of Rényi–Tsallis Entropy Maximization

Hisa-Aki Tanaka *, Masaki Nakagawa and Yasutada Oohama

Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan; m.nakagawa@ist.osaka-u.ac.jp (M.N.); oohama@uec.ac.jp (Y.O.)

* Correspondence: htanaka@uec.ac.jp; Tel.: +81-42-443-5174

Received: 30 April 2019; Accepted: 27 May 2019; Published: 30 May 2019



Abstract: The well-known Hölder’s inequality has been recently utilized as an essential tool for solving several optimization problems. However, such an essential role of Hölder’s inequality does not seem to have been reported in the context of generalized entropy, including Rényi–Tsallis entropy. Here, we identify a direct link between Rényi–Tsallis entropy and Hölder’s inequality. Specifically, we demonstrate yet another elegant proof of the Rényi–Tsallis entropy maximization problem. Especially for the Tsallis entropy maximization problem, only with the equality condition of Hölder’s inequality is the q -Gaussian distribution uniquely specified and also proved to be optimal.

Keywords: Rényi–Tsallis entropy; generalized entropy; optimization; Hölder’s inequality

1. Introduction

Tsallis entropy [1,2] has been recently utilized as a versatile framework for expanding the realm of Shannon–Boltzmann entropy for nonlinear processes, in particular, those that exhibit power-law behavior. It shares a structure in common with Rényi entropy [3], Daróczy entropy [4], and probability moment presented in Moriguti [5], since the essential part of all these functionals is $\int p^q(x)dx$ (or $\sum p_i^q$) for certain constrained probability density functions $p(x)$ (or p_i). This naturally has been of interest for a variety of issues in information theory and related areas. For instance, in his pioneering work, Campbell [6] stated that “Implicit in the use of average code length as a criterion of performance is the assumption that cost varies linearly with code length. This is not always the case.” Then, Campbell [6] introduced a nonlinear average length measure defined as

$$L(t) = \frac{1}{t} \log_D \sum_{i=1}^N p_i D^{tl_i},$$

being an extension of the one by Shannon,

$$L_0 = \sum_{i=1}^N p_i l_i,$$

in which D is the size of the alphabet, p_i is the probability for a source to produce symbols x_i , l_i is the length of a codeword c_i mapped from symbol x_i (using D letters of the alphabet) in the context of source coding, and t is an arbitrary parameter ($0 < t < \infty$). One of the surprising facts proved in [6] is that the lower bound to the moment-generating function of code lengths, namely, $L(t)$, is given by

$H_{\frac{1}{1+t}}(p)$, namely, Rényi entropy of order $(1+t)^{-1}$ of the source $p = \{p_i\}_{i=1}^N$. Moreover, Ref. [6] also realizes that, if

$$l_i = \frac{1}{1+t} \log_D \frac{1}{p_i} + \frac{t}{1+t} H_{\frac{1}{1+t}}(p),$$

which is a mixture of the Shannon code length $\log_D \frac{1}{p_i}$ and Rényi entropy of order $(1+t)^{-1}$, we have the lower bound $L(t) = H_{\frac{1}{1+t}}(p)$. So far, Baer [7] has further generalized this result and constructed an algorithm for finding optimal binary codes under quasiarithmetic penalties. In addition, new extensions of [6] were obtained by Bercher [8] and by Bunte and Lapidoth [9].

Such an instance, where “a nonlinear measure” (i.e., generalized entropy) naturally arises, is also known for channel capacities. Daróczy [4] first analyzed a generalized channel capacity, which is a natural consequence of his extension of Shannon entropy (i.e., Daróczy entropy). This result has initiated extensive work in this direction. For instance, Landsberg and Vedral [10] first introduced Rényi entropy and Tsallis entropy for a binary symmetric channel, and they suggested the possibility of “super-Shannon channel capacities.” More recently, Ilić, Djordjević, and Kueppers [11] obtained new expressions for generalized channel capacities by introducing Daróczy–Tsallis entropy even for a weakly symmetric channel, binary erasure channel, and z-channel. Similar extensions have been explored for rate distortion theory. For instance, Venkatesan and Plastino [12] developed nonextensive rate distortion theory by introducing Tsallis entropy and constructed a minimization algorithm for generalized mutual information. More recently, Girardin and Lhote [13] covered the setting in [12] in a general framework of generalized entropy rates, which includes Rényi–Tsallis entropy.

In the context of generalized entropy just described, the q -Gaussian distribution [1,2] often emerges as a maximizer of Rényi–Tsallis entropy under certain constraints, and, hence, it has been extensively studied. Since the q -Gaussian effectively models power-law behavior with a one-parameter q , its utility is widespread in various areas, including new random number generators proposed by Thistleton, Marsh, Nelson, and Tsallis [14] and by Umeno and Sato [15]. In addition to such an important application in communication systems, queuing theory has recently incorporated the q -Gaussian, reflecting the heavy-tailed traffic characteristics observed in broadband networks [16–19]. For instance, Karmeshu and Sharma [16] introduced Tsallis entropy maximization, and, there, the q -Gaussian emerges as the queue length distributions, which suggests that Jaynes’ maximum entropy principle [20–22] can be generalized to a framework of Tsallis entropy.

Some of the above issues are formulated as nonlinear optimizations with “a nonlinear measure” under certain constraints (which depend on each issue). As mentioned above, Rényi–Tsallis entropy and q -Gaussian is one such instance. In other words, the q -Gaussian maximizes Tsallis entropy under certain constraints. Therefore, it is useful to obtain a deeper understanding of such nonlinear optimization problems. In this study, we find a direct link between Rényi–Tsallis entropy and Hölder’s inequality that leads to yet another elegant proof of Rényi–Tsallis entropy maximization. The idea of the proof is different from those offered in previous studies (for instance, [23–27]) as explained below. Interestingly, the technique developed in this study might possibly be useful for tackling more complicated problems regarding optimization issues in information theory and other research areas, such as the conditional Rényi entropy (as in [28–30]), for instance.

Previous studies [23–27] are based on a common standpoint, the generalization of the moment–entropy inequality (cf. [25,26]). Namely, they intend to generalize the situation that a continuous random variable with a given second moment and maximal Shannon entropy is a Gaussian distribution (cf. [3], Theorem 8.6.5). In doing so, a *generalized relative entropy* is devised, which takes a different form (and has a different name) depending on the problem. First of all, Tsukada and Suyari’s beautiful work [23] has given proofs for Rényi entropy maximization, which is also known as a bound of Moriguti’s probability moment [5] (as posed in **R1** in Section 2). Namely, they prove that the q -Gaussian distribution [1,2] is a unique optimal solution by utilizing the fact that all feasible solutions constitute a convex set. Although [23] does not explicitly construct a generalized relative entropy, the

essential structure of the proofs inherits the one in the proof of the moment-entropy inequality ([3], Theorem 8.6.5)).

Moreover, they have identified an explicit one-to-one correspondence between feasible solutions to the problems of Rényi entropy maximization and Tsallis entropy maximization, which is also shown in ([31], p. 754). This implies that an ‘indirect’ proof to Tsallis entropy maximization (as posed in T1 in Section 2) has been first obtained in [23]. In contrast to this proof, the first ‘direct’ proof to Tsallis entropy maximization is obtained in Furuichi’s elegant work [24]. The proof in [24] utilizes nonnegativity of the *Tsallis relative entropy* defined between the q -Gaussian distribution (i.e., a possible maximizer) and any other feasible solution. On the other hand, the remarkable work of Lutwak, Yang, and Zhang first clarified that *generalized Gaussians* maximize λ -Rényi entropy power under a constraint on the p -th moment of the distribution, for univariate distributions [25] and for the associated n -dimensional extensions [26]. The essential point in the proofs in [25,26] is construction of *relative λ -Rényi entropy power*, which is nonnegative and takes a quite different form compared to the *Tsallis relative entropy* in [24]. (More precisely, in [25], they prove nonnegativity of the relative λ -Rényi entropy $\log N_\lambda[f, g]$ ([25], Lemma 1). Starting from this nonnegativity: $\log N_\lambda[f, Gt] \geq 1$, they construct a series of inequalities that saturate at the generalized Gaussian ([25], Lemma 2). Note, however, that, as observed in this $N_\lambda[f, Gt]$, they start by giving a candidate of the maximizer ab initio, which is the generalized Gaussian Gt .) Furthermore, Vignat, Hero, and Costa [32] obtained a general, sharp result using the *Bregman information divergence* for an n -dimensional extension of Tsallis entropy. In addition to [25,26,32], Eguchi, Komori, and Kato’s interesting results [27] include the same n -dimensional extension to Tsallis entropy. (Ref. [32] has also identified an elegant structure regarding the projective divergence and the γ -loss functions in maximum likelihood estimation.) Similar to [24–26,32], the key component of the proof in [27] is the *projective power divergence*, which again takes a quite different form compared to the ones in [24–26,32]. To prove nonnegativity of the generalized relative entropy, Refs. [25–27] utilize Hölder’s inequality, but Refs. [23,24,32] do not. Namely, Hölder’s inequality has been an auxiliary useful tool, and it has never played an essential role in these previous studies. In addition to the construction of generalized relative entropies, the optimal q -Gaussian distribution needs to be ‘given ab initio’ [23–27,32], inheriting the framework showing that the Gaussian distribution maximizes Shannon entropy ([3], Theorem 8.6.5).

Now natural questions arise: is it possible to systematically solve the problems of Rényi–Tsallis entropy maximization in a different (and hopefully simpler) way than the previous study? In addition, is it possible to ‘construct’ the q -Gaussian distribution? These questions are positively answered from a new viewpoint as follows. First, only by the equality (i.e., saturation) condition of Hölder’s inequality, the q -Gaussian distribution is specified, and, at the same time, its optimality is proved by Hölder’s inequality for a Tsallis entropy maximization of $1 < q < 3$ (Theorem 1) and of $0 \leq q < 1$ (Theorems 2 and 3). This clarifies how and why the q -Gaussian distribution emerges as the maximizer in an explicit way for the first time in the literature. (To the authors’ knowledge, such a characterization of the q -Gaussian distribution has never been reported.) However, for a Rényi entropy maximization of $q > 1$ (Theorem 4) and of $\frac{1}{3} < q < 1$ (Theorem 5), the q -Gaussian distribution is specified with the aid of the equality condition of Hölder’s inequality. In addition, the proof of its optimality requires a simple inequality inspired from Moriguti [5]. Note that we do not intend to provide an explicit characterization of the q -Gaussian distribution in terms of the parameter q , since numerous previous studies (including [23–27]) have already clarified this. Nevertheless, regarding Tsallis entropy maximization when $q = 0$, which has previously been studied in [2], a rigorous result (as in Theorem 3) is now obtained for the first time thanks to Hölder’s inequality. (For instance, in the framework of [24], the case for $q = 0$ cannot be incorporated because the Tsallis relative entropy is not defined adequately.)

We note that Hölder’s inequality has been recently utilized as an essential tool for optimization in Campbell [6], Bercher [8], and Bunte and Lapidot [9]; on source coding, in Bercher [33,34]; on generalized Cramér–Rao inequalities; and in Tanaka [35,36] on a physical limit of injection locking. However, such an essential role of Hölder’s inequality does not seem to be reported in the context

of generalized entropy, including Rényi entropy (cf. [37]), except for the use as a means for proving nonnegativity of a generalized relative entropy, as mentioned above.

In what follows, Section 2 introduces basic definitions required for the analysis. Section 3 includes the main results regarding Rényi–Tsallis entropy maximization problems, and it also contains an explanation on the link to Moriguti’s argument in [5]. Section 4 lists the proofs to the results presented in Section 3. Finally, one Appendix at the end provides further supplementary information.

2. Basic Definitions and Problem Formulation

In this section, we first define Tsallis entropy [1,24] and Rényi entropy ([3], pp. 676–679). Next, we reformulate Rényi–Tsallis entropy maximization problems in a unified way. Finally, we introduce Hölder’s inequality in relation to the problems in this study.

2.1. Tsallis Entropy and Rényi Entropy

Tsallis entropy is beautifully presented in the context of q -analysis (cf. [1], p. 41) as follows. First, the q -exponential function \exp_q , whose domain and range satisfies

$$\exp_q : \begin{cases} \left[\frac{-1}{1-q}, \infty \right) \rightarrow \mathbb{R}^+ \cup \{0\} & \text{if } q < 1, \\ \left(-\infty, \frac{1}{q-1} \right] \rightarrow \mathbb{R}^+ & \text{if } q > 1, \end{cases}$$

is defined by

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

While, the inverse of the q -exponential function, namely q -logarithmic function, is defined by

$$\ln_q x = \frac{x^{1-q} - 1}{1 - q}.$$

Note that, as $q \rightarrow 1$, we have $\exp_q x \rightarrow e^x$ and $\ln_q x \rightarrow \ln x$. We also note that the above definition of $\exp_q x$ and $\ln_q x$ has been recently revised by Oikonomou and Bagci [38]. (In [38], they have further developed ‘complete’ q -exponentials and q -logarithms.) Then, the Tsallis entropy H_q^{Tsallis} is defined by

$$H_q^{\text{Tsallis}}[p] = - \int_{-\infty}^{\infty} p^q(x) \ln_q p(x) dx = - \langle\langle p^q(x) \ln_q p(x) \rangle\rangle, \quad (1)$$

for univariate probability density functions (PDFs) p on \mathbb{R} , which is a natural generalization of Boltzmann–Gibbs entropy and Shannon entropy. Hereafter, $\langle\langle \cdot \rangle\rangle = \int_{-\infty}^{\infty} \cdot dx$ in (1) is used for notational simplicity. The reason why $\langle\langle \cdot \rangle\rangle$ is used, instead of $\langle \cdot \rangle$, is due to the fact that $\langle \cdot \rangle$ is generally used for the expectation value. On the other hand, Rényi entropy is well-known and can be found in textbooks of information theory (cf. [3], pp. 676–679), which is defined simply by

$$H_q^{\text{Rényi}}[p] = \frac{\ln \langle\langle p^q(x) \rangle\rangle}{1 - q} \quad (0 < q < \infty \quad (q \neq 1)).$$

Finally, we note that only differential entropies (i.e., continuous probability distributions) are considered in this study, although our technique with Hölder’s inequality can be applied to discrete probability distributions.

2.2. Problem Formulation

Let \mathcal{D} be the set of all PDFs on \mathbb{R} . We then define the set, as introduced in [24],

$$\mathcal{C}_q = \left\{ p \mid p \in \mathcal{D} \text{ and } \frac{\langle x^2 p^q(x) \rangle}{\langle p^q(x) \rangle} < \infty \right\} (\subset \mathcal{D}).$$

Following the problem formulation in [1,2,23,24,31], we first introduce the Tsallis entropy maximization problem for univariate PDFs p on \mathbb{R} :

$$\mathbf{T1:} \quad \underset{p \in \mathcal{C}_q}{\text{maximize}} \quad H_q^{\text{Tsallis}}[p] = -\langle p^q(x) \ln_q p(x) \rangle = \frac{1 - \langle p^q(x) \rangle}{q-1} \quad (\text{for } q \geq 0 \ (q \neq 1)) \quad (2a)$$

$$\text{subject to } \langle p(x) \rangle = 1, \quad (2b)$$

$$\frac{\langle x^2 p^q(x) \rangle}{\langle p^q(x) \rangle} = \langle x^2 \tilde{p}_q(x) \rangle = \sigma^2, \quad (2c)$$

in which q and σ^2 have fixed values, and $\tilde{p}_q(x) = p^q(x)/\langle p^q(x) \rangle$. Note that $\langle p^q(x) \rangle < \infty$ and $\langle x^2 p^q(x) \rangle < \infty$ are assumed in **T1**. $\tilde{p}_q(x)$ is often called the escort probability [27,31]. This somewhat unusual form of expectation $\langle x^2 \tilde{p}_q(x) \rangle = \sigma^2$ is called the q -normalized expectation [31], which has been usually assumed in Tsallis statistics. In contrast to the q -normalized expectation, as [31] pointed out, the usual expectation $\langle x^2 p(x) \rangle = \sigma^2$ is also valid in Tsallis statistics. We note the Tsallis entropy maximization problem under the constraint of this usual expectation is considered later in problem **R2**.

For problem **T1**, using Tsallis relative entropy, Furuichi [24] first proved that for $0 < q < 3$ the q -Gaussian distribution $p(x) = \frac{1}{Z_q} \exp_q(-\beta_q x^2)$ maximizes the Tsallis entropy among any univariate PDFs in \mathcal{C}_q , where Z_q and β_q are constants determined by q and σ .

Here, we formulate a slightly generalized optimization problem **T2**, as follows. First, replace (2c) with

$$\langle x^2 p^q(x) \rangle - \sigma^2 \langle p^q(x) \rangle = \langle (x^2 - \sigma^2) p^q(x) \rangle = 0.$$

Note that now, as opposed to **T1**, it is not necessarily required that both $\langle x^2 p^q(x) \rangle$ and $\langle p^q(x) \rangle$ are finite, and hence, \mathcal{C}_q is not required, and it is replaced with \mathcal{D} . Next, notice that Tsallis entropy is maximal at $p(x)$, such that $\langle p^q(x) \rangle$ is minimal (or correspondingly, maximal at $p(x)$, such that $\langle p^q(x) \rangle$ is maximal) for $q > 1$ (correspondingly, for $0 \leq q < 1$). Then, by introducing an additional arbitrary parameter λ_q , **T1** is reformulated as

$$\mathbf{T2:} \quad \underset{p \in \mathcal{D}}{\text{minimize (or maximize)}} \quad T_q[p; \lambda_q] = \sigma^2 \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2) p^q(x) \rangle$$

$$= \langle p^q(x) [\lambda_q x^2 + (1 - \lambda_q) \sigma^2] \rangle \quad (\lambda_q \in \mathbb{R}, \text{ for } q > 1 \text{ (correspondingly, for } 0 \leq q < 1)) \quad (3a)$$

$$\text{subject to } \langle p(x) \rangle = 1, \quad (3b)$$

$$\langle (x^2 - \sigma^2) p^q(x) \rangle = 0, \quad (3c)$$

where the constant σ^2 is multiplied with $\langle p^q(x) \rangle$ in the first term of (3a) simply due to notational convenience for later analysis in Section 4.

As opposed to the Tsallis entropy maximization problem **T1**, the Rényi entropy maximization problem is usually considered under the constraint of the usual expectation $\langle x^2 p(x) \rangle = \sigma^2$, in other words,

$$\mathbf{R1:} \quad \underset{p \in \mathcal{D}}{\text{maximize}} \quad H_q^{\text{Rényi}}[p] = \frac{\ln \langle p^q(x) \rangle}{1-q} \quad (\text{for } 0 < q < \infty \ (q \neq 1)) \quad (4a)$$

$$\text{subject to } \langle p(x) \rangle = 1, \quad (4b)$$

$$\langle x^2 p(x) \rangle = \sigma^2, \quad (4c)$$

which is equivalent to :

$$\underset{p \in \mathcal{D}}{\text{minimize (or maximize)}} \langle p^q(x) \rangle \quad (q > 1 \text{ (correspondingly, } 0 \leq q < 1)) \quad (5a)$$

$$\text{subject to } \langle p(x) \rangle = 1, \quad (5b)$$

$$\langle x^2 p(x) \rangle = \sigma^2. \quad (5c)$$

We note this very problem for $q > 1$ was first posed and solved by Moriguti in 1952 [5]. (Later in [39], cases $q > 1$ and $0 < q < 1$ are both analyzed in an n -dimensional spherical symmetric extension of [5] with the same approach as [5].) Similar to **T1**, by introducing an additional parameter λ_q and the constraint

$$\langle x^2 p(x) \rangle - \sigma^2 \langle p(x) \rangle = \langle (x^2 - \sigma^2) p(x) \rangle = 0, \quad (6)$$

which is obtained from (5b) and (5c), **R1** is now reformulated as

$$\begin{aligned} \mathbf{R2:} \quad & \underset{p \in \mathcal{D}}{\text{minimize (or maximize)}} R_q[p; \lambda_q] = \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2) p(x) \rangle \\ & = \langle p^q(x) [1 + \lambda_q (x^2 - \sigma^2) p^{1-q}(x)] \rangle \quad (\text{for } q > 1 \text{ (correspondingly, for } 0 \leq q < 1), \lambda_q \in \mathbb{R}) \quad (7a) \\ & \text{subject to } \langle p(x) \rangle = 1, \quad (7b) \\ & \quad \langle (x^2 - \sigma^2) p(x) \rangle = 0. \quad (7c) \end{aligned}$$

As we observe (3a) in **T2** and (7a) in **R2**, both become the inner products of two functions; $p^q(x)$ and $\lambda_q x^2 + (1 - \lambda_q) \sigma^2$ and $p^q(x)$ and $1 + \lambda_q (x^2 - \sigma^2) p^{1-q}(x)$, respectively. This suggests a direct link to Hölder's inequality.

2.3. Hölder's Inequality for Later Analysis

Here, we provide minimum information about Hölder's inequality for later analysis in Section 3 and Section 4. The standard Hölder's inequality is given by

$$\|fg\|_1 \leq \|f\|_\alpha \|g\|_\beta, \quad (8)$$

with $1 \leq \alpha, \beta \leq \infty$ and $\alpha^{-1} + \beta^{-1} = 1$ (cf. [40] for the one-dimensional case and [41] for general measurable functions). In general, f and g are measurable functions defined on a subset $S \subseteq \mathbb{R}^n$ and $\mu(S) > 0$, and we employ a compact notation as

$$\begin{aligned} \|f\|_\alpha &= \left(\int_S |f|^\alpha d\mu \right)^{\frac{1}{\alpha}}, \\ \|g\|_\beta &= \left(\int_S |g|^\beta d\mu \right)^{\frac{1}{\beta}}. \end{aligned}$$

Although $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are no longer norms for $\alpha, \beta < 1$, now in the context of this study, we set $\alpha = q^{-1}$ and $\beta = (1 - q)^{-1}$. Then, Hölder's inequality (8) is given in the following form:

$$\|fg\|_1 \leq \|f\|_{\frac{1}{q}} \|g\|_{\frac{1}{1-q}} \quad (0 \leq q \leq 1). \quad (9)$$

For the case $0 < q < 1$, the equality in (9) holds if and only if there exists constants A and B , not both 0 (cf. [40], p. 140), (More specifically, if f is null (i.e., $f(s) = 0$ (a.e. $s \in S$)), then $B = 0$. In addition, if g is null, $A = 0$.) such that

$$A |f(s)|^{\frac{1}{q}} = B |g(s)|^{\frac{1}{1-q}} \quad (\text{a.e. } s \in S). \quad (10)$$

In addition, for the exceptional case $q = 0$ (as well as $q = 1$), we can argue a condition for the equality in (9) separately, as shown in Section 4.3, although the expression of (10) is no more valid for this case.

In contrast to (9), reverse Hölder's inequality is given by

$$\|fg\|_1 \geq \|f\|_{\frac{1}{q}} \|g\|_{\frac{-1}{q-1}} \quad (q > 1), \quad (11)$$

which is directly obtained from Hölder's inequality [40]. We note that f can be 0 over any subset $U \subseteq S$. As for g , on the other hand, we assume $g(s) \neq 0$ for almost everywhere (a.e.) $s \in S$, taking care that $\frac{-1}{q-1} < 0$ in (11) (cf. [40], p. 140). Then, for the case $q > 1$, the equality in (11) holds if and only if there exists $A \geq 0$, such that

$$|f(s)| = A |g(s)|^{\frac{-q}{q-1}} \quad (\text{a.e. } s \in S \subseteq \mathbb{R}^n). \quad (12)$$

3. Main Results

In this study, we focus on the univariate PDFs on \mathbb{R} , and we consider $f(x)$ and $g(x)$ defined on \mathbb{R} as a special case of general $f(s)$ and $g(s)$ in Section 2.3. Hereafter, we refer to (10) and (12), as the equality condition of Hölder's inequality and reverse Hölder's inequality, respectively. Thanks to these equality conditions, we obtained our results systematically.

Let $p(x)$ be a univariate PDF defined on \mathbb{R} . Assume that $p(x)$ is a measurable function which is integrable with respect to x . In addition, let $B(\cdot, \cdot)$ denote the Beta function (cf. [42], p. 253). Then, we can form the following statements.

Theorem 1. (Tsallis entropy maximization for $1 < q < 3$): Suppose $1 < q < 3$. $p_{\text{opt}}(x)$, defined by

$$p_{\text{opt}}(x) = \frac{1}{Z_q} \exp_q \left(-\beta_q x^2 \right) \quad \text{with } Z_q = \sqrt{\frac{3-q}{q-1}} \sigma B \left(\frac{1}{q-1}, \frac{1}{2}, \frac{1}{2} \right) \text{ and } \beta_q = \frac{1}{(3-q)\sigma^2},$$

is the unique maximizer of the Tsallis entropy $H_q^{\text{Tsallis}}[p]$ in (2a) under the constraints $\langle p(x) \rangle = 1$ of (3b) and $\langle (x^2 - \sigma^2)p^q(x) \rangle = 0$ of (3c) in T2.

Corollary 1. For $q \geq 3$, Tsallis entropy $H_q^{\text{Tsallis}}[p]$ is bounded, but has no maximizer. Namely, there exist PDFs $p(x)$, such that $H_q^{\text{Tsallis}}[p] \rightarrow \frac{1}{q-1} \leq \frac{1}{2}$, in other words, $\langle p^q(x) \rangle \rightarrow 0$. (The idea for constructing such PDFs is from Tsukada and Suyari [23], where they proved that **R1** for $q \leq \frac{1}{3}$ becomes unbounded, i.e., $\langle p^q(x) \rangle \rightarrow +\infty$.)

The proof of this theorem (and corollary) is given in Section 4.1. As mentioned in Section 1, the above statement itself has already appeared in [24]. However, our proof is quite different to the one in [24], in the sense that it does not require *generalized relative entropy*, and the maximizer is explicitly 'specified' (not 'given ab initio'). Namely, reverse Hölder's inequality aids in finding the optimal solution.

The outline of the proof is as follows. First, for $1 < q < 3$, the maximization of the Tsallis entropy $H_q^{\text{Tsallis}}[p]$, in other words, the minimization of $T_q[p; \lambda_q]$ in (3a), is related to reverse Hölder's inequality in (11). Second, we observe that $T_q[p; \lambda_q]$ has the lower bound through reverse Hölder's inequality. Third, the minimizer $p_{\text{opt}}(x)$ achieving this bound is explicitly and uniquely constructed from the equality condition (12):

$$p_{\text{opt}}(x) = A^{\frac{1}{q}} \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{-1}{q-1}}, \quad (13)$$

where $A = \left[\frac{3-q}{2} \sigma^2 \right]^{\frac{q}{q-1}} \left[\sqrt{\frac{3-q}{q-1}} \sigma B \left(\frac{1}{q-1}, \frac{1}{2}, \frac{1}{2} \right) \right]^{-q}$ and $\lambda_{q,\text{opt}} = \frac{1}{2}(q-1)$.

Remark 1. Even if we assume the additional constraint: $p \in \mathcal{C}_q (\subset \mathcal{D})$ in **T2**, the proof of this theorem (as well as of Theorems 2 and 3) remains the same, since we do not require the finiteness of $\langle\langle x^2 p^q(x) \rangle\rangle$ and $\langle\langle p^q(x) \rangle\rangle$ (i.e., $p \in \mathcal{C}_q$) in the proof.

Remark 2. Another simple proof for optimality of p_{opt} is given as follows. The idea is due to Moriguti's argument (cf. [5], p. 288), where $p_{\text{opt}}^q(x)$ and any $p^q(x)$ are directly related by the Taylor expansion for each $x \in \mathbb{R}$:

$$p^q(x) = p_{\text{opt}}^q(x) + q p_{\text{opt}}^{q-1}(x)[p(x) - p_{\text{opt}}(x)] + \frac{q(q-1)}{2} p_{\text{int}}^{q-2}(x)[p(x) - p_{\text{opt}}(x)]^2, \quad (14)$$

where $p_{\text{int}}(x) (\geq 0)$ has a value between $p(x)$ and $p_{\text{opt}}(x)$. Substituting (13) into the second term of the right-hand side of (14), we have

$$\begin{aligned} & [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2] [p^q(x) - p_{\text{opt}}^q(x)] \\ &= q A^{\frac{q-1}{q}} [p(x) - p_{\text{opt}}(x)] + [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2] \frac{q(q-1)}{2} p_{\text{int}}^{q-2}(x) [p(x) - p_{\text{opt}}(x)]^2. \end{aligned} \quad (15)$$

With the constraints (3b) and (3c): $\langle\langle x^2 p_{\text{opt}}^q(x) \rangle\rangle = \sigma^2 \langle\langle p_{\text{opt}}^q(x) \rangle\rangle$, $\langle\langle x^2 p^q(x) \rangle\rangle = \sigma^2 \langle\langle p^q(x) \rangle\rangle$, and $\langle\langle p(x) \rangle\rangle = \langle\langle p_{\text{opt}}(x) \rangle\rangle = 1$, integrating (15) over \mathbb{R} , for any $p(x)$, we find

$$\sigma^2 [\langle\langle p^q(x) \rangle\rangle - \langle\langle p_{\text{opt}}^q(x) \rangle\rangle] = \frac{q(q-1)}{2} \langle\langle [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2] p_{\text{int}}^{q-2}(x) [p(x) - p_{\text{opt}}(x)]^2 \rangle\rangle \geq 0, \quad (16)$$

since $\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 > 0$ follows from $0 < \lambda_{q,\text{opt}} (= \frac{1}{2}(q-1)) < 1$. Therefore, p_{opt} in (13) is a unique optimal solution to **T2**, as the equality holds only if $p = p_{\text{opt}}$.

Theorem 2. (Tsallis entropy maximization for $0 < q < 1$): Suppose $0 < q < 1$. $p_{\text{opt}}(x)$, as defined by

$$p_{\text{opt}}(x) = \begin{cases} \frac{1}{Z_q} \exp_q(-\beta_q x^2) & (x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases}$$

with $\bar{S}_{q,\text{opt}} = \left[-\sqrt{\frac{3-q}{1-q}} \sigma, \sqrt{\frac{3-q}{1-q}} \sigma \right]$, $Z_q = \sqrt{\frac{3-q}{1-q}} \sigma B\left(\frac{1}{1-q} + 1, \frac{1}{2}\right)$, and $\beta_q = \frac{1}{(3-q)\sigma^2}$,

is the unique maximizer of the Tsallis entropy $H_q^{\text{Tsallis}}[p]$ in (2a) under the constraints $\langle\langle p(x) \rangle\rangle = 1$ of (3b) and $\langle\langle (x^2 - \sigma^2) p^q(x) \rangle\rangle = 0$ of (3c) in **T2**.

The proof of this theorem is given in Section 4.2. In the case for $0 < q < 1$, the maximization of $T_q[p; \lambda_q]$ is recast as Hölder's inequality in (9), where, similar to the argument in the proof of Theorem 1, construction of p_{opt} and $\lambda_{q,\text{opt}}$ and verification of its optimality are carried out simultaneously. The maximizer p_{opt} for $0 < q < 1$ is uniquely determined from the equality condition (10):

$$p_{\text{opt}}(x) = \begin{cases} \frac{A}{B} [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{1}{1-q}} & (x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases} \quad (17)$$

where $A/B = \left[\frac{3-q}{2} \sigma^2 \right]^{\frac{1}{q-1}} \left[\sqrt{\frac{3-q}{1-q}} \sigma B\left(\frac{1}{1-q} + 1, \frac{1}{2}\right) \right]^{-1}$ and $\lambda_{q,\text{opt}} = \frac{1}{2}(q-1) (< 0)$ are uniquely determined, and the associated $\bar{S}_{q,\text{opt}}$ is uniquely determined as $\bar{S}_{q,\text{opt}} = \left[-\sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}} \sigma, \sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}} \sigma \right]$.

Remark 3. Another simple proof for optimality of p_{opt} is given by following Moriguti's argument ([5], p. 288). Similar to the case for $1 < q < 3$ in Remark 2, (14) holds for $x \in \bar{S}_{q,\text{opt}} \subset \mathbb{R}$. As for $x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}$, $p_{\text{opt}}(x) = 0$ from (17). We then have

$$\begin{aligned} & \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right] \left[p^q(x) - p_{\text{opt}}^q(x) \right] \\ &= \begin{cases} qA^{\frac{q-1}{q}} [p(x) - p_{\text{opt}}(x)] + \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right] \frac{q(q-1)}{2} p_{\text{int}}^{q-2}(x) [p(x) - p_{\text{opt}}(x)]^2 & (x \in \bar{S}_{q,\text{opt}}) \\ \left\langle \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right] p^q(x) \right\rangle_{\mathbb{R} \setminus \bar{S}_{q,\text{opt}}} & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases} \end{aligned} \quad (18)$$

where $\langle \cdot \rangle_{\mathbb{R} \setminus \bar{S}_{q,\text{opt}}} = \int_{\mathbb{R} \setminus \bar{S}_{q,\text{opt}}} \cdot dx$ is used for notational simplicity. Integrating (18) over \mathbb{R} , for any p satisfying the constraints (3b) and (3c), we find

$$\begin{aligned} & \sigma^2 \left[\langle p^q(x) \rangle - \langle p_{\text{opt}}^q(x) \rangle \right] \\ &= qA^{\frac{q-1}{q}} \langle p(x) - p_{\text{opt}}(x) \rangle_{\bar{S}_{q,\text{opt}}} + \frac{q(q-1)}{2} \left\langle \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right] p_{\text{int}}^{q-2}(x) [p(x) - p_{\text{opt}}(x)]^2 \right\rangle_{\bar{S}_{q,\text{opt}}} \\ & \quad + \left\langle \left[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right] p^q(x) \right\rangle_{\mathbb{R} \setminus \bar{S}_{q,\text{opt}}} \leq 0, \end{aligned} \quad (19)$$

since the first term on the right-hand side ≤ 0 as $\langle p(x) \rangle_{\bar{S}_{q,\text{opt}}} \leq 1$, the second term ≤ 0 from the definition of $\bar{S}_{q,\text{opt}}$ and the fact that $0 < q < 1$, and the third term ≤ 0 because of the definition of $\bar{S}_{q,\text{opt}}$. Since the equality in (19) holds only if $p = p_{\text{opt}}$, this implies that p_{opt} in (17) is a unique optimal solution to T2.

Theorem 3. (Tsallis entropy maximization for $q = 0$): Suppose $q = 0$. $p_{\text{opt}}(x)$, defined by

$$\begin{aligned} p_{\text{opt}}(x) &= \begin{cases} \text{arbitrary positive value} & (x \in \bar{S}_{q,\text{opt}} = [-\sqrt{3}\sigma, \sqrt{3}\sigma]) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases} \\ & \text{with } \langle p_{\text{opt}}(x) \rangle_{\bar{S}_{q,\text{opt}}} = 1, \end{aligned} \quad (20)$$

is the unique representation of the maximizer of the Tsallis entropy $H_q^{\text{Tsallis}}[p]$ in (2a) under the constraints $\langle p(x) \rangle = 1$ of (3b) and $\langle (x^2 - \sigma^2) p^0(x) \rangle = 0$ of (3c) in T2.

The proof of this theorem is given in Section 4.3, where the associated Hölder's inequality is given as $\|fg\|_1 \leq \|f\|_\infty \|g\|_1$, and we follow the arguments in the proof of Theorem 2 for $0 < q < 1$. (This exceptional case ($q = 0$) is also considered in T1, where the same result is obtained through a more direct graphical argument, after proving that any candidate $p_*(x)$ for the maximizer $p_{\text{opt}}(x)$ is defined only on a simply connected interval S_* that is symmetric about the origin O. The proof is straightforward but lengthy, so we omit it here.) However, as opposed to the case for $0 < q < 1$, the equality condition is not available in the form of (10) for $q = 0$, and we directly verify that

$$f_*(x) = \begin{cases} \text{sgn}[g(x)] & (\text{a.e. } x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}) \end{cases}$$

is the unique solution satisfying the equality in (9), as shown in Lemma 1. Namely, for any feasible solutions $p(x)$ satisfying the constraints (3b) and (3c), we find that $f_*(x)$ is associated with the unique maximizer $p_{\text{opt}}(x)$ of $T_q[p; \lambda_q]$ from (52), and hence, $p_{\text{opt}}(x)$ for $q = 0$ is obtained as in (20).

Remark 4. We note that the optimal solution shown in ([2], p. 2399, Figure 1) for $q = 0$, which is obtained by setting $q \rightarrow 0$ in (17), is a special case of (20).

Theorem 4. (Rényi entropy maximization for $q > 1$): Suppose $q > 1$. $p_{\text{opt}}(x)$, as defined by

$$p_{\text{opt}}(x) = \begin{cases} \frac{1}{Z_q^*} \exp_{q^*}(-\beta_q^* x^2) & (x \in \mathcal{S}_{\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \mathcal{S}_{\text{opt}}), \end{cases}$$

$$\text{with } q_* = 2 - q, \mathcal{S}_{\text{opt}} = \left[-\sqrt{\frac{3q-1}{q-1}}\sigma, \sqrt{\frac{3q-1}{q-1}}\sigma\right], Z_q^* = \sqrt{\frac{3q-1}{q-1}}\sigma B\left(\frac{q}{q-1}, \frac{1}{2}\right), \text{ and } \beta_q^* = \frac{1}{(3q-1)\sigma^2},$$

is the unique maximizer of Rényi entropy in (4a) under the constraints $\langle p(x) \rangle = 1$ of (5b) (or (7b)) and $\langle x^2 p(x) \rangle = \sigma^2$ of (5c) (or (7c)).

The proof of this theorem is given in Section 4.4. The minimization of $R_q[p; \lambda_q]$ in (7b) for $q > 1$ is related to reverse Hölder's inequality in (11). In contrast to those of Theorems 1–3, the proof of Theorem 4, which can be found in Section 4.4, follows from two steps. In the first step, we construct a candidate for the minimizer (i.e., $p_{\text{opt}}(x)$, see (61) below), whose support becomes \mathcal{S}_{opt} , and we determine the associated $\lambda_{q,\text{opt}}$ and \mathcal{S}_{opt} through the equality condition of reverse Hölder's inequality. In doing so, as shown in Figure 1, we introduce a subset of feasible solutions $p(x)$, in other words, \mathcal{Q} , which satisfies the constraints (5b) and (5c), and an additional constraint: $p^q(x) + \lambda_q(x^2 - \sigma^2)p(x) > 0$ ($x \in \mathbb{R}$). In the second step, after obtaining a candidate $p_{\text{opt}}(x) \in \mathcal{Q}$, we verify that this $p_{\text{opt}}(x)$ is indeed the unique minimizer of $R_q[p; \lambda_q]$ by directly comparing $\langle\langle p_{\text{opt}}^q(x) \rangle\rangle$ and $\langle\langle p^q(x) \rangle\rangle$ for any feasible solutions $p(x)$ satisfying the constraints (5b) and (5c).

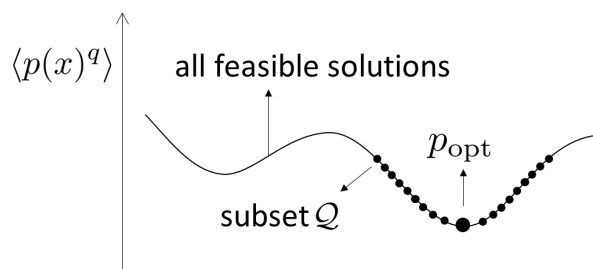


Figure 1. This figure illustrates how our approach for Theorem 4 works in a possible structure of our optimization problem. The whole curve represents all feasible solutions, and the dotted points represent the subset \mathcal{Q} in all feasible solutions.

Remark 5. We note that the first proof for this optimality of p_{opt} has been given in Moriguti [5], in which the essential idea is the Taylor expansion shown in the argument below (18).

Theorem 5. (Rényi entropy maximization for $\frac{1}{3} < q < 1$): Suppose $\frac{1}{3} < q < 1$. $p_{\text{opt}}(x)$, defined by

$$p_{\text{opt}}(x) = \frac{1}{Z_q^*} \exp_{q^*}(-\beta_q^* x^2)$$

$$\text{with } q_* = 2 - q, Z_q^* = \sqrt{\frac{3q-1}{1-q}}\sigma B\left(\frac{1}{1-q}, \frac{1}{2}\right) \text{ and } \beta_q^* = \frac{1}{(3q-1)\sigma^2},$$

is the unique maximizer of Rényi entropy (4a) under the constraints (5b) and (5c) (or (7b) and (7c)).

The proof of this theorem is given in Section 4.5. Maximization of $R_q[p; \lambda_q]$ is related to Hölder's inequality in (9) and the proof follows two steps, similar to the proof for Theorem 4. In the first step, we construct a candidate for the maximizer (i.e., $p_{\text{opt}}(x)$, given below by (72)) and determine $\lambda_{q,\text{opt}}$ through the equality condition of Hölder's inequality. In the second step, after obtaining a candidate p_{opt} , we verify that this p_{opt} is indeed the unique maximizer of $R_q[p; \lambda_q]$ by directly comparing $\langle\langle p_{\text{opt}}^q(x) \rangle\rangle$ and

$\langle p^q(x) \rangle$ for any feasible solutions $p(x)$ satisfying the constraints (5b) and (5c). This verification is done as in the proofs for Theorem 4. Although omitted here, using essentially the same argument as in Remark 1, another simple proof based on Moriguti [5] is possible.

Remark 6. Tsukada and Suyari [23] have proved that **R1** for $0 < q \leq \frac{1}{3}$ becomes unbounded. As for the exceptional case of $q = 0$, the upper and lower bounds of $\langle p^q(x) \rangle (= \langle p^0(x) \rangle)$ are argued as follows. First, if we consider the Gaussian distribution that satisfies (5b) and (5c), this gives us $\langle p^0(x) \rangle = \langle 1 \rangle_{\mathbb{R}} = \infty$, and it implies there is no maximizer. Next, consider a particular distribution given by

$$\Delta(x) = \begin{cases} \delta^{-1} & (x \in [\bar{\sigma}, \bar{\sigma} + \delta]) \\ 0 & (\text{otherwise}), \end{cases} \quad (21)$$

with $\delta > 0$. This $\Delta(x)$ satisfies $\langle \Delta(x) \rangle = \delta^{-1}\delta = 1$ in (5b), and it also satisfies $\langle x^2\Delta(x) \rangle = \sigma^2$ in (5c) when δ is arbitrary small, in other words,

$$\forall \sigma, \delta (\ll \sigma) \exists \bar{\sigma} \quad \langle x^2\Delta(x) \rangle = \bar{\sigma}^2 + \delta\bar{\sigma} + \frac{\delta^2}{3} = \sigma^2, \quad (22)$$

and, this particular distribution gives $\langle \Delta^0(x) \rangle = \langle 1 \rangle_{[\bar{\sigma}, \bar{\sigma} + \delta]} = \delta \rightarrow 0$ ($\delta \rightarrow 0$), which implies there is no minimizer. Therefore, problem **R1** (and **R2**) has no maximizer nor minimizer for $q = 0$.

4. Proof of Main Results

Following the outlines leading to Theorems 1–5 in Section 3, here we give their proofs.

4.1. Proof of Theorem 1

Proof. Let p be arbitrary feasible solutions to **T2** for $1 < q < 3$, and let p_{opt} be its optimal solution, which is eventually constructed in (25). Let $\lambda_{q,\text{opt}}$ be a particular value of the additional parameter λ_q in **T2**, which is associated with p_{opt} and is eventually constructed in (29). Then, for any p and a particular $\lambda_{q,\text{opt}}$ ($= \frac{1}{2}(q-1)$ in (29)), we define f and g as

$$f(x) = p^q(x) (\geq 0), \quad (23a)$$

$$g(x) = \lambda_{q,\text{opt}}x^2 + (1 - \lambda_{q,\text{opt}})\sigma^2 (> 0). \quad (23b)$$

First, we show $T_q[p; \lambda_q]$ is minimized in the following way:

$$T_q[p; \lambda_q] = T_q[p; \lambda_{q,\text{opt}}] = \langle fg \rangle \quad (24a)$$

$$= \|fg\|_1 \geq \|p^q\|_{\frac{1}{q}} \|g\|_{\frac{q-1}{q-1}} \quad (24b)$$

$$= \|g\|_{\frac{q-1}{q-1}} \text{ (the lower bound)}. \quad (24c)$$

The first “=” in (24a) follows from the fact that $T_q[p; \lambda_q] = \sigma^2 \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2)p^q(x) \rangle$ in (3a) is independent from the value of λ_q , since any feasible solution p satisfies $\langle (x^2 - \sigma^2)p^q(x) \rangle = 0$ in (3c), and the second “=” in (24a) is immediate from (23). The “=” in (24b) follows from $p^q(x) \geq 0$ and $g(x) > 0$ in (23b), since in (29) $\lambda_{q,\text{opt}} = \frac{1}{2}(q-1)$, and it satisfies $0 < \lambda_{q,\text{opt}} < 1$. The “ \geq ” in (24b) follows from reverse Hölder’s inequality (11). The “=” in (24c) follows from $\|p^q\|_{\frac{1}{q}} = \langle |p(x)|^q \rangle^{\frac{1}{q}} = 1$. (24c) implies that $T_q[p; \lambda_q]$ has the lower bound (i.e., the Tsallis entropy $H_q^{\text{Tsallis}}[p]$ in (1) has the upper bound).

Next, we construct a maximizer p_{opt} achieving this bound and show its uniqueness, which is done by checking the conditions where the “ \geq ” in (24) become “=”; the only “ \geq ” in (24b) becomes

“=” if and only if the equality condition (12) is satisfied. Now, we rewrite the equality condition (12) (after assuming $S = \mathbb{R}$ in (12)) by using (23), which constructs (a candidate of) p_{opt} :

$$p_{\text{opt}}(x) = A^{\frac{1}{q}} [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{-1}{q-1}} = \frac{1}{Z_q} \exp_q \left(-\beta_q x^2 \right), \quad (25)$$

where A and $\lambda_{q,\text{opt}}$ ($= \frac{1}{2}(q-1)$) are uniquely determined, and hence, Z_q and β_q are also uniquely determined, as shown in the following calculations. We note the formula ([42], p. 253):

$$\int_0^\infty \frac{dx}{x^\alpha (1+x^\gamma)^\beta} = \frac{1}{\gamma} B \left(\beta - \frac{1-\alpha}{\gamma}, \frac{1-\alpha}{\gamma} \right)$$

is repeatedly used for the integrations below. (More precisely, for (26), we set $\alpha = 0$ (< 1), $\beta = \frac{1}{q-1}$ (> 0), $\gamma = 2$ (> 0), and $\beta\gamma > 1 - \alpha$ is satisfied. For (27b), we set $\alpha = -2$, 0 (< 1), $\beta = \frac{q}{q-1}$ (> 0), $\gamma = 2$ (> 0), and $\beta\gamma > 1 - \alpha$ is satisfied.) First, by substituting p_{opt} in (25) into the constraint (3b), and for $1 < q < 3$ (finiteness of the left-hand side of (3b) requires the condition $q < 3$), we have

$$\langle\langle p_{\text{opt}}(x) \rangle\rangle = A^{\frac{1}{q}} [(1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{1}{1-q}} \sqrt{\frac{1 - \lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}}}} \sigma B \left(\frac{1}{q-1} - \frac{1}{2}, \frac{1}{2} \right) = 1. \quad (26)$$

On the other hand, substituting (25) into $n \langle\langle p^q(x) \rangle\rangle$ and $\langle\langle x^2 p^q(x) \rangle\rangle$ in the constraint (3c), we have

$$\langle\langle p_{\text{opt}}^q(x) \rangle\rangle = A [(1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{q}{1-q}} \sigma \left[\sqrt{\frac{1 - \lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}}}} B \left(\frac{1}{q-1} + \frac{1}{2}, \frac{1}{2} \right) \right], \quad (27a)$$

$$\langle\langle x^2 p_{\text{opt}}^q(x) \rangle\rangle = A [(1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{q}{1-q}} \sigma^3 \left[\left(\frac{1 - \lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}}} \right)^{\frac{3}{2}} B \left(\frac{1}{q-1} - \frac{1}{2}, \frac{3}{2} \right) \right], \quad (27b)$$

and substitution of (27a) and (27b) into (3c) yields

$$A [(1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{q}{1-q}} \sigma^3 \left[\left(\frac{1 - \lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}}} \right)^{\frac{3}{2}} B \left(\frac{1}{q-1} - \frac{1}{2}, \frac{3}{2} \right) - \sqrt{\frac{1 - \lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}}}} B \left(\frac{1}{q-1} + \frac{1}{2}, \frac{1}{2} \right) \right] = 0. \quad (28)$$

Then, using the formula $B(x, y+1) = \frac{y}{x} B(x+1, y)$ ([42], p. 254) in (28), $\lambda_{q,\text{opt}}$ is uniquely determined as

$$\frac{\lambda_{q,\text{opt}}}{1 - \lambda_{q,\text{opt}}} = \frac{q-1}{3-q}, \text{ i.e., } \lambda_{q,\text{opt}} = \frac{1}{2}(q-1), \quad (29)$$

and from (26) and (29) A is uniquely determined as

$$A = \left[\frac{3-q}{2} \sigma^2 \right]^{\frac{q}{q-1}} \left[\sqrt{\frac{3-q}{q-1}} \sigma B \left(\frac{1}{q-1} - \frac{1}{2}, \frac{1}{2} \right) \right]^{-q}. \quad (30)$$

In (25), equating the second term to the third term, β_q is uniquely determined as

$$\beta_q = \frac{\lambda_{q,\text{opt}}}{1 - \lambda_{q,\text{opt}}} (q-1)^{-1} \sigma^{-2} = \frac{1}{(3-q) \sigma^2},$$

and from (29) and (30), Z_q is uniquely determined as

$$Z_q = \left\{ A^{\frac{1}{q}} [(1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{1}{1-q}} \right\}^{-1} = \sqrt{\frac{3-q}{q-1}} \sigma B \left(\frac{1}{q-1} - \frac{1}{2}, \frac{1}{2} \right).$$

This proves that p_{opt} in (25) is a unique minimizer to **T2** for $1 < q < 3$.

Finally, we prove the Corollary to Theorem 1 in Section 3. For $q \geq 3$, let p_* be the distribution satisfying (3b) and (3c), defined as

$$p_*(x) = Z^{-1}(|x| + \alpha)^{-(1+\varepsilon)},$$

where a normalization factor $Z = \int (|x| + \alpha)^{-(1+\varepsilon)} dx$ and $\alpha, \varepsilon > 0$. What we are going to prove is $\langle\langle p_*^q(x) \rangle\rangle \rightarrow 0$ ($\varepsilon \rightarrow 0$), which is done as follows. First, straightforward integrations yield:

$$Z = 2\alpha^{-\varepsilon}\varepsilon^{-1}, \quad (31a)$$

$$\langle\langle p_*^q(x) \rangle\rangle = 2\alpha^{1-\bar{q}}(\bar{q}-1)^{-1}Z^{-q}, \quad (31b)$$

$$\langle\langle x^2 p_*^q(x) \rangle\rangle = 4\alpha^{3-\bar{q}}(\bar{q}-1)^{-1}(\bar{q}-2)^{-1}(\bar{q}-3)^{-1}Z^{-q}, \quad (31c)$$

where $\bar{q} = (1 + \varepsilon)q > 3$. Second, from the constraint in (3c), $2(\bar{q}-2)^{-1}(\bar{q}-3)^{-1}\alpha^2 = \sigma^2$ is obtained, and this shows that α becomes finite and is determined by ε, q , and σ . Finally, substituting (31a) into (31b), we obtain

$$\langle\langle p_*^q(x) \rangle\rangle = 2^{1-q}(\bar{q}-1)^{-1}\varepsilon^q, \quad (32)$$

and hence $\langle\langle p_*^q(x) \rangle\rangle \rightarrow 0$ ($\varepsilon \rightarrow 0$), since for $q \geq 3$ in (32) $(\bar{q}-1)^{-1} \rightarrow (q-1)^{-1}$ and $\varepsilon^q \rightarrow 0$ ($\varepsilon \rightarrow 0$) in (32). \square

4.2. Proof of Theorem 2

Proof. Let p be arbitrary feasible solutions to **T2** for $0 < q < 1$, and let p_{opt} be its optimal solution, which is eventually constructed in (38). Let $\lambda_{q,\text{opt}}$ be a particular value of the additional parameter λ_q in **T2**, which is associated with p_{opt} and is eventually constructed in (42). Then, for any p and a particular $\lambda_{q,\text{opt}}$ ($= \frac{1}{2}(q-1)$ in (42)), we define f and g as

$$f(x) = p^q(x), \quad (33a)$$

$$g(x) = \lambda_{q,\text{opt}}x^2 + (1 - \lambda_{q,\text{opt}})\sigma^2, \quad (33b)$$

and we define an interval

$$\bar{S}_{q,\text{opt}} = \left[-\sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}}\sigma, \sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}}\sigma \right].$$

First, we show that $T_q[p; \lambda_q]$ is maximized in the following way:

$$T_q[p; \lambda_q] = T_q[p; \lambda_{q,\text{opt}}] = \langle\langle fg \rangle\rangle \quad (34a)$$

$$\leq \langle\langle fg \rangle\rangle_{\bar{S}_{q,\text{opt}}} = \langle\langle |fg| \rangle\rangle_{\bar{S}_{q,\text{opt}}} \quad (34b)$$

$$= \|fg\|_{1, \bar{S}_{q,\text{opt}}} \leq \|f\|_{\frac{1}{q}, \bar{S}_{q,\text{opt}}} \|g\|_{\frac{1}{1-q}, \bar{S}_{q,\text{opt}}} \quad (34c)$$

$$\leq \|g\|_{\frac{1}{1-q}, \bar{S}_{q,\text{opt}}} \quad (\text{the upper bound}), \quad (34d)$$

where $\langle\langle \cdot \rangle\rangle = \int_{-\infty}^{\infty} \cdot dx$, $\langle\langle \cdot \rangle\rangle_{\bar{S}_{q,\text{opt}}} = \int_{\bar{S}_{q,\text{opt}}} \cdot dx$, and $\|\cdot\|_{\alpha, \bar{S}_{q,\text{opt}}} = \left(\int_{\bar{S}_{q,\text{opt}}} |\cdot|^\alpha dx \right)^{\frac{1}{\alpha}}$. The first “=” in (34a) follows from the fact that $T_q[p; \lambda_q] = \sigma^2 \langle\langle p^q(x) \rangle\rangle + \lambda_q \langle\langle (x^2 - \sigma^2)p^q(x) \rangle\rangle$ in (3a) is independent from the value of λ_q , since any feasible solution p satisfies $\langle\langle (x^2 - \sigma^2)p^q(x) \rangle\rangle = 0$ in (3c), and the second “=” in (34a) is immediate from (33). The “ \leq ” in (34b) is obtained from the following observation. By plotting the graph of $g(x)$ for (any negative) $\lambda_{q,\text{opt}}$, we observe that $\bar{S}_{q,\text{opt}}$ is the set of x on which $g(x)$ becomes

positive. For any f and g in (34a), we also observe from this graph that $\langle fg \rangle_{\bar{S}_{q,\text{opt}}} > \langle fg \rangle_{S_*}$ for any set $S_* (\subset \mathbb{R})$ but $\bar{S}_{q,\text{opt}}$, since

$$f(x) \geq 0 \ (\forall x \in \mathbb{R}), \ g(x) \geq 0 \ (x \in \bar{S}_{q,\text{opt}}), \text{ and } g(x) < 0 \ (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}). \quad (35)$$

On the other hand, the “=” in (34b) is immediate from $f(x)g(x) \geq 0 \ (\forall x \in \bar{S}_{q,\text{opt}})$. The first “=” in (34c) follows from the definition of $\|\cdot\|_{1,\bar{S}_{q,\text{opt}}}$, and the inequality in (34c) follows from Hölder’s inequality (9). In view of (33a), the final “≤” in (34d) follows from

$$\|f\|_{\frac{1}{q},\bar{S}_{q,\text{opt}}} = \left(\int_{\bar{S}_{q,\text{opt}}} |p(x)| \, dx \right)^q \leq \left(\int_{-\infty}^{\infty} |p(x)| \, dx \right)^q = 1,$$

and the resulting $\|g\|_{\frac{1}{1-q},\bar{S}_{q,\text{opt}}}$ implies the upper bound of $T_q[p; \lambda_q]$ if $\lambda_{q,\text{opt}}$ exists for a given q and σ .

Next, we construct a maximizer p_{opt} achieving this bound and show its uniqueness, which is done by checking the conditions where all three “≤” in (34) become “=”. As for the “≤” in (34b), it becomes “=” if and only if $p(x)$ becomes positive only in $\bar{S}_{q,\text{opt}}$. Namely,

$$f(x) = p^q(x) \geq 0 \ (x \in \bar{S}_{q,\text{opt}}) \quad \text{and} \quad f(x) = 0 \ (\text{a.e. } x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}). \quad (36)$$

In other words, the “≤” in (34b) becomes “<” if the above condition (36) is violated, which is easily verified from the graph of $g(x)$ and the above argument for the “≤” in (34b). On the other hand, in (34c), the “≤” becomes “=” if and only if the equality condition (10) is satisfied for $x \in \bar{S}_{q,\text{opt}}$, in other words,

$$Ap(x) = B \left| \lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2 \right|^{\frac{1}{1-q}}. \quad (37)$$

If p satisfies (36), it is immediate that $A \neq 0$ and $B \neq 0$ in (37), since $[\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{1}{1-q}} \geq 0 \ (x \in \bar{S}_{q,\text{opt}})$, and A and B are not both 0 (cf. [40], p. 140). The conclusion is that, if these two conditions (36) and (37) are satisfied, a maximizer p_{opt} achieving the upper bound of $T_q[p; \lambda]$ is uniquely determined:

$$p_{\text{opt}}(x) = \begin{cases} \frac{A}{B} [\lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2]^{\frac{1}{1-q}} = \frac{1}{Z_q} \exp_q(-\beta_q x^2) & (x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases} \quad (38)$$

in which A/B , $\lambda_{q,\text{opt}}$, $Z_q = \sqrt{\frac{3-q}{1-q}} \sigma B \left(\frac{1}{1-q} + 1, \frac{1}{2} \right)$, $\beta_q = \frac{1}{(3-q)\sigma^2}$, and $\bar{S}_{q,\text{opt}} = \left[-\sqrt{\frac{3-q}{1-q}} \sigma, \sqrt{\frac{3-q}{1-q}} \sigma \right]$ are uniquely determined, as shown in the following calculations. We note the formula ([42], p. 253):

$$\int_0^1 x^\alpha (1-x^\gamma)^\beta dx = \frac{1}{\gamma} B \left(1 + \beta, \frac{1+\alpha}{\gamma} \right)$$

is repeatedly used for the integrations below. (More precisely, for (39), we set $\alpha = 0 \ (> -1)$, $\beta = \frac{1}{1-q} \ (> -1)$, and $\gamma = 2 \ (> 0)$. For (40b), we set $\alpha = 2, 0 \ (> -1)$, $\beta = \frac{q}{1-q} \ (> -1)$, and $\gamma = 2 \ (> 0)$.) By substituting $p_{\text{opt}}(x)$ in (38) into the constraint (3b), we have

$$\langle p_{\text{opt}}(x) \rangle = \frac{A}{B} \left[(1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{1}{1-q}} \sqrt{\frac{\lambda_{q,\text{opt}} - 1}{\lambda_{q,\text{opt}}}} \sigma B \left(\frac{1}{1-q} + 1, \frac{1}{2} \right) = 1. \quad (39)$$

On the other hand, by substituting (38) into $\langle p^q(x) \rangle$ and $\langle x^2 p^q(x) \rangle$ in the constraint (3c), we have

$$\langle p_{\text{opt}}^q(x) \rangle = \left(\frac{A}{B} \right)^q \left[(1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{q}{1-q}} \sigma \left[\sqrt{\frac{\lambda_{q,\text{opt}} - 1}{\lambda_{q,\text{opt}}}} B \left(\frac{1}{1-q}, \frac{1}{2} \right) \right], \quad (40a)$$

$$\langle x^2 p_{\text{opt}}^q(x) \rangle = \left(\frac{A}{B} \right)^q \left[(1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{q}{1-q}} \sigma^3 \left[\left(\frac{\lambda_{q,\text{opt}} - 1}{\lambda_{q,\text{opt}}} \right)^{\frac{3}{2}} B \left(\frac{1}{1-q}, \frac{3}{2} \right) \right], \quad (40b)$$

and substitution of (40a) and (40b) into (3c) yields

$$\left(\frac{A}{B} \right)^q \left[(1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{1}{1-q}} \sigma^3 \left[\left(\frac{\lambda_{q,\text{opt}} - 1}{\lambda_{q,\text{opt}}} \right)^{\frac{3}{2}} B \left(\frac{1}{1-q}, \frac{3}{2} \right) - \sqrt{\frac{\lambda_{q,\text{opt}} - 1}{\lambda_{q,\text{opt}}}} B \left(\frac{1}{1-q}, \frac{1}{2} \right) \right] = 0. \quad (41)$$

Then, using the formula $B(x, y+1) = \frac{y}{x+y} B(x, y)$ ([42], p. 254) in (41), $\lambda_{q,\text{opt}}$ is uniquely determined as

$$\frac{\lambda_{q,\text{opt}}}{\lambda_{q,\text{opt}} - 1} = \frac{1-q}{3-q}, \text{ i.e., } \lambda_{q,\text{opt}} = \frac{1}{2}(q-1) \quad (42)$$

and from (39) and (42) A/B is uniquely determined as

$$\frac{A}{B} = \left[\frac{3-q}{2} \sigma^2 \right]^{\frac{1}{q-1}} \left[\sqrt{\frac{3-q}{1-q}} \sigma B \left(\frac{1}{1-q} + 1, \frac{1}{2} \right) \right]^{-1}.$$

In (38), equating the second term to the third term, β_q is uniquely determined as

$$\beta_q = \frac{\lambda_{q,\text{opt}}}{1 - \lambda_{q,\text{opt}}} (q-1)^{-1} \sigma^{-2} = \frac{1}{(3-q)\sigma^2}, \quad (43)$$

and from (29) and (30), Z_q is uniquely determined as

$$Z_q = \left\{ \frac{A}{B} \left[(1 - \lambda_{q,\text{opt}}) \sigma^2 \right]^{\frac{1}{1-q}} \right\}^{-1} = \sqrt{\frac{3-q}{1-q}} \sigma B \left(\frac{1}{1-q} + 1, \frac{1}{2} \right). \quad (44)$$

Thus, from (43), and (44), p_{opt} is uniquely obtained as in (38).

To see that p_{opt} makes all “ \leq ” in (34) “=”, finally, we check the last “ \leq ” in (34d) becomes “=”, which is immediate since in (34c) $\|f\|_{\frac{1}{q}, \bar{S}_{q,\text{opt}}} = \int_{\bar{S}_{q,\text{opt}}} p_{\text{opt}}(x) dx = 1$. Therefore, it is concluded that $T_q[p; \lambda_q]$ is uniquely maximized by p_{opt} in (38) for $0 < q < 1$. \square

4.3. Proof of Lemma 1 and Theorem 3

Let p be arbitrary feasible solutions to **T2** for $q = 0$, and let p_{opt} be its optimal solution, which is eventually constructed in (53). Let $\lambda_{q,\text{opt}}$ be a particular value of the additional parameter λ_q in **T2**, which is associated with p_{opt} and is eventually constructed in (54). Then, for any p and a particular $\lambda_{q,\text{opt}}$ ($= -\frac{1}{2}$ in (54)), we define f and g as

$$f(x) = p^0(x) = \begin{cases} 1 & \text{if } 0 < p(x) < \infty \\ 0 & \text{if } p(x) = 0, \end{cases} \quad (45a)$$

$$g(x) = \lambda_{q,\text{opt}} x^2 + (1 - \lambda_{q,\text{opt}}) \sigma^2, \quad (45b)$$

and we define an interval

$$\bar{S}_{q,\text{opt}} = \left[-\sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}}\sigma, \sqrt{\frac{\lambda_{q,\text{opt}}-1}{\lambda_{q,\text{opt}}}}\sigma \right]. \quad (46)$$

In (45a), as a convention, we take $0^0 = 0$, and $p(x) < \infty$ (a.e. $x \in \bar{S}_{q,\text{opt}}$). Then, $\|f\|_\infty = 1$ follows from (45a). Now, we define f_* as

$$f_*(x) = \begin{cases} \text{sgn}[g(x)] & (\text{a.e. } x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}). \end{cases} \quad (47)$$

We note this particular $f_*(x) = \text{sgn}[g(x)]$ is proved to be the unique maximizer of $\|fg\|_{1,\bar{S}_{q,\text{opt}}}$ in the following Lemma 1 (as a minor modification of Lemma 4 in [35]).

Lemma 1. (cf. Lemma 4 in [35]). Let S be an arbitrary subset in \mathbb{R} , with $\mu(S) > 0$. For $f \in L^\infty(S)$ and $g \in L^1(S)$, assume $g(x) \neq 0$, a.e. on S . Then, $f_*(x) = \text{sgn}[g(x)]$ (a.e. $x \in S$) is the unique maximizer of the functional $\|fg\|_{1,S}$ in (9).

Proof of Lemma 1. First, thanks to Hölder's inequality, see (9), $\|fg\|_{1,S}$ is maximized by f_* , since

$$\|f_*g\|_{1,S} = \langle \text{sgn}[g(x)] g(x) \rangle_S = \langle |g(x)| \rangle_S = \|g\|_{1,S}.$$

Second, the unique representation of this maximizer f_* is shown by proof by contradiction, as follows. Suppose another maximizer \bar{f}_* exists and it maximizes $\|fg\|_{1,S}$, in other words, $\|\bar{f}_*g\|_{1,S} = \|g\|_{1,S}$. Then, for any given $g \in L^1(S)$, the following is satisfied:

$$\|f_*g\|_{1,S} - \|\bar{f}_*g\|_{1,S} = 0. \quad (48)$$

Now, using the identities $f_*(x)g(x) = \text{sgn}[g(x)] \cdot g(x) \geq 0$ and $|g(x)| = f_*(x)g(x)$, we obtain $|f_*(x)g(x)| - |\bar{f}_*(x)g(x)| = f_*(x)g(x) - |\bar{f}_*(x)||g(x)|$ and $f_*(x)g(x) - |\bar{f}_*(x)||g(x)| = f_*(x)g(x) - |\bar{f}_*(x)|f_*(x)g(x)$, respectively, resulting in the equality

$$|f_*(x)g(x)| - |\bar{f}_*(x)g(x)| = f_*(x)g(x) - |\bar{f}_*(x)|f_*(x)g(x). \quad (49)$$

Substituting (49) into the left-hand side of (48) and using $|g(x)| = f_*(x)g(x)$, (48) is rewritten as

$$\langle (1 - |\bar{f}_*(x)|) f_*(x) g(x) \rangle_S = \langle (1 - |\bar{f}_*(x)|) |g(x)| \rangle_S = 0. \quad (50)$$

Now, keeping $0 \leq \bar{f}_*(x) \leq 1$ and the assumption that $g(x) \neq 0$, a.e. on S in mind, (50) implies

$$\begin{aligned} |\bar{f}_*(x)| &= 1, \quad \text{or equivalently} \\ \bar{f}_*(x) &= \sigma(x), \quad \text{a.e. on } S, \end{aligned}$$

where σ takes either -1 or 1 . However, among such functions \bar{f}_* having either -1 or 1 values, it is clear that $\text{sgn}[g(x)] (= f_*)$ is the only one that makes $\langle fg \rangle_S$ maximal. Thus, no \bar{f}_* can exist except for f_* , and the uniqueness of the maximizer f_* is verified. \square

Proof of Theorem 3. First, we show that $T_q[p; \lambda_q]$ is maximized in the following way:

$$T_q[p; \lambda_q] = T_q[p; \lambda_{q,\text{opt}}] = \langle fg \rangle \quad (51a)$$

$$\leq \langle fg \rangle_{\bar{S}_{q,\text{opt}}} = \langle |fg| \rangle_{\bar{S}_{q,\text{opt}}} \quad (51b)$$

$$= \|fg\|_{1,\bar{S}_{q,\text{opt}}} \leq \|f\|_{\infty,\bar{S}_{q,\text{opt}}} \|g\|_{1,\bar{S}_{q,\text{opt}}} \quad (51c)$$

$$\leq \|g\|_{1,\bar{S}_{q,\text{opt}}} \quad (\text{the upper bound}), \quad (51d)$$

where $\langle \cdot \rangle = \int_{-\infty}^{\infty} \cdot dx$, $\langle \cdot \rangle_{\bar{S}_{q,\text{opt}}} = \int_{\bar{S}_{q,\text{opt}}} \cdot dx$, and $\| \cdot \|_{1,\bar{S}_{q,\text{opt}}} = \int_{\bar{S}_{q,\text{opt}}} | \cdot | dx$, and $\| f \|_{\infty,\bar{S}_{q,\text{opt}}}$ is the infinity norm of $f(x)$ ($x \in \bar{S}_{q,\text{opt}}$), in other words, the essential supremum of $|f(x)|$ ($x \in \bar{S}_{q,\text{opt}}$). The first “=” in (51a) follows from the fact that $T_q[p; \lambda_q] = \sigma^2 \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2)p^q(x) \rangle$ in (3a) is independent from the value of λ_q , since any feasible solution $p(x)$ satisfies $\langle (x^2 - \sigma^2)p^q(x) \rangle = 0$ in (3c), and the second “=” in (51a) is immediate from (45). The “ \leq ” in (51b) is obtained from the same argument of the inequality (34b) and (35) in the proof of Theorem 2. On the other hand, the equality in (51b) is immediate from $f(x)g(x) \geq 0$ ($\forall x \in \bar{S}_{q,\text{opt}}$). The first “=” in (51c) follows from the definition of $\| \cdot \|_{1,\bar{S}_{q,\text{opt}}}$, and the “ \leq ” in (51c) follows from the Hölder’s inequality (9). The final “ \leq ” in (51d) follows from the definition of $f(x)$ in (45a), in other words, $\| f \|_{\infty,\bar{S}_{q,\text{opt}}} \leq 1$, and the resulting $\| g \|_{1,\bar{S}_{q,\text{opt}}}$ implies the upper bound of $T_q[p; \lambda_q]$ if $\lambda_{q,\text{opt}}$ exists for given q and σ .

Next, we construct a maximizer $p_{\text{opt}}(x)$ achieving this bound and show its uniqueness, which is done by checking the conditions where all three “ \leq ” in (51) become “=” . As for the first “ \leq ” in (51b), it becomes “=” if and only if $p(x)$ becomes positive only in $\bar{S}_{q,\text{opt}}$. Namely,

$$f(x) = p^0(x) = 0 \text{ or } 1 \text{ (} x \in \bar{S}_{q,\text{opt}} \text{) and } f(x) = 0 \text{ (a.e. } x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}} \text{),} \quad (52)$$

in other words, the “ \leq ” in (51b) becomes “ $<$ ” if the above condition (52) is violated, which is easily verified from the graph of $g(x)$ and the above argument for the “ \leq ” in (51b). On the other hand, in (51c), the second “ \leq ” becomes “=” if and only if $f(x) = \text{sgn}[g(x)]$ (a.e. $x \in \bar{S}_{q,\text{opt}}$) due to Lemma 1 (simply by replacing S with $\bar{S}_{q,\text{opt}}$ in Lemma 1). The final “ \leq ” in (51d) becomes “=” if and only if $\| f \|_{\infty,\bar{S}_{q,\text{opt}}} = 1$ in (51c). From these three conditions, f is uniquely determined as f_* in (47), and from (45a) the associated maximizer p_{opt} for $q = 0$ is obtained as

$$p_{\text{opt}}(x) = \begin{cases} \text{arbitrary positive value} & (x \in \bar{S}_{q,\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \bar{S}_{q,\text{opt}}), \end{cases} \quad (53)$$

where $p_{\text{opt}}(x)$ should satisfy $\langle p_{\text{opt}}(x) \rangle_{\bar{S}_{q,\text{opt}}} = 1$. Finally, substituting (53) into the constraint (3c), we have $\langle (x^2 - \sigma^2)p_{\text{opt}}^0(x) \rangle = \langle x^2 - \sigma^2 \rangle_{\bar{S}_{q,\text{opt}}} = 0$, and from (46) $\lambda_{q,\text{opt}}$ and $\bar{S}_{q,\text{opt}}$ are uniquely obtained as

$$\lambda_{q,\text{opt}} = -\frac{1}{2} (< 0) \quad \text{and} \quad \bar{S}_{q,\text{opt}} = [-\sqrt{3}\sigma, \sqrt{3}\sigma], \quad (54)$$

respectively. This shows the uniqueness of the representation of p_{opt} in (53). (This exceptional case $q = 0$ is also argued in T1, where the same result is obtained through a more direct graphical argument, after proving that any candidate $p_*(x)$ for the maximizer $p_{\text{opt}}(x)$ is defined only on a simply connected interval S_* that is symmetric about the origin O. The proof is straightforward but lengthy, and we omit it here.) \square

4.4. Proof of Theorem 4

Proof. Let p be arbitrary feasible solutions to R2 for $q > 1$, and let p_{opt} be its optimal solution, which is eventually constructed in (61). Let $\lambda_{q,\text{opt}}$ be a particular value of the additional parameter λ_q in R2, which is associated with p_{opt} and is eventually constructed in (65). First, for any p and a particular $\lambda_{q,\text{opt}}$ in (65), we define f and g as

$$f(x) = p^q(x), \quad (55a)$$

$$g(x) = 1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x), \quad (55b)$$

and we define a set \underline{S}_q in \mathbb{R} :

$$\mathbb{S}_q = \left\{ x \mid p^q(x) + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) (= f(x)g(x)) > 0 \right\}. \quad (56)$$

Next, we introduce a subset \mathcal{Q} of the feasible solutions p ,

$$\mathcal{Q} = \left\{ p(x) \mid (5b), (5c), \text{ and } p^q(x) + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) \geq 0 \ (\forall x \in \mathbb{R}) \right\}, \quad (57)$$

which is proved to be non-empty in Appendix A.1.

First, we show that the following holds: if $p \in \mathcal{Q}$,

$$R_q[p; \lambda_q] = R_q[p; \lambda_{q,\text{opt}}] = \left\langle \left\langle p^q(x) + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) \right\rangle \right\rangle_{\mathbb{S}_q} \quad (58a)$$

$$= \left\langle \left\langle p^q(x) + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) \right\rangle \right\rangle_{\mathbb{S}_q} = \|fg\|_{1,\mathbb{S}_q} \quad (58b)$$

$$\geq \|f\|_{\frac{1}{q},\mathbb{S}_q} \|g\|_{\frac{1}{1-q},\mathbb{S}_q}, \quad (58c)$$

where $\langle \cdot \rangle = \int_{-\infty}^{\infty} \cdot dx$, $\langle \cdot \rangle_{\mathbb{S}_q} = \int_{\mathbb{S}_q} \cdot dx$, and $\|\cdot\|_{\alpha,\mathbb{S}_q} = \left(\int_{\mathbb{S}_q} |\cdot|^\alpha dx \right)^{\frac{1}{\alpha}}$. The first “=” in (58a) follows from the fact that $R_q[p; \lambda_q] = \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2)p(x) \rangle$ in (7c) is independent from the value of λ_q , since any feasible solution $p(x)$ satisfies $\langle (x^2 - \sigma^2)p(x) \rangle = 0$ in (6), and the second “=” in (58a) is immediate from the definitions (56) and (57). The first “=” in (58b) is also immediate from (56). The “ \geq ” in (58c) follows from reverse Hölder’s inequality (11).

If $R_q[p; \lambda_q]$ achieves the lower bound, and $\|f\|_{\frac{1}{q},\mathbb{S}_q} \|g\|_{\frac{1}{1-q},\mathbb{S}_q}$ in (58c) saturates at this bound for $p_{\text{opt}} (\in \mathcal{Q})$ and $\lambda_{q,\text{opt}}$, then, from (58), the following has to be satisfied:

$$R_q[p_{\text{opt}}; \lambda_{q,\text{opt}}] = \|f\|_{\frac{1}{q},\mathbb{S}_q} \|g\|_{\frac{1}{1-q},\mathbb{S}_q} = \text{the lower bound.}$$

Therefore, to construct a candidate p_{opt} achieving this bound, we consider the condition where the “ \geq ” in (58c) becomes “=” . Namely, we rewrite the equality condition (12) by using (55):

$$p^q(x) = A \left[1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x) \right]^{\frac{q}{1-q}} > 0 \quad (x \in \mathbb{S}_q), \text{ i.e.,} \quad (59a)$$

$$p^{1-q}(x) = C \left[1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x) \right] > 0 \quad (x \in \mathbb{S}_q), \quad (59b)$$

where $C = A^{\frac{1-q}{q}} > 0$. From (59b), it is immediate that

$$p^{1-q}(x) \left[1 + C\lambda_{q,\text{opt}}(\sigma^2 - x^2) \right] = C \quad (x \in \mathbb{S}_q), \quad (60)$$

and, hence,

$$p(x) = \left(\frac{1}{C} \right)^{\frac{1}{q-1}} \left[1 + C\lambda_{q,\text{opt}}(\sigma^2 - x^2) \right]^{\frac{1}{q-1}} = \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} \right]^{\frac{1}{q-1}} \quad (x \in \mathbb{S}_q),$$

since $[1 + C\lambda_{q,\text{opt}}(\sigma^2 - x^2)] > 0$ because $C > 0$ and $p^{1-q}(x) > 0$ ($x \in \mathbb{S}_q$) in (60). On the other hand, for $p \in \mathcal{Q}$, it is also immediate that $p(x) = 0$ ($x \in \mathbb{R} \setminus \mathbb{S}_q$) from (56) and (57). Thereby, a candidate of the minimizer (in \mathcal{Q}) is constructed as:

$$p_{\text{opt}}(x) = \begin{cases} \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} \right]^{\frac{1}{q-1}} = \frac{1}{Z_q^*} \exp_{q^*}(-\beta_q^* x^2) & (x \in \mathbb{S}_{\text{opt}}) \\ 0 & (x \in \mathbb{R} \setminus \mathbb{S}_{\text{opt}}), \end{cases} \quad (61)$$

where $q^* = 2 - q$. (The distribution in (61) is called the q^* -Gaussian [31].) In (61), $\mathcal{S}_{\text{opt}} = \left[-\sqrt{\sigma^2 + \frac{1}{C\lambda_{q,\text{opt}}}}, \sqrt{\sigma^2 + \frac{1}{C\lambda_{q,\text{opt}}}} \right]$ is now specified and $\lambda_{q,\text{opt}}$, Z_q^* , and β_q^* are uniquely determined as $\lambda_{q,\text{opt}} = \frac{1}{2C\sigma^2} \frac{q-1}{q} > 0$, $Z_q^* = \sqrt{\frac{3q-1}{q-1}} \sigma B\left(\frac{q}{q-1}, \frac{1}{2}\right)$, and $\beta_q^* = \frac{1}{(3q-1)\sigma^2}$, respectively, as shown in the following calculations. We note the formula ([42], p. 253):

$$\int_0^1 x^\alpha (1-x^\gamma)^\beta dx = \frac{1}{\gamma} B\left(1+\beta, \frac{1+\alpha}{\gamma}\right)$$

is repeatedly used for the integrations below. (More precisely, for (62) we set $\alpha = 0$ (> -1), $\beta = \frac{1}{q-1}$ (> -1), and $\gamma = 2$ (> 0).) For (63), we set $\alpha = 2$ (> -1), $\beta = \frac{1}{1-q}$ (> -1), and $\gamma = 2$ (> 0). Substituting (61) into the constraint (5b), we have

$$\langle\langle p_{\text{opt}}(x) \rangle\rangle = r^{\left(\frac{2}{q-1}+1\right)} \lambda_{q,\text{opt}}^{\frac{1}{q-1}} B\left(\frac{q}{q-1}, \frac{1}{2}\right) = 1, \quad (62)$$

where r is defined by $r^2 = \sigma^2 + \frac{1}{\lambda_{q,\text{opt}}C}$. Note that $r^2 = \frac{3q-1}{q-1}\sigma^2 > 0$ as shown in (65). On the other hand, substituting (61) into the constraint (5c), we have

$$\langle\langle x^2 p_{\text{opt}}(x) \rangle\rangle = r^{\left(\frac{2}{q-1}+3\right)} \lambda_{q,\text{opt}}^{\frac{1}{q-1}} B\left(\frac{q}{q-1}, \frac{3}{2}\right) = \sigma^2. \quad (63)$$

Substitution of (62) and (63) (multiplied with σ^2 to its both sides) into (7a) yields

$$r^2 B\left(\frac{q}{q-1}, \frac{3}{2}\right) = \sigma^2 B\left(\frac{q}{q-1}, \frac{1}{2}\right), \text{ i.e., } r^2 = \frac{3q-1}{q-1} \sigma^2 \left(= \sigma^2 + \frac{1}{\lambda_{q,\text{opt}}C}\right), \quad (64)$$

in which $B(x, y+1) = \frac{y}{x+y} B(x, y)$ ([42], p. 254) is used. Thus, $\lambda_{q,\text{opt}}$ is obtained as

$$\lambda_{q,\text{opt}} = \frac{1}{2C\sigma^2} \frac{q-1}{q}. \quad (65)$$

Since

$$r^2 = \sigma^2 + \frac{1}{\lambda_{q,\text{opt}}C} = \frac{3q-1}{q-1} \sigma^2.$$

While, from (62), or (63), and (64) that determines r with q and σ , $\lambda_{q,\text{opt}} (> 0)$ is uniquely obtained for any given $q (> 1)$ and σ , and hence, C is also uniquely determined from (65):

$$C = \frac{1}{2\sigma^2} \frac{q-1}{q} r^{q+1} B^{q-1}\left(\frac{q}{q-1}, \frac{1}{2}\right).$$

Next, we obtain Z_q^* and β_q^* from (61) and (65):

$$\begin{aligned} \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} \right]^{\frac{1}{q-1}} &= \left(\frac{3q-1}{2Cq} \right)^{\frac{1}{q-1}} \left[1 - \frac{1-q_*}{(3q-1)\sigma^2} x^2 \right]^{\frac{1}{q_*}} \\ &= \frac{1}{Z_q^*} \exp_{q_*} \left[-\frac{x^2}{(3q-1)\sigma^2} \right], \end{aligned}$$

which yields

$$\beta_q^* = \frac{1}{(3q-1)\sigma^2}, Z_q^* = \left(\frac{3q-1}{2Cq} \right)^{\frac{1}{1-q}} = \sqrt{\frac{3q-1}{q-1}} \sigma B\left(\frac{q}{q-1}, \frac{1}{2}\right).$$

Therefore, from (61), (62), and (65), $p_{\text{opt}}(x)$ is uniquely determined as in (61). Note that $p_{\text{opt}}(x) \in \mathcal{Q}$, since

$$1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2) p_{\text{opt}}^{1-q}(x) = \frac{1}{C} \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} \right]^{-1} > 0 \quad (x \in \mathbb{S}_{\text{opt}})$$

is immediate from (61).

Next, to prove that the candidate in (61) is the unique minimizer, we directly compare $\langle p^q(x) \rangle$ and $\langle p_{\text{opt}}^q(x) \rangle$ in the following way:

$$\begin{aligned} & \langle p^q(x) \rangle - \langle p_{\text{opt}}^q(x) \rangle \\ &= \left\langle p^q(x) - p_{\text{opt}}^q(x) + q\lambda_{q,\text{opt}}(x^2 - \sigma^2) [p(x) - p_{\text{opt}}(x)] \right\rangle - \frac{q}{C} \langle p(x) - p_{\text{opt}}(x) \rangle \end{aligned} \quad (66a)$$

$$= \left\langle p^q(x) + q\lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) - \frac{q}{C}p(x) \right\rangle_{\mathbb{R} \setminus \mathbb{S}_{\text{opt}}} \quad (66b)$$

$$+ \left\langle p^q(x) - p_{\text{opt}}^q(x) - qp_{\text{opt}}^{q-1}(x)[p(x) - p_{\text{opt}}(x)] \right\rangle_{\mathbb{S}_{\text{opt}}}. \quad (66c)$$

Note that (66a) follows from $\langle p_{\text{opt}}(x) \rangle = \langle p(x) \rangle = 1$ in (5b) and $\langle x^2 p_{\text{opt}}(x) \rangle = \langle x^2 p(x) \rangle = \sigma^2$ in (5c), (66b) follows from (61), in other words, $p_{\text{opt}}(x) = 0$ ($x \in \mathbb{R} \setminus \mathbb{S}_{\text{opt}}$), and (66c) follows from (66a) by using $q\lambda_{q,\text{opt}}(x^2 - \sigma^2) - \frac{q}{C} = -qp_{\text{opt}}^{q-1}(x)$ ($x \in \mathbb{S}_{\text{opt}}$), which is immediate from (61). Finally, substituting $\lambda_{q,\text{opt}} = \frac{1}{2C\sigma^2} \frac{q-1}{q}$ in (65) into (66b), we obtain

$$\left\langle p^q(x) + \frac{q-1}{2C\sigma^2} p(x) \left(x^2 - \frac{3q-1}{q-1} \sigma^2 \right) \right\rangle_{\mathbb{R} \setminus \mathbb{S}_{\text{opt}}} \geq 0,$$

in which the equality holds if and only if $p(x) = 0$ ($x \in \mathbb{R} \setminus \mathbb{S}_{\text{opt}}$), since

$$\mathbb{S}_{\text{opt}} = \left[-\sqrt{\sigma^2 + \frac{1}{C\lambda_{q,\text{opt}}}}, \sqrt{\sigma^2 + \frac{1}{C\lambda_{q,\text{opt}}}} \right] = \left[-\sqrt{\frac{3q-1}{q-1}} \sigma, \sqrt{\frac{3q-1}{q-1}} \sigma \right],$$

and hence,

$$\frac{q-1}{2C\sigma^2} \left(x^2 - \frac{3q-1}{q-1} \sigma^2 \right) > 0 \quad (x \in \mathbb{R} \setminus \mathbb{S}_{\text{opt}}).$$

On the other hand, the term (66c) can be expressed as

$$\left\langle p_{\text{opt}}^q(x) \left\{ \left[\frac{p(x)}{p_{\text{opt}}(x)} \right]^q - q \frac{p(x)}{p_{\text{opt}}(x)} + q-1 \right\} \right\rangle_{\mathbb{S}_{\text{opt}}}. \quad (67)$$

Because, for $q > 1$, $h(X) = X^q - qX + q-1 \geq 0$ for any $X \geq 0$, and because $h(X) = 0$ only when $X = 1$, (67) is nonnegative and it becomes 0 if and only if $X = 1$, in other words, $p(x) = p_{\text{opt}}(x)$ ($x \in \mathbb{S}_{\text{opt}}$). This proves that $p_{\text{opt}}(x)$ in (61) is the unique minimizer to **R2** for $q > 1$. \square

4.5. Proof of Theorem 5

Proof. Let p be arbitrary feasible solutions to **R2** for $\frac{1}{3} < q < 1$, and let p_{opt} be its optimal solution, which is eventually constructed in (72). Let $\lambda_{q,\text{opt}}$ be a particular value of the additional parameter λ_q in **R2**, which is associated with p_{opt} and is eventually constructed in (76). Then, for any p and a particular $\lambda_{q,\text{opt}}$ in (76), we define f and g as

$$f(x) = p^q(x), \quad (68a)$$

$$g(x) = 1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x). \quad (68b)$$

First, we show the following holds for any feasible solution p ,

$$R_q[p; \lambda_q] = R_q[p; \lambda_{q,\text{opt}}] = \langle f, g \rangle \quad (69a)$$

$$\leq \|fg\|_1 \leq \|f\|_{\frac{1}{q}} \|g\|_{\frac{1}{1-q}} \quad (69b)$$

$$= \|g\|_{\frac{1}{1-q}}. \quad (69c)$$

The first “=” in (69a) follows from the fact that $R_q[p; \lambda_q] = \langle p^q(x) \rangle + \lambda_q \langle (x^2 - \sigma^2)p(x) \rangle$ in (7c) is independent from the value of λ_q , since any feasible solution $p(x)$ satisfies $\langle (x^2 - \sigma^2)p(x) \rangle = 0$ in (6), and the second “=” in (69a) is immediate from (68). The first “≤” in (69b) follows from $f(x)g(x) \leq |f(x)g(x)|$ ($\forall x \in \mathbb{R}$), since $f(x)$ is always nonnegative but $g(x)$ in (68b) can be negative on some intervals in \mathbb{R} by choosing certain $p(x)$. The second “≤” in (69b) follows from Hölder’s inequality (9). The final “=” in (69c) follows from $\|f\|_{\frac{1}{q}} = \|p^q\|_{\frac{1}{q}} = \langle |p(x)| \rangle^q = 1$ in (69b).

Next, if $R_q[p; \lambda_q]$ achieves the upper bound, and $\|g\|_{\frac{1}{1-q}}$ in (69c) saturates at this bound for p_{opt} and $\lambda_{q,\text{opt}}$, then from (69) the following has to be satisfied:

$$R_q[p_{\text{opt}}; \lambda_{q,\text{opt}}] = \|fg\|_1 = \|f\|_{\frac{1}{q}} \|g\|_{\frac{1}{1-q}} = \text{the upper bound.}$$

Therefore, to construct a candidate p_{opt} achieving this bound, we consider the condition where the two “≤” in (69) become “=”. As for the first “≤” in (69b), it becomes “=” if $g(x) > 0$ (a.e. $x \in \mathbb{R}$), in other words,

$$1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x) > 0 \quad (\text{a.e. } x \in \mathbb{R}), \quad (70)$$

which is eventually verified in (78). On the other hand, the second “≤” in (69b) becomes “=” if and only if the equality condition (10) is satisfied, $Ap(x) = B|1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x)|^{\frac{1}{1-q}}$, in other words,

$$p^{1-q}(x) = C \left| 1 + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p^{1-q}(x) \right| \quad (\forall x \in \mathbb{R}), \quad (71)$$

where $C = (B/A)^{1-q} > 0$. Note that $A \neq 0$ and $B \neq 0$ because of (70). From (70) and (71), a candidate of the maximizer is uniquely constructed:

$$p_{\text{opt}}(x) = \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} \right]^{\frac{-1}{1-q}} = \frac{1}{Z_q^*} \exp_{q^*} \left(-\beta_q^* x^2 \right) \quad (\forall x \in \mathbb{R}), \quad (72)$$

in which $C = (B/A)^{1-q}$, $\lambda_{q,\text{opt}} = \frac{1}{2C\sigma^2} \frac{q-1}{q}$ (< 0), $Z_q^* = \sqrt{\frac{3q-1}{1-q}} \sigma B \left(\frac{1}{1-q} - \frac{1}{2}, \frac{1}{2} \right)$, and $\beta_q^* = \frac{1}{(3q-1)\sigma^2}$ are uniquely determined, and $\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} > 0$ is verified, as shown in the following calculations. We note the formula ([42], p. 253):

$$\int_0^\infty \frac{dx}{x^\alpha(1+x^\gamma)^\beta} = \frac{1}{\gamma} B\left(\beta - \frac{1-\alpha}{\gamma}, \frac{1-\alpha}{\gamma}\right)$$

is repeatedly used for the integrations below. (More precisely, for (73) we set $\alpha = 0$ (< 1), $\beta = \frac{1}{1-q}$ (> 0), $\gamma = 2$ (> 0), and $\beta\gamma > 1 - \alpha$ is satisfied.) For (74), we set $\alpha = -2$ (< 1), $\beta = \frac{1}{1-q}$ (> 0), $\gamma = 2$ (> 0), and $\beta\gamma > 1 - \alpha$ is satisfied. Substituting (72) into the constraint (5b), we have

$$\langle\langle p_{\text{opt}}(x) \rangle\rangle = r^{\left(\frac{2}{q-1}+1\right)} (-\lambda_{q,\text{opt}})^{\frac{1}{q-1}} B\left(\frac{1}{1-q} - \frac{1}{2}, \frac{1}{2}\right) = 1, \quad (73)$$

where r is defined by $r^2 = -\left(\sigma^2 + \frac{1}{\lambda_{q,\text{opt}}C}\right)$. Note that $r^2 = \frac{3q-1}{1-q}\sigma^2 > 0$ as shown in (76). On the other hand, substituting (72) into the constraint (5c), we have

$$\langle\langle x^2 p_{\text{opt}}(x) \rangle\rangle = r^{\left(\frac{2}{q-1}+3\right)} (-\lambda_{q,\text{opt}})^{\frac{1}{q-1}} B\left(\frac{1}{1-q} - \frac{3}{2}, \frac{3}{2}\right) = \sigma^2. \quad (74)$$

Substitution of (73) and (74) after multiplying σ^2 to both sides into (7a) yields

$$r^2 B\left(\frac{1}{1-q} - \frac{3}{2}, \frac{3}{2}\right) = \sigma^2 B\left(\frac{1}{1-q} - \frac{1}{2}, \frac{1}{2}\right), \text{ i.e., } r^2 = \frac{3q-1}{1-q}\sigma^2, \quad (75)$$

in which $B(x+1, y-1) = \frac{x}{y-1}B(x, y)$ ([42], p. 254) is used. Thus, we first obtain $\lambda_{q,\text{opt}}$ as

$$\lambda_{q,\text{opt}} = \frac{1}{2C\sigma^2} \frac{q-1}{q}, \quad (76)$$

since

$$r^2 = -\left(\sigma^2 + \frac{1}{\lambda_{q,\text{opt}}C}\right) = \frac{3q-1}{1-q}\sigma^2.$$

Meanwhile, from (73), or (74), and (75) that determines r with q and σ , $\lambda_{q,\text{opt}}$ (< 0) is uniquely determined for any q (< 1) and σ , and hence, C (> 0) is also uniquely determined from (76):

$$C = \frac{1}{2\sigma^2} \frac{1-q}{q} r^{q+1} B^{q-1}\left(\frac{1}{1-q} - \frac{1}{2}, \frac{1}{2}\right).$$

Next, we obtain Z_q^* and β_q^* from (72) and (76):

$$\begin{aligned} \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C}\right]^{\frac{-1}{1-q}} &= \left(\frac{3q-1}{2Cq}\right)^{\frac{1}{q-1}} \left[1 - \frac{1-q_*}{(3q-1)\sigma^2} x^2\right]^{\frac{1}{1-q_*}} \\ &= \frac{1}{Z_q^*} \exp_{q^*} \left[-\frac{x^2}{(3q-1)\sigma^2}\right], \end{aligned}$$

which yields

$$\beta_q^* = \frac{1}{(3q-1)\sigma^2}, \quad Z_q^* = \left(\frac{3q-1}{2Cq}\right)^{\frac{1}{1-q}} = \sqrt{\frac{3q-1}{1-q}} \sigma B\left(\frac{1}{1-q} - \frac{1}{2}, \frac{1}{2}\right).$$

Now, we verify (70) is satisfied by p_{opt} . Note that if $\frac{1}{3} < q < 1$, using (76) we have

$$\lambda_{q,\text{opt}}(\sigma^2 - x^2) + \frac{1}{C} = \frac{1}{C} \left(\frac{1-q}{2\sigma^2 q} x^2 + \frac{3q-1}{2q}\right) > 0 \quad (77)$$

and hence (70) is satisfied by p_{opt} :

$$1 + \lambda_{q,\text{opt}} (x^2 - \sigma^2) p_{\text{opt}}^{1-q}(x) = \frac{1}{C} \left[\lambda_{q,\text{opt}} (\sigma^2 - x^2) + \frac{1}{C} \right]^{-1} > 0 \quad (\text{a.e. } x \in \mathbb{R}), \quad (78)$$

which is immediate from (71), (72), and (77). Thus, p_{opt} is uniquely determined as in (72).

Finally, to prove that this candidate (72) is the unique maximizer, we directly compare $\langle\langle p^q(x) \rangle\rangle$ and $\langle\langle p_{\text{opt}}^q(x) \rangle\rangle$ as follows. Similar to (66), the following holds here:

$$\begin{aligned} & \langle\langle p^q(x) \rangle\rangle - \langle\langle p_{\text{opt}}^q(x) \rangle\rangle \\ &= \langle\langle p^q(x) - p_{\text{opt}}^q(x) + q\lambda_{q,\text{opt}}(x^2 - \sigma^2)[p(x) - p_{\text{opt}}(x)] \rangle\rangle - \frac{q}{C} \langle\langle p(x) - p_{\text{opt}}(x) \rangle\rangle \end{aligned} \quad (79a)$$

$$= \langle\langle p^q(x) - p_{\text{opt}}^q(x) - qp_{\text{opt}}^{q-1}(x)[p(x) - p_{\text{opt}}(x)] \rangle\rangle. \quad (79b)$$

Note that (79a) follows from $\langle\langle p_{\text{opt}}(x) \rangle\rangle = \langle\langle p(x) \rangle\rangle = 1$ in (5b) and $\langle\langle x^2 p_{\text{opt}}(x) \rangle\rangle = \langle\langle x^2 p(x) \rangle\rangle = \sigma^2$ in (5c), and (79b) follows from (79a) by using $q\lambda_{q,\text{opt}}(x^2 - \sigma^2) - \frac{q}{C} = -qp_{\text{opt}}^{q-1}(x)$ ($x \in \mathbb{R}$), which is immediate from (72). Because for $0 < q < 1$, $h(X) = X^q - qX + q - 1 \leq 0$ for any $X \geq 0$ and because $h(X) = 0$ only when $X = 1$, (79b) is not positive, and it becomes 0 if and only if $X = 1$, in other words, $p(x) = p_{\text{opt}}(x)$ ($x \in \mathbb{R}$). This proves that $p_{\text{opt}}(x)$ in (72) is the unique maximizer to **R2** for $\frac{1}{3} < q < 1$. \square

5. Conclusions and discussion

We obtained a new insight about a direct link between generalized entropy and Hölder's inequality, and yet another proof for Rényi–Tsallis entropy maximization; the q -Gaussian distribution is directly obtained from the equality condition of Hölder's inequality, and its optimality is proved by Hölder's inequality through Moriguti's argument. The simplicity in the proofs of Tsallis entropy maximization (Theorem 1, 2, and 3) is worth noting; essentially, several lines of inequalities (including Hölder's inequality) are sufficient for the proof.

As an analogy, what we have described in this study can be explained as mountain climbing; as for Tsallis entropy maximization, the top of the mountain, in other words, the upper/lower bound is clearly seen from the starting point. Namely, the bounds in (24c), (34d), and (51d) are explicitly given by q and σ . Therefore, all we need to do is to keep climbing to the top, in other words, to construct a series of inequalities (24), (34), and (51) that saturate at the bound. On the other hand, for Rényi entropy maximization, the top of the mountain is not clearly seen from the starting point. Namely, the upper/lower bound is not given only by q and σ but contains $p(x)$, as in (58c) or (69c). Even in such a case, Hölder's inequality is still useful for finding a peak of the mountain, in other words, it leads to a candidate of the global optimal, and then we verify this candidate is really the top by using a GPS (global positioning system). In addition, this GPS is obtained as in (66) or (79), thanks to Moriguti [5].

Our technique with Hölder's inequality plus the additional parameter λ_q can be useful for other inequalities (e.g., Young's inequality), and it seems an interesting open problem to clarify what sort of optimization problems can be solved from such a technique.

Author Contributions: Conceptualization, H.-A.T.; methodology, H.-A.T.; writing—original draft preparation, H.-A.T.; writing—review and editing, H.-A.T. and M.N.; supervision, Y.O.; funding acquisition, H.-A.T.

Funding: This work has been supported by the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT) (Grant No. 26286086) and by the Support Center for Advanced Telecommunications Technology Research (SCAT).

Acknowledgments: The authors would like to express their gratitude to the referees for their careful reading of an earlier draft of this paper and valuable suggestions for improving the paper. The authors are indebted to Hiroki Suyari, Makoto Tsukada, Hideki Takayasu, Hayato Waki, Hayato Chiba, Yutaka Jitsumatsu, Fumito Mori, Yasuhiro Tsubo, Takashi Shimada, Akitoshi Takayasu, Masahide Kashiwagi, and Shinichi Oishi for their enlightening suggestions. One of the authors (H. T.) appreciates Norikazu Takahashi for his critical reading of the

manuscript and valuable suggestions which improved the manuscript. H. T. appreciates Jürgen Kurths and Istvan Z. Kiss for their inspiring suggestion that motivated this work, and H. T. also appreciates Constantino Tsallis for his critical comments at Social Modeling and Simulations + Econophysics Colloquium (SMSEC2014). One of the authors (H. T.) would like to dedicate this work to the memory of Sigeiti Moriguti.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Example of Non-Empty set \mathcal{Q}

Here, we illustrate an example of \mathcal{Q} introduced in Section 4.4. Having obtained p_{opt} and $\lambda_{q,\text{opt}}$ in Section 4.4, an element $p \in \mathcal{Q}$, which satisfies (5b), (5c), and

$$p^q(x) + \lambda_{q,\text{opt}}(x^2 - \sigma^2)p(x) \geq 0 \quad (\forall x \in \mathbb{R}) \quad (\text{A1})$$

is constructed from p_{opt} in (61) as follows. Figure A1 shows how we are going to construct p from p_{opt} ; the basic idea is that such p is obtained only by slightly modifying p_{opt} at its edge while keeping the constraint (A1). First, we choose small adjacent intervals I_1 and I_2 inside the interval $\left[\sigma, \sqrt{\frac{3q-1}{q-1}}\sigma\right]$. This choice is consistent to the fact that the constraint (A1) is equivalent to

$$p(x) \geq \left[\lambda_{q,\text{opt}}(\sigma^2 - x^2), 0\right]_+^{\frac{1}{q-1}} \quad (\text{as shown by the red dotted line in Figure A1}),$$

and hence $p(x)$ can be 0 in $\left[\sigma, \sqrt{\frac{3q-1}{q-1}}\sigma\right]$ (as observed in the inset of Figure A1). Second, we shift I_1 , I_2 , and the associated value of $p_{\text{opt}}(x)$ originally defined on I_1 and I_2 , altogether, while keeping the original value of $\langle (x^2 - \sigma^2)p(x) \rangle$ at 0. As shown in the inset of Figure A1, an option for this shift is: I_1 to the right and I_2 to the left. Such an option for small shifts always exists because of the continuity of integration $\langle (x^2 - \sigma^2)p(x) \rangle$ with respect to I_1 and I_2 . Note that the resulting p shown in the inset of Figure A1 satisfies (5b), (5c), and (A1). The above constructed p constitutes a non-empty set \mathcal{Q} and it is straightforwardly verified to be convex.

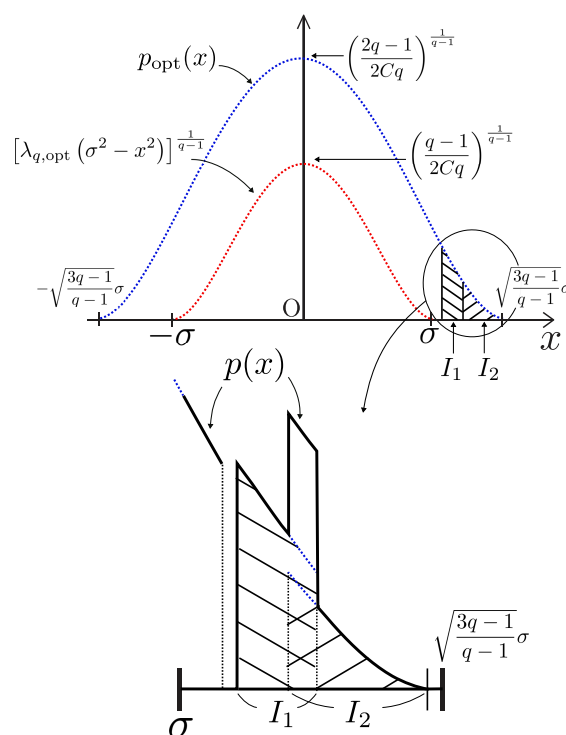


Figure A1. Construction of $p \in \mathcal{Q}$ from p_{opt} .

References

1. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 200; ISBN 978-0-387-85358-1.
2. Prato, D.; Tsallis, C. Nonextensive foundation of Lévy distributions. *Phys. Rev. E* **1999**, *60*, 2398–2401, doi:10.1103/PhysRevE.60.2398.
3. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; ISBN 978-0-471-24195-9.
4. Daróczy, Z. Generalized information functions. *Inf. Control* **1970**, *16*, 36–51.
5. Moriguti, S. A lower bound for a probability moment of any absolutely continuous distribution with finite variance. *Ann. Math. Stat.* **1952**, *23*, 286–289.
6. Campbell, L.L. A coding theorem and Rényi's entropy. *Inf. Control* **1965**, *8*, 423–429.
7. Baer, M.B. Source coding for quasarithmetic penalties. *IEEE Trans. Inf. Theory* **2006**, *52*, 4380–4393, doi:10.1109/TIT.2006.881728.
8. Bercher, J.-F. Source coding with scaled distributions and Rényi entropy bounds. *Phys. Lett. A* **2009**, *373*, 3235–3238, doi:10.1016/j.physleta.2009.07.015.
9. Bunte, C.; Lapidoth, A. Encoding tasks and Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 5065–5076, doi:10.1109/TIT.2014.2329490.
10. Landsberg, P.T.; Vedral, V. Distributions and channel capacities in generalized statistical mechanics. *Phys. Lett. A* **1998**, *247*, 211–217, doi:10.1016/S0375-9601(98)00500-3.
11. Ilić, V.M.; Djordjević, I.B.; Küeppers, F. On the Daróczy–Tsallis capacities of discrete channels. In *Sciforum Electronic Conference Series, Proceedings of the 2nd International Electronic Conference on Entropy and Its Applications, 15–30 November 2015*; MDPI: Basel, Switzerland, 2015; B004, pp. 1–11.
12. Venkatesan, R.C.; Plastino, A. Generalized statistics framework for rate distortion theory. *Phys. A* **2009**, *388*, 2337–2353, doi:10.1016/j.physa.2009.02.003.
13. Girardin, V.; Lhote, L. Rescaling entropy and divergence rates. *IEEE Trans. Inf. Theory* **2015**, *61*, 5868–5882, doi:10.1109/TIT.2015.2476486.
14. Thistleton, W.J.; Marsh, J.A.; Nelson, K.; Tsallis, C. Generalized Box–Müller method for generating q -Gaussian random deviates. *IEEE Trans. Inf. Theory* **2007**, *53*, 4805–4810, doi:10.1109/TIT.2007.909173.
15. Umeno, K.; Sato, A. Chaotic method for generating q -Gaussian random variables. *IEEE Trans. Inf. Theory* **2013**, *59*, 3199–3209, doi:10.1109/TIT.2013.2241174.
16. Karmeshu; Sharma, S. Queue length distribution of network packet traffic: Tsallis entropy maximization with fractional moments. *IEEE Commun. Lett.* **2006**, *10*, 34–36, doi:10.1109/LCOMM.2006.1576561.
17. Sharma, S.; Karmeshu. Power law characteristic and loss probability: finite buffer queueing systems. *IEEE Commun. Lett.* **2009**, *13*, 971–973, doi:10.1109/LCOMM.2009.12.091768.
18. Singh, A.K.; Karmeshu. Power law behavior of queue size: Maximum entropy principle with shifted geometric mean constraint. *IEEE Commun. Lett.* **2014**, *18*, 1335–1338, doi:10.1109/LCOMM.2014.2331292.
19. Singh, A.K.; Singh, H.P.; Karmeshu. Analysis of finite buffer queue: Maximum entropy probability distribution with shifted fractional geometric and arithmetic means. *IEEE Commun. Lett.* **2015**, *19*, 163–166, doi:10.1109/LCOMM.2014.2377236.
20. Jaynes, E.T. On the rationale of maximum-entropy methods. *Proc. IEEE* **1982**, *70*, 939–952, doi:10.1109/PROC.1982.12425.
21. Shore, J.E.; Johnson, R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37, doi:10.1109/TIT.1980.1056144.
22. Shore, J.E.; Johnson, R.W. Properties of cross-entropy minimization. *IEEE Trans. Inf. Theory* **1981**, *27*, 427–482, doi:10.1109/TIT.1981.1056373.
23. Tsukada, M.; Suyari, H.; Kato, M. On the probability distribution maximizing generalized entropies. In *Proceedings of 2005 Symposium on Applied Functional Analysis: Information Sciences and Related Fields*; Murohashi, T., Takahashi, W., Tsukada, M., Eds.; Yokohama Publisher: Yokohama, Japan, 2007; pp. 99–111.
24. Furuichi, S. On the maximum entropy principle and the minimization of the Fisher information in Tsallis statistics. *J. Math. Phys.* **2009**, *50*, 013303:1–013303:13, doi: 10.1063/1.3063640.
25. Lutwak, E.; Yang, D.; Zhang, G. Cramér–Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Trans. Inf. Theory* **2005**, *51*, 473–478, doi:10.1109/TIT.2004.840871.

26. Lutwak, E.; Yang, D.; Zhang, G. Moment-entropy inequalities for a random vector. *IEEE Trans. Inf. Theory* **2007**, *53*, 1603–1607, doi:10.1109/TIT.2007.892780.
27. Eguchi, S.; Komori, O.; Kato, S. Projective power entropy and maximum Tsallis entropy distributions. *Entropy* **2011**, *13*, 1746–1764, doi:10.3390/e13101746.
28. Watanabe, S.; Oohama, Y. Secret key agreement from vector Gaussian sources by rate limited public communication. *IEEE Trans. Inf. Forensic Secur.* **2011**, *6*, 541–550, doi:10.1109/ISIT.2010.5513738.
29. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810, doi:10.1109/TIT.2014.2357799.
30. Sakai, Y.; Iwata, K. Sharp bounds on Arimoto's conditional Rényi entropies between two distinct orders. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2985–2989, doi:10.1109/ISIT.2017.8007075.
31. Suyari, H.; Tsukada, M. Law of error in Tsallis statistics. *IEEE Trans. Inf. Theory* **2005**, *51*, 753–757, doi:10.1109/TIT.2004.840862.
32. Vignat, C.; Hero III, A.O.; and Costa, J.A. About closedness by convolution of the Tsallis maximizers. *Phys. A* **2004**, *340*, 147–152, doi:10.1016/j.physa.2004.04.001.
33. Bercher, J.-F. On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized q -Gaussian distributions. *J. Phys. A Math. Gen.* **2012**, *45*, 255303:1–255303:15, doi:10.1088/1751-8113/45/25/255303
34. Bercher, J.-F. On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q -Gaussian distributions. *J. Phys. A Math. Theor.* **2013**, *46*, 095303:1–095303:18, doi:10.1088/1751-8113/46/9/095303.
35. Tanaka, H.-A. Optimal entrainment with smooth, pulse, and square signals in weakly forced nonlinear oscillators. *Phys. D* **2014**, *288*, 1–22, doi:10.1016/j.physd.2014.07.003.
36. Tanaka, H.-A. Synchronization limit of weakly forced nonlinear oscillators. *J. Phys. A Math. Theor.* **2014**, *47*, 402002:1–402002:10, doi:10.1088/1751-8113/47/40/402002.
37. Dembo, A.; Cover, T.M.; Thomas, J.A. Information theoretic inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518, doi:10.1109/18.104312.
38. Oikonomou, T.; Bagci, G.B. A note on the definition of deformed exponential and logarithm functions. *J. Math. Phys.* **2009**, *50*, 103301:1–103301:9, doi:10.1063/1.3227657.
39. Dehesa, J.S.; Galvez, F.J.; Porras, I. Bounds to density-dependent quantities of D -dimensional many-particle systems in position and momentum spaces: Applications to atomic systems. *Phys. Rev. A* **1989**, *40*, 35–40, doi:10.1103/PhysRevA.40.35.
40. Hardy, G.; Littlewood, J.E.; Pólya, G. *Inequalities*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1988; ISBN 978-0-521-35880-4.
41. Rudin, W. *Real and Complex Analysis*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1987; pp. 63–65, ISBN 10 0070542341, ISBN 13 9780070542341.
42. Whittaker, E.T.; Watson, G.N. *A Course of Modern Analysis*, 4th ed.; Cambridge University Press: Cambridge, UK, 1927.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).