Entropy **2001** , *3*, 280–292

entropy
ISSN 1099-4300
www.mdpi.org/entropy/

A Possible Extension of Shannon's Information Theory

Takuya Yamano

Institut für Theoretische Physik, Universität zu Köln, Zülpicher Str.77, D-50937 Köln, Euroland
Department of Applied Physics, Faculty of Science, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo,152-8551, Japan
E-mail: tyamano@mikan.ap.titech.ac.jp

Received: 7 August 2001/ Accepted: 31 October 2001/ Published: 21 November 2001

Abstract:
As a possible generalization of Shannon's information theory, we review the formalism based on the non-logarithmic information content parametrized by a real number q, which exhibits nonadditivity of the associated uncertainty. Moreover it is shown that the establishment of the concept of the mutual information is of importance upon the generalization.
Keywords: Information theory; Tsallis entropy; Nonadditivity; Source coding theorem; Mutual information.

---

1 Introduction

The recent development of the statistical mechanics based on the non-Gibbsian entropy, or Tsallis's nonextensive statistical mechanics [1] has intensified interests in search for a possible extension of Shannon information theory. This is mainly due to the similarity of the form of the entropy function involved. In fact one can develop the non-Shannon (but includes Shannon's as a special case) information theory to some extent from a formal point of view. The quantitative measure of information is of importance when we transmit the strings of messages through a channel, moreover, coding and decoding the messages can be considered to be manipulation of the amount of information. Therefore we need a definition of the information content. Shannon wrote in his original paper[2]:
If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and ...
Here it is worth to know the motivation of adopting the information measure introduced by Hartley[3]. He proposed that the information is to be proportional to the the number of possible events (the term 'selections' is used in his original paper). By comparing the number of possible sequences of two systems, he states that to take the logarithm of the number is most practical. Therefore, Shannon provided the measure of uncertainty of information contained as an entropy which is the celebrated form $-\sum_i p_i \log p_i$ [2, 4], where the p_i is the probability of the i-th state of the system considered.

We shall note that the above statement by Shannon may seem to allow some possibilities of generalized formalism for theory of communication, or information theory. Up to now, many forms of information function differed from Shannon have been devised with parameters[5, 6, 7]. Among them, let us mention that similar forms to Tsallis's are that of Havrda et.al. and of Daróczy[6]. However the discussions about the coding theorem and other important elements for information theory based on such a generalized function seem to lacking. In this paper we assume that nonadditivity of the uncertainty of the system arises from a certain property of messages. We briefly shows the present trial of a possible extension of Shannon's information theory.

2 Set up

Before we go into details, we first define the information source. This is common in both an additive source and a nonadditive one and described as follows. A source consists of a finite set called a source alphabet $S = \{x_1, \dots x_n\}$ and of the probability set $P = \{p(x_1), \dots, p(x_n)\}$

assigned to them. The entropy function H depends only on the P and not on the elements of the alphabets.
Here, it would be worth mentioning the following fact. I.e. the entropy function is uniquely determined as Shannon's form due not to the definition of the mean value of the information but to the additivity of the uncertainty of the information that the source has. To see this clearly, let us take the following explanation. Let the elements $S = \{x_1, \dots x_n\}$ partition into k nonempty disjoint groups $G_1, \dots, G_k$, each of which has the size $|G_i| = g_i$ and $\sum_i g_i = n$. We first chose a group G_i with probability $p(G_i) = g_i/n$, then we pick an alphabet with equal probability from the group G_i. Since the conditional probability that a certain alphabet x_j is in G_m is given as

$$p(x_j \mid G_i) = \left\{ \begin{array}{cc} \frac{1}{g_m} & (i = m) \\ 0 & (i \neq m) \end{array} \right. , \tag{1}$$

the probability of choosing x_j is written with Bayes' relation as

$$p(x_j) = \sum_{i=1}^{n} p(x_j \mid G_i) p(G_i) = \frac{1}{g_m} \frac{g_m}{n} = \frac{1}{n}. \tag{2}$$

Therefore we see that it is the same as the probability when we chose x_j directly from S with equal probability. Although a group that an element belongs has been identified, we still have the uncertainty involved in choosing the element from the group. The uncertainty in choosing an element from G_i is $H(1/g_i, \dots, 1/g_i)$. Here, suppose that the average uncertainty be given by taking e.g. an escort probability [8] with a parameter $q \in \mathcal{R}$ instead of the usual average

$$\sum_{i=1}^{k} p^q(G_i) H(\frac{1}{g_i}, \dots, \frac{1}{g_i}) / \sum_{i=1}^{k} p^q(G_i) = \sum_{i=1}^{k} \left(\frac{g_i}{n}\right)^q H(\frac{1}{g_i}, \dots, \frac{1}{g_i}) / \sum_{i=1}^{k} \left(\frac{g_i}{n}\right)^q . \tag{3}$$

The case $q = 1$ means to take the usual average. On the other hand, we have the uncertainty in choosing one group G_i from k groups, that is $H(g_1/n, \dots, g_k/n)$. When we impose the additivity of the uncertainty, the uncertainty in choosing directly from the source S with equal probability $H(1/n, \dots, 1/n)$ can be provided as the sum of the two uncertainties,

$$H(\frac{1}{n}, \dots, \frac{1}{n}) = H(\frac{g_1}{n}, \dots, \frac{g_k}{n}) + \sum_{i=1}^{k} \left(\frac{g_i}{n}\right)^q H(\frac{1}{g_i}, \dots, \frac{1}{g_i}) / \sum_{i=1}^{k} \left(\frac{g_i}{n}\right)^q . \tag{4}$$

When $g_i = l$ for all $i = 1, \dots, k$ therefore $lk = n$, no matter what the value q takes, r.h.s of the above relation reduces to $H(l/n, \dots, l/n) + H(1/l, \dots, 1/l)$. The proof of the uniqueness of the logarithmic form for H after this stage (with two more important properties:H is a continuous function of its argument $\{p_i\}(p_i \in [0, 1])$ for all i and we have greater uncertainty when there are more outcomes, $H(\frac{1}{n}, \dots, \frac{1}{n}) < H(\frac{1}{n+1}, \dots, \frac{1}{n+1}))$ can be seen in the standard textbook e.g. [9] . The important point in the above discussion is that the logarithmic form is not stem from the way we take the average of the uncertainty but from the additivity of uncertainties. The additivity postulate of the uncertainty leads to logarithmic property of the information acquired. It is worthwhile to note that the uniqueness theorem for Tsallis entropy has been discussed in Ref.[10, 11]. These are not based on the additivity of the entropy function but on the pseudoadditivity[12]. We are concerned with the formulation when we discard the additivity postulate in the present consideration. As a starting point, we begin with the definition of the amount of information associated with the occurrence of a certain information from the source. Then we shall define the pertinent information entropy based on it. Upon the generalization of the information content parametrized with nonadditivity, there seem two unavoidable notions that we should bear in mind. First point is that we should retain Shannon's (or Hartley's) information as a special case. Second, the introduction of appropriate definition for the mutual information is needed to develop the channel coding theorem. We shall describe this point later in Section 3. Hereafter we change notations that denote elements of an alphabet to avoid burdensome suffixes: x belongs to an alphabet $\mathcal{H}$ and y to another $\mathcal{Y}$.

We borrow from the definition used in Tsallis nonextensive statistical mechanics. In Tsallis's formalism of statistical mechanics, we base our discussions on the q-logarithmic function [13, 14, 15]. Then we may regard the amount of information as the q-logarithm of the probability $I_q(p) \equiv -\ln_q p(x)$ where $\ln_q x = (x^{1-q} - 1)/(1-q)$. In fact, $I_q(p)$ is a monotonically decreasing function as $-\ln p$ is for all q. In this case, we measure the information not by *bit* but by *nat*. In the limit $q \to 1$, the information content recovers $-\ln p$.

One should also notice that the escort average or the normalized $q-$average[12] of the information content (information entropy) gives the modified form of Tsallis entropy[16, 17, 18]

$$\frac{-\sum_{x\in\mathcal{H}} p^q(x)\ln_q p(x)}{\sum_{x\in\mathcal{H}} p^q(x)} = \frac{1-\sum_{x\in\mathcal{H}} p^q(x)}{(q-1)\sum_{x\in\mathcal{H}} p^q(x)}. \tag{5}$$

We note that this form takes as the Tsallis one divided by a factor $\sum_i p_i^q$. We shall hereafter denote the entropy function by indexing q as $H_q(X)$.

In a similar way, we define the joint entropy of X and Y and a nonadditive conditional entropy $H_q(Y \mid x)$ as

$$H_q(X,Y) := \frac{1-\sum_{x\in\mathcal{H},y\in\mathcal{Y}} p^q(x,y)}{(q-1)\sum_{x\in\mathcal{H},y\in\mathcal{Y}} p^q(x,y)}, \quad \langle\frac{1}{1+(q-1)H_q(Y \mid x)}\rangle_q^{(X)} := [1+(q-1)H_q(Y\mid X)]^{-1} \tag{6}$$

where $\langle\cdot\rangle_q^{(X)}$ represents the escort average with respect to $p(x)$. Let us next define the mutual information $I_q(Y;X)$. Here, we shall follow the usual definition of the reduction in uncertainty due to another variable (the entropy subtracted by the conditional entropy). That is,

$$I_q(Y;X) \equiv H_q(Y) - H_q(Y \mid X) = \frac{H_q(X) + H_q(Y) - H_q(X,Y) + (q-1)H_q(X)H_q(Y)}{1+(q-1)H_q(X)}. \tag{7}$$

In order to be in consistent with the usual additive mutual information, we impose non-negativity for it. It also converges to the usual mutual information $I(Y;X) = H(Y) - H(Y \mid X) = H(X) + H(Y) - H(X,Y)$ in the additive limit($q \to 1$). The mutual information of a random variable with itself is the entropy itself $I_q(X;X) = H_q(X)$, and when X and Y are independent variables, $I_q(Y;X) = 0$. Some people claim that the mutual information in any generic extension (a) ought to obey symmetry under interchange of two variables(sender X and receiver Y) and (b) must vanish when the two events are statistically independent. However, there seems no clear reason to keep these properties when we measure the information as the non-logarithmic form. We could also say that since the pseudoadditivity states that the probability with factorized form does not construct the sum of the pertinent entropy, it may take a stand point that the mutual information may not possess the above two properties at the same time. We will discuss this point in the next Section.
We here only give the relations which the H_q satisfies (the reader who has interest in proofs is referred to [18]).

- $H_q(X,Y) = H_q(X) + H_q(Y \mid X) + (q-1)H_q(X)H_q(Y \mid X)$
- $H_q(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} \left[1 + (q-1)H_q(X_{i-1}, \cdots, X_1)\right] H_q(X_i \mid X_{i-1}, \cdots, X_1)$
- $H_q(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} \left[1 + (q-1)H_q(X_{i-1}, \ldots, X_1)\right] H_q(X_i)$
- $[1 + (q-1)H_q(X)]H_q(Y \mid X) = H_q(Y,Z \mid X) - H_q(Z \mid Y,X) + (q-1)\{H_q(X)H_q(Y,Z \mid X) - H_q(X,Y)H_q(Z \mid Y,X)\}$

where Z in the last relation denotes the third events. All these relations reduce to the ones which are satisfied in the usual Shannon's information theory in the limit $q \to 1$. Next, we give the generalized Kullbak-Leibler(KL) entropy between two probability distributions $p_i(x)$ and $p'_i(x)$ as follows [18, 19, 20],

$$D_q[p(x) \parallel p'(x)] := \langle \ln_q p'(x) - \ln_q p(x) \rangle_q^{(X)}. \tag{8}$$

The KL entropy or the relative entropy is a measure of the distance between two probability distributions[21]. We note that the above generalized KL entropy satisfies the form invariant structure which has introduced in Ref.[17]. Then the following information inequality holds

$$D_q[p(x) \parallel p'(x)] \begin{cases} \geq 0 & (q > 0) \\ < 0 & (q < 0) \end{cases} , \tag{9}$$

with equality if and only if $p(x) = p'(x)$ for all $x \in \mathcal{H}$ and $D_0[p(x) \parallel p'(x)] = 0$. The positivity in the case of $q > 0$ can be considered a necessary property to develop the information theory.

3 Source coding theorem

We review and give the nonadditive version of the Shannon's source coding theorem. Source coding theorem together with the channel coding theorem can be considered a pillar of information theory today. We usually want to encode the source letters in such a way that the expected codeword length becomes as short as possible. We recapitulate how the source coding theorem is described by writing a usual Shannon's case in a parallel way. We write expressions of the nonadditive version in parentheses in this section.

Let us denote codeword lengths of each letters as $l_1, l_2, \cdots$. We recall that any code which is uniquely decoded and satisfies prefix conditions over the alphabet size D must be content with the Kraft inequality[22]

$$\sum_i^M D^{-l_i} \leq 1 \tag{10}$$

where M is the number of codewords. The minimum expected codeword length $\langle L \rangle = \sum_i p_i l_i$ ($\langle L \rangle_q = \sum_i p_i^q l_i / \sum_i p_i^q$) can be achieved by using the Lagrange multiplier method with constraint of the Kraft inequality. The resulting optimal codeword length l_i^* becomes $-\log_D p_i$ ($\log_D(\sum_i p_i^q) - q \log_D p_i$). The following property is crucial i.e., the codeword length can not be below the entropy,

$$H(X) \leq \langle L \rangle, \quad (H_q(X) \leq \langle L \rangle_q) \tag{11}$$

with equality if and only if $p_i = D^{-l_i}$ ($p_i = [1-(1-q)l_i]^{\frac{1}{1-q}}$). When we use an incorrect or a non-optimal information, we have more complexity in its description. In other words, the use of a wrong distribution, say r, involved in each source letter incurs a penalty (or an increment of the expected description length) of the relative entropy $D[p \parallel r]$ ($D_q[p \parallel r]$) in the average codeword length,

$$\langle L \rangle = H(X) + D[p \parallel r], \quad (\langle L \rangle_q = H_q(X) + D_q[p \parallel r]). \tag{12}$$

In fact, the l_i^*'s may not be integers then we round them up to the smallest integers larger than l_i^*'s. We represent these codeword lengths as $l_i = \lceil \log_D \frac{1}{p_i} \rceil$ ($\lceil \log_D (\sum_i p_i^q) - q \log_D p_i \rceil$). We can show that the average codeword length resides within 1 *nat* against $H(p) + D[p \parallel r]$ $(H_q(p) + D_q[p \parallel r])$ due to the present definition of the l_i, that is, $\log_D \frac{1}{p_i} \leq l_i < \log_D \frac{1}{p_i} + 1$ $(\log_D(\sum_i p_i^q) - q \log_D p_i \leq l_i < \log_D(\sum_i p_i^q) - q \log_D p_i + 1)$. When we transmit a sequence of n letters from the source, each of which is generated independently and in identically distributed manner, the average codeword length per one letter $\langle L_n \rangle$ is bounded as $H(X) \leq \langle L_n \rangle < H(X) + 1/n$. This bound relation can be attributed to the additive property of the joint entropy $H(X_1, \ldots, X_n) = \sum_i H(X_i) = nH(X)$. In the nonadditive case, on the other hand, $H(X)$ is replaced by $\frac{\sum_{i=1}^{n}[1+(q-1)H_q(X_{i-1},\ldots,X_1)]H_q(X_i)}{n}$. This form arises from the property of $H_q(X_1, \cdots, X_n)$ we have listed in the previous section. This can be considered to complete the source coding theorem. We only mention the fact that the Fano's inequality which provides an upper bound to the conditional entropy, when the system has a noisy channel, can be naturally generalized within the present formalism[18].

4 Mutual information and channel capacity

We address the generalized mutual information we have promised in the Section 2 and explore consequences of two alternative definitions of mutual information by calculating the simple model of information channel i.e., the binary symmetric channel(BSC). Recently, Landsberg et.al.[16] have focused on a possible constraint on the generalization of entropy concept and have compared channel capacities based on Shannon, Renyi, Tsallis, and another generalized entropy by applying to a BSC for a fixed error probability e of communication. It would be worth noting that the last generalized entropy introduced by Landsberg et al. was presented independently by Rajagopal et al.[17] from a consideration of a form invariant structure of nonadditive entropy.

The concept of the mutual information can be regarded as the important ingredient to evaluate a communication channel in information theory. The mutual information is a measure of the amount of information that one random variable contains about another random variable. In other wards, $I_q(Y; X)$ expresses the reduction in uncertainty of Y due to the acquisition of knowledge of X[22]. In previous considerations in Section 2 and in Ref.[16], there is a point that should be investigated about the definition of the mutual information upon generalization. In Ref.[16] they devised to take the average of conditional entropy $H(Y \mid X = x)$ using usual weighting $p(x)$ from the requisition that the capacity based on the any information entropy should be 0 for $e = 1/2$ as Shannon's case. However this selection possesses asymmetry of the mutual information in X and Y unless we choose the Shannon's $H(Y \mid X = x)$. The same applies to the case of the definition we have adopted.

Let us try to compare the two possible definitions of the generalized mutual information by checking the channel capacity of the BSC. The one holds symmetry in X and Y and the other is borrowed from the generalized KL entropy.

To capture and elucidate the present problem clearly let us review the usual mutual information by Shannon. The mutual information can be defined as the relative entropy between the joint probability distribution $p(x, y)$ and the product of the two probability distributions $p(x)p(y)$. That is, we can regard that the mutual information is a marker how far the joint probability deviates from the independence of the two i.e.,

$$I(X;Y) = D[p(x,y) \parallel p(x)p(y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{13}$$

Hence the mutual information is symmetric in X and Y, $I(X;Y) = I(Y;X)$. Moreover its positivity can be proved. Here, the most remarkable feature associated with the later discussion is the fact that we can rewrite the $I(X;Y)$ from Eq.(13) as

$$\begin{aligned} I(X;Y) &= -\sum_x p(x) \log p(x) + \left(\sum_{x,y} p(x,y) \log p(x|y) \right) \\ &= H(X) - H(X|Y) \end{aligned} \tag{14}$$

Therefore we can take the definition of the mutual information either as $H(X) - H(X \mid Y)$ or as $D[p(x,y) \parallel p(x)p(y)]$. However, the situation totally differs in the context of the generalized mutual information. More specifically, the definition by the generalized KL entropy $D_q[p(x,y) \parallel p(x)p(y)]$ [18] does not give the expression $H_q(X) - H_q(X \mid Y)$. Accordingly we need to determine which standpoint we take as the definition.
Now we first consider the mutual information indexed by q as

$$I_q(X;Y) \stackrel{\text{def}}{=} H_q(X) - [1 + (q-1)H_q(Y)]H_q(X \mid Y). \tag{15}$$

Then we immediately see that the mutual information becomes symmetric in X and Y,

$$I_q(X;Y) = I_q(Y;X) = H_q(X) + H_q(Y) - H_q(X,Y). \tag{16}$$

In BSC we have the alphabet $\mathcal{H} = \{0, 1\}$. The input code x of 0 and 1 can be received as a 1 and a 0 of the output code y respectively with the error probability e due to the external noise[22]. Then the conditional probabilities $p(y \mid x)$ are given as $p(0 \mid 1) = p(1 \mid 0) = e$ and $p(0 \mid 0) = p(1 \mid 1) = 1 - e$ respectively. If we write the input (output) probability which we observe a 0 and a 1 as p_0 and p_1 (p'_0 and p'_1) respectively, then we have the joint probabilities $p_{xy} = p(y \mid x)p_x$; $p_{00} = (1-e)p_0$, $p_{01} = ep_0$, $p_{10} = ep_1$, and $p_{11} = (1-e)p_1$.
Using this, the nonadditive joint entropy is calculated as $(1 - 1/(p_0^q + p_1^q))(e^q + (1-e)^q))/(q-1)$. We define a generalized channel capacity as follows

$$C_q \stackrel{\text{def}}{=} \max_{p(x)} I_q(X;Y). \tag{17}$$

Since the output probabilities p'_0 and p'_1 are given as

$$p'_0 = p_{00} + p_{10} = p_0 + (p_1 - p_0)e, \quad p'_1 = p_{01} + p_{11} = p_1 + (p_0 - p_1)e \tag{18}$$

respectively, the summation $\sum_i p^q(y_i)$ is calculated with $p_1 = 1 - p_0$ as $((1-2e)p_0 + e)^q + ((2e - 1)p_0 + 1 - e)^q$. Therefore from Eq(16) the mutual information is expressed as

$$I_q(X;Y) = \frac{1}{q-1}\{\frac{1}{(p_0^q + (1-p_0)^q)}\left(1 - \frac{1}{e^q + (1-e)^q}\right) + \frac{1}{f(p_0, e)} - 1\} \tag{19}$$

where we have put $f(p_0, e) = ((1-2e)p_0 + e)^q + ((2e-1)p_0 + 1 - e)^q$. The $I_q(X;Y)$ can be taken its maximum when $p_0 = p_1 = 1/2$ for some ranges of q and e. Then the capacity is found to be

$$C_q = \frac{1}{q-1}\left(2^q - 1 - \frac{2^{q-1}}{e^q + (1-e)^q}\right). \tag{20}$$

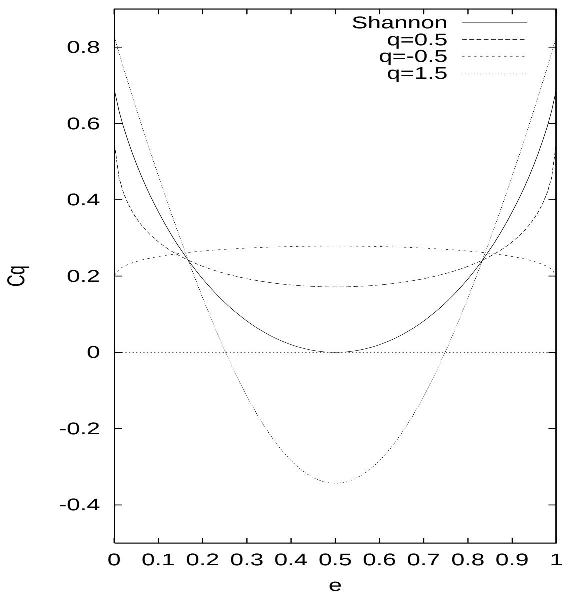

Figure 1: The channel capacity of BSC is plotted against the error probability for Shannon and some different values of q defined by Eq.(16).

Fig.1 shows the channel capacity C_q against e for the case of Shannon and for different values of q. The capacity does not become zero for $e = 1/2$ except for Shannon. It exceeds the Shannon's capacity for the intermediate noise level ($q < 1$) and is below Shannon when the channel is very noisy(or less noisy). We note that the capacity can be negative also for our new definition of the mutual information for some ranges of q and e and we can have the capacity above Shannon for extreme cases of e. However, the negative channel capacity has a difficulty to understand in the usual communication framework as pointed out in the definition of Ref.[16].

Next we take a stance where the generalized mutual information should be defined in terms of the generalized KL entropy[18] in a complete analogous manner to the Shannon's case i.e.,

$$\begin{aligned} I_q(X;Y) &\stackrel{\text{def}}{=} D_q[p(x,y) \parallel p(x)p(y)] \\ &= \frac{1}{1-q}\left[1-\sum_{x,y} p(x,y)\left(\frac{p(x)p(y)}{p(x,y)}\right)^{1-q}\right] \Big/ \sum_{x,y} p^q(x,y). \end{aligned} \tag{21}$$

Then Eq.(21) in BSC is calculated as

$$I_q(X;Y) = \frac{1}{1-q}\frac{1-h(p_0,e)-g(p_0,e)}{(p_0^q+(1-p_0)^q)(e^q+(1-e)^q)} \tag{22}$$

where we have put $h(p_0,e) = [(1-e)^q p_0 + e^q(1-p_0)][(1-2e)p_0+e]^{1-q}$ and $g(p_0,e) = [(1-e)^q(1-p_0)+e^q p_0][(2e-1)p_0+1-e]^{1-q}$. Since the maximization of $I_q(X;Y)$ can be achieved when $p_0 = p_1 = 1/2$ as in the previous definition, we can calculate C_q as

$$C_q = \frac{1}{q-1}\left(2^{2q-2} - \frac{2^{q-1}}{e^q+(1-q)^q}\right). \tag{23}$$

Fig.2 shows the channel capacity by Eq.(23) for some values of q and Shannon's capacity as a reference. The zero capacity is achieved for all cases when $e = 1/2$. This capacity contrasts with the one by the first definition Eq(15); the negative capacity emerges for $q < 0$ and it goes beyond(underrun) the Shannon's capacity when $q > 1$ ($q < 1$).

In conclusion, we have summarized the present developments of a possibility of extension of Shannon's information theory with a nonadditive information content. As far as the source coding theorem is concerned, we can successfully extend the theorem with nonadditive information content of the form $-\ln_q p(x)$ with *escort* average. Moreover we have devised new possible definitions of the mutual information in nonadditive context. One of which is due to the entropy expression Eq.(15) that preserves the symmetry in the input and the output. The other has constructed by the generalized KL entropy. As an application of this, the behavior of the channel capacity of BSC has shown for different noise characteristics. The generalized channel capacity can entail negativity for $q > 1$ in the former case and for $q < 0$ in the latter case. In the nonadditive context, we may allow the mutual information to be asymmetric, however, if we decide to take the standpoint that any generalization of mutual information has to retain symmetry, we would have to accept the modification of the intuitively reasonable property of Shannon's capacity. On the other hand, the mutual information by the generalized KL entropy may suggest plausible definition in that the zero capacity is achieved for all q when $e = 1/2$. However the consistent definition of the generalized mutual information that both approaches give the same results would be welcomed. The necessity for investigation on whether or not the capacity based on the statistics which breaks the additivity can be valid for the real communication channel remains unchanged.

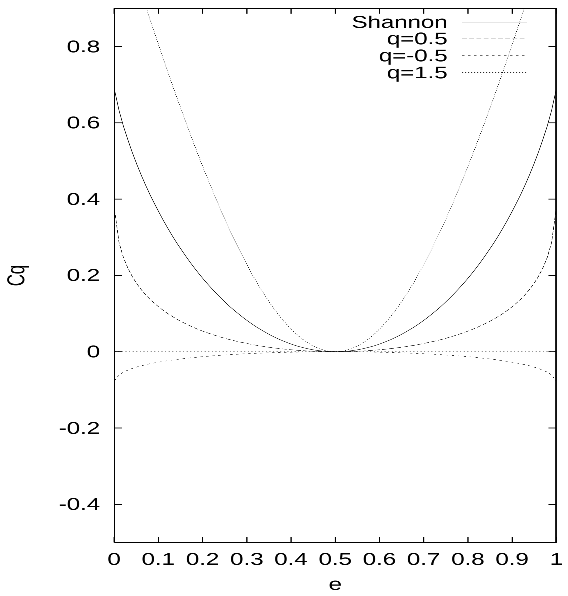

Figure 2: The channel capacity of BSC is plotted against the error probability for Shannon and some different values of q defined by Eq.(21)

Acknowledgements

The author acknowledges the financial support from the DAAD(Deutscher Akademischer Austauschdienst) for staying the Universität zu Köln and would like to thank Prof.D.Stauffer for his kind hospitality.

References

[1] C. Tsallis, Possible generalization of Boltzmann- Gibbs statistics, J. Stat. Phys. **52**, 479 (1988); E.M.F. Curado and C. Tsallis, Generalized statistical mechanics: connection with thermodynamics, J.Phys.A **24**, L69 (1991); Corrigenda: **24**, 3187 (1991) and **25**, 1019 (1992).

[2] C.E.Shannon, A Mathematical Theory of Communication, Bell Syst. Tech. J. **27**, 379 (1948); *ibid* 623 (1948).

[3] R.V.L Hartley, Transmission of information, Bell Syst. Tech. J. **7**, 535 (1928).

[4] C.E. Shannon and W.Weaver, *The Mathematical Theory of Communication*(University of Illinois Press, Urbana,1963).

[5] A. Renyi, *Probability Theory*, North-Holland, Amsterdam (1970).

[6] J.Havrda and F. Charvat, Kybernetica **3**, 30 (1967); Z.Daróczy, Information and Control, **16**,36 (1970).

[7] B.D. Sharma and D.P. Mittal, J. Math. Sci. **10**, 28 (1975).

[8] C.Beck and F.Schlögl, Thermodynamics of chaotic Systems, Cambridge University Press, Cambridge (1993).

[9] A.I. Khinchin, Mathematical Foundation of Information Theory, Dover, New York, 1957.

[10] R.J.V. dos Sanntos, J. Math. Phys. **38**, 4104 (1997).

[11] S.Abe, Phys. Lett. A **271**, 74 (2000).

[12] C. Tsallis, R.S. Mendes and A.R. Plastino, Physica A **261**, 534 (1998).

[13] C. Tsallis, *Quimica Nova* **17**, 468 (1994).

[14] E.P. Borges, J.Phys.A **31**, 5281 (1998).

[15] T. Yamano, unpublised.

[16] P.T. Landsberg and V. Vedral, Phys. Lett. A **247**, 211 (1998).

[17] A.K. Rajagopal and S. Abe, Phys. Rev. Lett. **83**, 1711 (1999).

[18] T.Yamano, Information theory based on nonadditive information content, Phys. Rev. E,**63**, 046105(2001).

[19] C.Tsallis, Phys. Rev. E **58**, 1442 (1998).

[20] L. Borland, A.R. Plastino, C. Tsallis, J. Math. Phys. **39**, 6490 (1998); Errata:*ibid* **40**, 2196 (1999).

[21] S. Kullback and R.A. Leibler, Ann. Math. Stat. **22**, 79 (1951); S. Kullback, Information Theory and Statistics (Wiley,New York,1959).

[22] T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley,New York,1991).