Note on bounds for symmetric divergence measures

S.Furuichi\textsuperscript{1,a)}, K.Yanagi\textsuperscript{2} and K.Kuriyama\textsuperscript{3}

\textsuperscript{1}Nihon University  \\
\textsuperscript{2}Josai University  \\
\textsuperscript{3}Yamaguchi University

\textsuperscript{a)}Corresponding author: furuichi@chs.nihon-u.ac.jp

Abstract. In the paper \cite{1}, the tight bounds for symmetric divergence measures are derived by applying the results established in the paper \cite{2}. In this article, we are going to report two kinds of extensions for the above results, namely classical $q$-extension and non-commutative (quantum) extension.

INTRODUCTION

In the paper \cite{1}, the tight bounds for symmetric divergence measures are derived by applying the results established in the paper \cite{2}. In the paper \cite{1}, the minimization problem for Bhattacharyya coefficient, Chernoff information, Jensen-Shannon divergence and Jeffrey’s divergence under the constraint on total variation distance. In this article, we are going to report two kinds of extensions for the above results, namely classical $q$-extension and non-commutative (quantum) extension. The parametric $q$-extension means that Tsallis entropy $H_q(X) \equiv \sum_x p(x)^q - p(x)^{1-q}$ [3] converges to Shannon entropy when $q \to 1$. Namely, all results with the parameter $q$ recover the usual (standard) Shannon’s results when $q \to 1$. We give here list of our extensions as follows.

(i) The lower bound for Jensen-Shannon-Tsallis divergence is given by applying the results in \cite{2}.
(ii) The lower bound for Jeffrey-Tsallis divergence is given by applying the results in \cite{2} and deriving $q$-Pinsker’s inequality for $q \geq 1$. This implies new upper bounds of $\sum_{u \in U} |p(u) - Q_d(u)|$.
(iii) The lower bound for quantum Chernoff information is given by the known relation between the trace distance and fidelity.
(iv) The lower bound for quantum Jeffrey divergence is given by applying the monotonicity (data processing inequality) of quantum $f$-divergence.

$q$-EXTENDED CASES

Here we review some quantities. The total variation distance between two probability distributions $P(x)$ and $Q(x)$ is defined by

$$d_{TV}(P, Q) \equiv \frac{1}{2} \sum_x |P(x) - Q(x)| = \frac{1}{2} \|P - Q\|_1,$$

where $\| \cdot \|_1$ represents $l_1$ norm. The $f$-divergence introduced by Csiszár in \cite{4} is defined by

$$D_f(P\|Q) \equiv \sum_x Q(x)f\left(\frac{P(x)}{Q(x)}\right),$$

where $f$ is convex function and $f(1) = 0$. If we take $f(t) = -t \ln q^{t-1}$, where $\ln_q(x) \equiv \frac{x^{1-q} - 1}{1-q}$ is $q$-logarithmic function defined for $x \geq 0$ and $q \neq 1$, then $f$-divergence is equal to the Tsallis relative entropy (Tsallis divergence) defined by
Applying this lemma, we can prove the following proposition, which condition is same to the paper [1] except for the

\[ D_q(P\|Q) = -\sum_x P(x) \ln \frac{Q(x)}{P(x)} = \sum_x \frac{P(x) - P(x)^q Q(x)^{1-q}}{1-q}. \]

In this section, we use the result established by Gilardoni in [2] for the symmetric divergence.

**Theorem (Gilardoni, 2006 [2])** We suppose \( D_f \) is symmetric divergence (which condition is known as \( f(u) = uf(1/u) + c(u-1), u \in (0,\infty) \) and \( c \) is constant number) and \( f : (0,\infty) \to \mathbb{R} \) with \( f(1) = 0 \). Then we have

\[ \inf_{P,Q:d_f(P,Q)=e} D_f (P \| Q) = (1-e) f \left( \frac{1+e}{1-e} \right) - 2f''(1)e \]

As corollaries of the above theorem, we obtain the following two propositions. We define the Jensen-Shannon-Tsallis divergence as

\[ C_q(P, Q) \equiv D_q \left( P \left\| \frac{P+Q}{2} \right\| \right) + D_q \left( Q \left\| \frac{P+Q}{2} \right\| \right). \]

Then \( D_{f_q}(P\|Q) = C_q(P, Q) \) with \( f_q(t) = -\ln \frac{t^{q+1}}{q} - \ln \frac{t^{q+1}}{q} \), \( f_q \) is convex with \( f_q(1) = 0 \) and \( C_q(P, Q) = C_q(Q, P) \). Thus we have the following proposition which is \( q \)-parametric extension of Proposition 3 in [1].

**Proposition 1**

\[ \min_{P,Q:d_f(P,Q)=e} C_q(P, Q) = -(1-e) \ln_q \frac{1}{1-e} - (1+e) \ln_q \frac{1}{1+e}. \]

The equality is archived when \( P = \left( \frac{1-e}{2}, \frac{1+e}{2} \right), Q = \left( \frac{1+e}{2}, \frac{1-e}{2} \right) \).

We also define Jeffrey-Tsallis divergence as

\[ J_q(P, Q) \equiv \frac{1}{2} \{ D_q (P \| Q) + D_q (Q \| P) \}. \]

Then \( D_{f_q}(P\|Q) = J_q(P, Q) \) with \( f_q(t) = \frac{(q-1)\ln_q t}{q} \), \( f_q \) is convex with \( f_q(1) = 0 \) and \( J_q(P, Q) = J_q(Q, P) \). Thus we have the following proposition which is \( q \)-parametric extension of Proposition 4 in [1].

**Proposition 2**

\[ \min_{P,Q:d_f(P,Q)=e} J_q(P, Q) = -\frac{1}{2} \left( (1+e) \ln_q \frac{1-e}{1+e} + (1-e) \ln_q \frac{1+e}{1-e} \right). \]

The equality is archived when \( P = \left( \frac{1-e}{2}, \frac{1+e}{2} \right), Q = \left( \frac{1+e}{2}, \frac{1-e}{2} \right) \).

Here we are able to prove the following lemma, which may be named \( q \)-Pinsker’s inequality.

**Lemma 1**

\[ D_q(P\|Q) \geq \frac{1}{2} d_{TV}(P, Q)^2 \text{ for } q \geq 1. \]

**Proof:** The proof is easily done by the fact that \( \log t \leq \frac{t-1}{t} \), \((t > 0, r > 0)\) implies \(-\log \frac{1}{y} \leq \ln \frac{1}{t} \), \((t > 0, q > 1)\). Putting \( r = q - 1 \). Thus we have

\[ -\ln_q \frac{y}{x} - (1-x) \ln_q \frac{1-x}{1-y} \geq -x \log \frac{y}{x} - (1-x) \log \frac{1-y}{1-x} \geq 2(x-y)^2 \]

for \( 0 < x, y < 1, q \geq 1 \). Thus we have this lemma by data processing inequality.

As remark, the above \( q \)-Pinsker inequality does not hold for the case \( 0 < q < 1 \), since we have counter-examples. Applying this lemma, we can prove the following proposition, which condition is same to the paper [1] except for the
Theorem 1 Consider a memoryless stationary source with alphabet $\mathcal{U}$ with probability distribution $P$ and assume that a uniquely decodable code with an alphabet size $d$. For $q \geq 1$, we have

$$\frac{1}{2} \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l}(u)| \leq \min \left\{ 1, \sqrt{\frac{\Delta_{d,q} \log_e d}{2}} \right\}.$$ 

Where $\Delta_{d,q} = H_d(q) - H_{d,q}(\mathcal{U}), H_d(q) = -\frac{(c_{d,l})^{-1}}{\log_e d} \sum_{u \in \mathcal{U}} p(u)q \log_q d^{-\log_q d - 1}$, $H_{d,q}(\mathcal{U}) = -\frac{1}{\log_e d} \sum_{u \in \mathcal{U}} p(u)q \log_q p(u)$, and $Q_{d,l}(u)$ is the extended parameter $q$.

Proof: We give the sketch of the proof of this proposition. Firstly $\sum_{u \in \mathcal{U}} |p(u) - Q_{d,l}(u)| \leq 2$ is trivial. By Lemma 1, we have

$$D_q\left( P \mid\mid Q_{d,l} \right) \geq D_q\left( P \mid\mid \hat{\mathcal{Q}_{d,l}} \right) \geq 2(P(A) - Q_{d,l}(A))^2 = 2 \left( \frac{1}{2} I_{TV}(P, \hat{Q}_{d,l}) \right)^2 = \frac{1}{2} \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l}(u)|^2,$$

where $A = \{ x : P(x) > Q_{d,l}(x) \}$, $Y = \phi(X)$ and $P$ and $\hat{\mathcal{Q}_{d,l}}$ are distributions of new random variable $Y$. By simple computations with formula $\ln_q \frac{1}{x} = x^{-1}(\ln_q y - \ln_q x)$, we have

$$D_q\left( P \mid\mid Q_{d,l} \right) = \sum_{u \in \mathcal{U}} p(u)^q \left( \ln_q p(u) - \ln_q Q_{d,l}(u) \right) = \sum_{u \in \mathcal{U}} p(u)^q \left( \ln_q p(u) - \ln_q \frac{d^{-\log_q d - 1}}{c_{d,l}} \right)$$

$$= - \log_e d \cdot H_{d,q}(\mathcal{U}) - \left( c_{d,l} \right)^{-1} \sum_{u \in \mathcal{U}} p(u)^q \left( \ln_q d^{-\log_q d - 1} - \ln_q c_{d,l} \right)$$

$$= - \log_e d \cdot H_{d,q}(\mathcal{U}) + \log_e d \cdot \ln_q 1 \sum_{u \in \mathcal{U}} p(u)^q \leq \log_e d \cdot \Delta_{d,q}$$

since the Kraft-McMillian inequality $c_{d,l} \leq 1$ was used. Thus we have

$$\frac{1}{2} \left( \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l}(u)| \right)^2 \leq \log_e d \cdot \Delta_{d,q}.$$ 

Remark 1 This theorem is a parametric extension of the inequality (32) in the paper [6] in the sense that the left hand side of our inequality contains the parameter $q \geq 1$. We also note that the condition $q \geq 1$ is corresponding to the result in our previous paper [6], so the condition $q \geq 1$ may not be so unnatural within our framework of this topic.

In addition, we compare our upper bound with parameter $q \geq 1$ obtained in Theorem 1 and that obtained in the paper [6]. Actually we give an example such that $\sqrt{\frac{\Delta_{d,1} \log_e d}{2}} \leq \sqrt{\frac{\Delta_{d,1} \log_e d}{2}}$, where $\Delta_{d,1}$ was used in the paper [6] as $\Delta_d$. Consider the following information source

$$\mathcal{U} = \left( \frac{u_1}{u_2}, \frac{u_2}{u_3}, \frac{u_3}{0.5}, \frac{u_3}{0.3}, \frac{u_3}{0.2} \right),$$

with $d = 2$. Then we have the code $u_1 \rightarrow \text{"0"'}$, $u_2 \rightarrow \text{"10"'}$, $u_3 \rightarrow \text{"110"'}$ by Shannon-Fano coding, so that $c_{d,l} = \frac{7}{8} < 1$ since $l_1 = 1, l_2 = 2, l_3 = 3$. By numerical computations, we have $\sqrt{\frac{\Delta_{d,1} \log_e d}{2}} \approx 0.225793$ and $\sqrt{\frac{\Delta_{d,1} \log_e d}{2}} = 0.272669$. This means there exists a code such that $\sqrt{\frac{\Delta_{d,1} \log_e d}{2}} \leq \sqrt{\frac{\Delta_{d,1} \log_e d}{2}}$, which shows our upper bound with the parameter $q \geq 1$ is tighter than the upper bound in the paper [6], in this example. We performed some numerical computations with a few information sources, then we could find the parameter $q \geq 1$ such that $\sqrt{\frac{\Delta_{d,1} \log_e d}{2}} \leq \sqrt{\frac{\Delta_{d,1} \log_e d}{2}}$ for the case $c_{d,l} < 1$. 


However, for the case $c_{d,l} = 1$ (e.g., Huffman code), the following proposition can be proven.

**Proposition 3** Let $q \geq 1$ and $c_{d,l} = 1$. Then we have the relation $\Delta_{d,l} \leq \Delta_{d,q}$.

Proof: We firstly prove the inequality $f_q(x,y) \geq 0$ for $q \geq 1, 0 < x, y \leq 1$, where $f_q(x,y) \equiv x(\log_y y - \log_x x) + x^q(\log_y x - \log_y y)$. Since $\frac{df_q(x,y)}{dy} = \frac{\partial}{\partial y} \left( \frac{\partial f_q(x,y)}{\partial y} \right) = 1$, if $x \leq y$, then $\frac{df_q(x,y)}{dy} \geq 0$ and if $x \geq y$, then $\frac{df_q(x,y)}{dy} \leq 0$, thus we have $f_q(x,y) \geq f_q(x,x) = 0$. Putting $x = p(u)$ and $y = d^{-\log_d u}$, taking summation on both sides by $u \in \mathcal{U}$ and dividing the both sides by $\log_d u$, we have

$$-\frac{1}{\log_d u} \sum_{u \in \mathcal{U}} p(u)^q \log_q d^{-\log_d u} + \frac{1}{\log_d u} \sum_{u \in \mathcal{U}} p(u) \log_q d^{-\log_d u} - \frac{1}{\log_d u} \sum_{u \in \mathcal{U}} p(u) \log_e p(u) + \frac{1}{\log_d u} \sum_{u \in \mathcal{U}} p(u)^q \log_q p(u) \geq 0.$$ 

When $c_{d,l} = 1$, we thus obtain the inequality $\Delta_{d,q} \geq \Delta_{d,1} = \pi_q - \pi_1 + H_{d,1}(\mathcal{U}) - H_{d,q}(\mathcal{U}) \geq 0$, taking account that the usual average code length can be rewritten as $\pi_1 = \sum_{u \in \mathcal{U}} p(u) \log q(u) = -\frac{1}{\log_d u} \sum_{u \in \mathcal{U}} p(u) \log_e d^{-\log_d u}$.

This proposition shows that for the special (but nontrivial) case $c_{d,l} = 1$, the upper bound $\sqrt{\Delta_{d,1} \log_d \frac{d}{2}}$ given in (32) of the paper [1] is always tighter than ours $\sqrt{\Delta_{d,1} \log_d \frac{d}{2}}$ (for $q \geq 1$) obtained in Theorem 1.

**NON-COMMUTATIVE CASES**

Let $\rho$ and $\sigma$ be density matrices (quantum states), which are positive semi-definite matrices and unit trace. Then the following quantities are well known in the field of quantum information or physics as trace distance and fidelity, respectively:

$$d(\rho, \sigma) \equiv \frac{1}{2} \text{Tr} |\rho - \sigma|, \quad F(\rho, \sigma) \equiv \text{Tr} \left[ \rho^{1/2} \sigma^{1/2} \right],$$

Where $|A| = (A^*A)^{1/2}$. Then we have the following propositions.

**Proposition 4** For the trace distance and fidelity, we have the following relation:

$$1 - d(\rho, \sigma) \leq F(\rho, \sigma) \leq \sqrt{1 - d(\rho, \sigma)^2}.$$ 

This relation is well known in the field of quantum information or quantum statistical physics, and this proposition is non-commutative extension of Proposition 1 in the paper [1].

By the easy calculations such as $C_0(\rho, \sigma) \equiv -\log \left( \min_{0 \leq s \leq 1} \text{Tr} \left[ \rho^s \sigma^{1-s} \right] \right) = -\min_{0 \leq s \leq 1} \left( \log \text{Tr} \left[ \rho^s \sigma^{1-s} \right] \right) \geq -\log \text{Tr} \left[ \rho^{1/2} \sigma^{1/2} \right] \geq -\log \text{Tr} \left[ \rho^{1/2} \sigma^{1/2} \right] = -\log F(\rho, \sigma) \geq -\frac{1}{2} \log \left( 1 - d(\rho, \sigma)^2 \right)$, we have the following proposition.

**Proposition 5** For the quantum Chernoff information, we have

$$\min_{\rho, \sigma, d(\rho, \sigma)=\epsilon} C_0(\rho, \sigma) = \begin{cases} -\frac{1}{2} \log \left( 1 - \epsilon^2 \right), \epsilon \in [0, 1) \\ +\infty, \quad \epsilon = 1 \end{cases}$$

The above proposition is also non-commutative extension of Proposition 2 in the paper [1].

The quantum Pinsker inequality on quantum relative entropy (divergence) and similar one are known (see e.g., [7] and [8], respectively)

$$D(\rho|\sigma) \equiv \text{Tr} [\rho (\log \rho - \log \sigma)] \geq \frac{1}{2} \text{Tr} [\rho - \sigma]^2.$$
and
\[ D(\rho|\sigma) \geq -2 \log \text{Tr} \left[ \rho^{1/2} \sigma^{1/2} \right] \geq \text{Tr} \left[ \rho^{1/2} - \sigma^{1/2} \right]^2 \]

To show our final result, we use the following well-known fact. See [7] for example.

**Lemma 2** Let \( E : B(\mathcal{H}) \to B(\mathcal{K}) \) be a state transformation. For an operator monotone decreasing function \( f : \mathbb{R}^+ \to \mathbb{R} \), the monotonicity holds:
\[
D_f (\rho|\sigma) \geq D_f (E(\rho)|E(\sigma))
\]
where \( D_f (\rho|\sigma) \equiv \text{Tr} \left[ \rho f (A) \right] \) is the quantum \( f \)-divergence, with \( \Delta_{\sigma,\rho} \equiv \Delta = LR \) is the relative modular operator such as \( L(A) = \sigma A \) and \( R(A) = A \rho^{-1} \).

**Theorem 2** The quantum Jeffrey divergence defined by \( J(\rho|\sigma) \equiv \frac{1}{2} \left[ D(\rho|\sigma) + D(\sigma|\rho) \right] \) has the following lower bound:
\[
J(\rho|\sigma) \geq d(\rho,\sigma) \log \left( \frac{1 + d(\rho,\sigma)}{1 - d(\rho,\sigma)} \right).
\]

**Proof:** By Lemma 2, Proposition 4 in the paper [1] and \( \|\rho - \sigma\|_1 = \|P - Q\|_1 \) (which will be shown in the end of proof), we have
\[
J(\rho|\sigma) \geq J(P|Q) \geq d_{TV}(P,Q) \log \left( \frac{1 + d_{TV}(P,Q)}{1 - d_{TV}(P,Q)} \right) = d(\rho,\sigma) \log \left( \frac{1 + d(\rho,\sigma)}{1 - d(\rho,\sigma)} \right).
\]
Here we note that \( f(t) = \frac{1}{t}(t - 1) \log t \) is operator convex which is equivalent to operator monotone decreasing and we have \( D_{f(t-1) \log}(\rho|\sigma) = J(\rho|\sigma) \), since \( (\Delta_{\sigma,\rho})^n (Y) = Y \log Y \rho^{-1} - \sigma Y \rho^{-1} \log \rho \).

Finally, we show \( \|\rho - \sigma\|_1 = \|P - Q\|_1 \). Let \( \mathcal{A} = C^*(\rho_1 - \rho_2) \) be commutative \( C^* \)-algebra generated by \( \rho_1 - \rho_2 \), \( M_n \) be the set of all \( n \times n \) matrices and set the map \( E : M_n \to \mathcal{A} \) as trace preserving, conditional expectation. If we take \( \rho_1 = E(\rho_1) \) and \( \rho_2 = E(\rho_2) \), then two elements \( (\rho_1 - \rho_2)_+ \) and \( (\rho_1 - \rho_2)_- \) of Jordan decomposition of \( \rho_1 - \rho_2 \), are commutative functional calculus of \( \rho_1 - \rho_2 \), and we have \( p_1 - p_2 = E(\rho_1 - \rho_2) = E((\rho_1 - \rho_2)_+ - (\rho_1 - \rho_2)_-) = E(\rho_1 - \rho_2)_+ - E(\rho_1 - \rho_2)_- = \rho_1 - \rho_2 \) which implies \( \|\rho - \sigma\|_1 = \|P - Q\|_1 \).

**ACKNOWLEDGMENTS**

The author (S. F.) was partially supported by JSPS KAKENHI Grant Number 16K05257.

**REFERENCES**

[1] I. Sason, Tight Bounds for Symmetric Divergence Measures and a Refined Bound for Lossless Source Coding, IEEE, TIT, Vol. 61(2015),pp.701–707.
[2] G. L. Gilardoni, On the minimum \( f \)-divergence for given total variation, C. R. Acad. Sci. Paris, Ser. I, Vol.343 (2006), pp.763–766.
[3] C.Tsallis, Possible generalization of Bolzmann-Gibbs statistics, J.Stat. Phys., Vol.52(1988), pp. 479–487.
[4] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, Stud. Sci. Math. Hungarica, Vol. 2(1967), pp. 299–318.
[5] S.Furuichi, K.Yanagi and K.Kuriyama, Fundamental properties of Tsallis relative entropy, J.Math.Phys., Vol.45(2004), pp.4868–4877.
[6] S.Furuichi, Information theoretical properties of Tsallis entropies, J.Math.Phys., Vol.47(2006), 023302.
[7] D.Petz, Quantum information theory and quantum statistics, Springer, 2004.
[8] E.A.Carlen and E.H. Lieb, Remainder terms for some quantum entropy inequalities, J. Math. Phys., Vol.55 (2014), 042201.
Appendix: Added notes related to Theorem 1

Actually we have \( \lim_{q \to 1} \pi_q = \sum_{u \in \mathcal{U}} p(u) l(u) \) which is the usual average code length, but the definition of \( \pi_q \) in Theorem 1 seems to be complicated and somewhat unnatural to understand its meaning. In order to overcome this problem, we may adopt the simple alternative definition for \( \pi_q \) instead of that in Theorem 1. Then we have the following proposition.

**Proposition A** Let \( q \geq 1 \) and \( c_{d,l,q} \leq 1 \). Then we have

\[
\frac{1}{2} \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l,q}(u)| \leq \min \left\{ 1, \sqrt{\frac{\Delta_{d,q} \log d}{2}} \right\}
\]

Where \( \Delta_{d,q} = \pi_q - H_{d,q}(\mathcal{U}) \pi_q = \sum_{u \in \mathcal{U}} p(u)^q l(u) \), \( H_{d,q}(\mathcal{U}) = -\frac{1}{\log d} \sum_{u \in \mathcal{U}} p(u)^q \log p(u) \), \( Q_{d,l,q}(u) = \frac{1}{c_{d,l,q}} \exp_q \left( \log d^{-l(u)} \right) \) and \( c_{d,l,q} = \sum_{u \in \mathcal{U}} \exp_q \left( \log d^{-l(u)} \right) \), where \( q \)-exponential function \( \exp_q(\cdot) \) is the inverse function of \( q \)-logarithmic function \( \log_q(\cdot) \) and its form is given in the proof of this proposition.

**Proof:** By the same way to the proof of Theorem 1, we have

\[
D_q \left( P \mid Q_{d,l,q} \right) \geq \frac{1}{2} \left( \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l,q}(u)| \right)^2,
\]

By simple computations with formula \( \log_q = y^{1-q}(\log y + \log_q 1) \), we have

\[
D_q \left( P \mid Q_{d,l,q} \right) = \sum_{u \in \mathcal{U}} p(u)^q \left( \log_q p(u) - \log_q Q_{d,l,q}(u) \right) = -\log_q d \cdot H_{d,q}(\mathcal{U}) - \sum_{u \in \mathcal{U}} p(u)^q \log_q \frac{\exp_q \left( \log d^{-l(u)} \right)}{c_{d,l,q}}
\]

\[
- \log_q d \cdot H_{d,q}(\mathcal{U}) - \sum_{u \in \mathcal{U}} p(u)^q \log_q d^{-l(u)} - \sum_{u \in \mathcal{U}} p(u)^q \left( \exp_q \left( \log d^{-l(u)} \right) \right)^{1-q} \log_q \frac{1}{c_{d,l,q}}
\]

\[
\leq \Delta_{d,q} \log_q d
\]

since \( d \geq 2, l(u) \geq 1 \) implies \( \log d^{-l(u)} \leq 0 \) thus we have \( 1 + (1 - q) \log d^{-l(u)} \geq 0 \), then the definition of \( q \)-exponential function

\[
\exp_q(x) = \begin{cases} (1 + (1 - q) x)^{1/q}, & \text{if } 1 + (1 - q) x > 0 \\ 0, & \text{otherwise} \end{cases}
\]

shows \( \exp_q(\log d^{-l(u)}) \geq 0 \) and \( c_{d,l,q} \leq 1 \) was used. Thus we have

\[
\frac{1}{2} \left( \sum_{u \in \mathcal{U}} |p(u) - Q_{d,l,q}(u)| \right)^2 \leq \Delta_{d,q} \log_q d.
\]

We could not remove the needless and meaningless condition \( c_{d,l,q} \leq 1 \) in the above proposition, unfortunately. It is known that the inequality \( c_{d,l,1} \leq 1 \) holds for the uniquely decodable code and the equality \( c_{d,l,1} = 1 \) holds if the code archives the entropy, namely \( \pi_1 = H_{d,1}(\mathcal{U}) \). In our proposition, we obtained \( q \)-parametric extension but it does not have any information theoretical meaning. We will have to consider about this problem in the future.