Mutual Information and Nonadditive Entropies: The Case of Tsallis Entropy

Amelia Carolina Sparavigna

Abstract: Mutual information of two random variables can be easily obtained from their Shannon entropies. However, when nonadditive entropies are involved, the calculus of the mutual information is more complex. Here we discuss the basic matter about information from Shannon entropy. Then we analyse the case of the generalized nonadditive Tsallis entropy.

Keywords: Mutual Information, Entropy, Tsallis Entropy, Generalized Additivity, Image Registration

1. Introduction
In many applications of engineering and telecommunication, it is often desired to increase or decrease the dependency of two random variables. This dependency is linked to the mutual information, which is its measure. The mutual information can be easily decomposed into a sum (with signs) of entropies [1], when the Shannon entropy is used. This entropy is additive, that is, when we have independent subsystems, X and Y, the entropy of their union is $S(X,Y)=S(X)+S(Y)$. When nonadditive entropies are involved, the approach to find the mutual information is not so simple. Moreover, it is often preferred to use the term “mutual entropy”, instead of “mutual information” [2,3].

The mutual entropy contains conditional entropies, which must be carefully defined when entropies are nonadditive [3]. Among the nonadditive entropies, we have the Tsallis entropy. It is a generalization of the standard Boltzmann–Gibbs entropy, introduced in 1988 as a basis for generalizing the standard statistical mechanics [4,5]. Due to its entropic index, which can be used as a tuning parameter, this entropy is involved in several applications, in particular for image processing and image registration [6]. Here we discuss the basic matter concerning the mutual information when Tsallis entropy is involved.

2. Mutual information
The mutual information of two random variables in subsystems X,Y is providing a measure of the mutual dependence of the variables. This can be viewed intuitively as a measure of information that X and Y are sharing. In particular, it measures how much knowing one of these variables reduces uncertainty about the other [7-9]. In [9], examples are explaining this quantity. If X and Y are independent, knowing X does not give any information about Y and vice versa: the mutual information is zero. If $Y=f(X)$ or $X=f(Y)$, where f indicates a deterministic function, all information conveyed by X (or Y) is shared with Y (or X): the mutual information is the same as the uncertainty contained in Y (or X) alone, which is measured by the entropy of Y (or X). The mutual information is then the entropy of Y (or X) [9].

Figure 1: Venn diagram of mutual information $I(X;Y)$ associated with correlated variables X and Y. The area contained by both circles is the joint entropy $H(X,Y)$. The lower circle is the individual entropy $H(X)$, such as the upper circle is that of $H(Y)$.

The physical meaning of the mutual information $I(X;Y)$ as “the reduction of the uncertainty of X due to knowledge of Y” (or vice versa) [9], can be depicted in a Venn diagram (Figure 1). In this


Mutual Information and Nonadditive Entropies: The Case of Tsallis Entropy

For X,Y independent variables

MT' (X; Y) = T(X) + T(Y) − T(X, Y)  \quad (4)

If X,Y are independent, we must have a mutual information equal to zero. However, from (4), we find:

MT'' (X; Y) = (q − 1)T(X)T(Y)  \quad (5)

This happens because the generalized additivity for independent subsystems is:

T(X, Y) = T(X) + T(Y) + (1 − q)T(X)T(Y).

(6)

As a result, for an entropic index different from 1, (4) should give a result different from zero. Therefore, (4) is not good for representing the mutual information I(X;Y).

In his paper, Tsallis is discussing the problem of correlated systems too [4]. He used the Rényi entropy for correlated systems:

\[ \Gamma = R(X) + R(Y) − R(X, Y) \quad (7) \]

Since the Rényi entropy is additive, if X,Y are independent, \( \Gamma \) is equal to zero. Let us note that it is function \( \Gamma \) which seems working as the mutual information.

However, for the non-additive Tsallis conditional entropy, a quite simple formula was given in [14]:

\[
T(X \mid Y) = \frac{T(X, Y) − T(Y)}{1 + (1−q)T(Y)} \quad (8a)
\]

\[
T(Y \mid X) = \frac{T(X, Y) − T(X)}{1 + (1−q)T(X)} \quad (8b)
\]

We could define the mutual entropy, as for the Shannon entropy, in the following manner:

\[
MT^{\Gamma} (X; Y) = T(X) − T(X \mid Y)
= T(X) − \frac{T(X, Y) − T(Y)}{1 + (1−q)T(Y)} \quad (9a)
\]

\[
MT^{\Gamma} (Y; X) = T(Y) − T(Y \mid X)
= T(Y) − \frac{T(X, Y) − T(X)}{1 + (1−q)T(X)} \quad (9b)
\]

For X,Y independent variables, using (9a) for instance:

\[
MT^{\Gamma} (X; Y) = (q − 1)T(X)T(Y)
\]

Using Tsallis entropy, it is possible to define more generalized entropies too, and, among them, the nonadditive entropies. How can we generalize the mutual information in this case? In [3], it is preferred the use of the so-called Tsallis mutual entropy, defined as:

\[
MT(X; Y) = T(X) − T(X \mid Y)
= T(Y) − T(Y \mid X) \quad (2)
\]

In (2), T is referring to the Tsallis entropy. According to [4], \( T(X,Y)=T(X)+T(Y|X) \) and \( T(Y,X)=T(Y)+T(X|Y) \). Let us remember that Tsallis entropy \( T \) and Rényi entropy \( R \), [13], are linked by the following equation:

\[
R = \frac{\ln[1 + (1−q)T]}{1−q} \quad (3)
\]

Here \( q \) is the entropic index. As \( q \) approaches 1, the Tsallis entropy becomes the Shannon entropy.

Let us try defining the Tsallis mutual entropy as:

\[
MT^{\Gamma} (X; Y) = T(X) + T(Y) − T(X, Y) \quad (4)
\]

If X,Y are independent, we must have a mutual information equal to zero. However, from (4), we find:

\[
MT^{\Gamma} (X; Y) = (q − 1)T(X)T(Y) \quad (5)
\]

This happens because the generalized additivity for independent subsystems is:

\[
T(X, Y) = T(X) + T(Y) + (1 − q)T(X)T(Y).
\]

(6)

As a result, for an entropic index different from 1, (4) should give a result different from zero. Therefore, (4) is not good for representing the mutual information I(X;Y).

In his paper, Tsallis is discussing the problem of correlated systems too [4]. He used the Rényi entropy for correlated systems:

\[
\Gamma = R(X) + R(Y) − R(X, Y) \quad (7)
\]

Since the Rényi entropy is additive, if X,Y are independent, \( \Gamma \) is equal to zero. Let us note that it is function \( \Gamma \) which seems working as the mutual information.

However, for the non-additive Tsallis conditional entropy, a quite simple formula was given in [14]:

\[
T(X \mid Y) = \frac{T(X, Y) − T(Y)}{1 + (1−q)T(Y)} \quad (8a)
\]

\[
T(Y \mid X) = \frac{T(X, Y) − T(X)}{1 + (1−q)T(X)} \quad (8b)
\]

We could define the mutual entropy, as for the Shannon entropy, in the following manner:

\[
MT^{\Gamma} (X; Y) = T(X) − T(X \mid Y)
= T(X) − \frac{T(X, Y) − T(Y)}{1 + (1−q)T(Y)} \quad (9a)
\]

\[
MT^{\Gamma} (Y; X) = T(Y) − T(Y \mid X)
= T(Y) − \frac{T(X, Y) − T(X)}{1 + (1−q)T(X)} \quad (9b)
\]

For X,Y independent variables, using (9a) for instance:
MT**(X; Y) \\
= \frac{T(X) + F_{X,Y} - T(X) - T(Y) - F_{X,Y} + T(Y)}{1 + (1-q)T(Y)} \\
= 0

Where \( F_{X,Y} = (1-q)T(X)T(Y) \). The same for (9b).

**Figure 2**: Venn diagram when \( Y \) is completely dependent on \( X \).

### 4. The problem of symmetry

Let us note that \( I(X; Y) \) must be symmetric, that is \( I(X; Y) = I(Y; X) \), “otherwise it would not be mutual information” [15]. Are (9a) and (9b) giving us the same mutual information? Let us consider the case when \( Y \) is completely dependent on \( X \), as depicted in the Figure 2.

In this case \( H(X,Y) = H(X) \), so we have \( T(X,Y) = T(X) \), and the mutual information is \( T(Y) \). In the following discussion, we have \( X \) larger that \( Y \). Then \( T(X) \) larger than \( T(Y) \). Let us consider (9a) and (9b):

\[
MT^+(X; Y) = \frac{1}{1 + (1-q)T(Y)} T(X) - T(Y) \] \quad (10a)
\[
MT^-(Y; X) = T(Y) \] \quad (10b)

Therefore, (9a)-(9b) do not fulfil the required symmetry. In fact, if we consider the Figure 2, it is clear that a non-symmetric situation exists.

Aiming to solve the case of Figure 2, we could modify these mutual Tsallis entropies so that:

\[
MT(X; Y) = (1 + (1-q)T(Y)) \left[ T(X) - \frac{T(X,Y) - T(Y)}{1 + (1-q)T(Y)} \right] \frac{1}{1 + (1-q)T(X)} \] \quad (11a)
\[
MT(Y; X) = (1 + (1-q)T(X)) \left[ T(Y) - \frac{T(X,Y) - T(X)}{1 + (1-q)T(X)} \right] \frac{1}{1 + (1-q)T(Y)} \] \quad (11b)

We can see that:

\[
MT(Y; X) = \frac{T(X) + T(Y) - T(X,Y) + (1-q)T(X)T(Y)}{1 + (1-q)T(X)} \] \quad (12)

So we have a mutual information, which is symmetric. If \( X \) and \( Y \) are independent: \( MT(X; Y) = 0 \).

For completely dependent variable and \( T(X,Y) = T(X) \):

\[
MT(X; Y) = MT(Y; X) = \frac{T(Y) + (1-q)T(X)T(Y)}{1 + (1-q)T(X)} = T(Y) \] \quad (13)

In the case \( X \) is coincident with \( Y \), we have \( T(X) = T(Y) \):

\[
MT(X; Y) = \frac{T(X) + (1-q)T(X)T(Y)}{1 + (1-q)T(X)} = T(Y) = T(X) \] \quad (14)

However, we can have also the case when \( X \) is totally inside \( Y \) (Figure 3).
Then, let us write the join entropy as $T(Y, X)$, to remark this situation. We have that the mutual information must be equal to $T(X)$. We obtain this in a symmetric entropy, when we define the mutual entropy as:

$$MT(X; Y) = MT(Y; X) = \frac{T(X) + T(Y) - T(Y, X) + (1-q)T(X)T(Y)}{1 + (1-q)T(Y)} \quad (15)$$

Then, when $T(Y, X) = T(Y)$, from (15) we obtain $MT(X; Y) = T(X)$. We can define, in the same manner as in Ref.4 is proposed the correlation term in the mutual Tsallis entropy:

$$MT(X; Y) = MT(Y; X) = \frac{T(X) + T(Y) - T(Y, X) + (1-q)T(X)T(Y)}{1 + (1-q)[\max\{T(X), T(Y)\}]} \quad (16)$$

This can be an expression of the mutual information for Tsallis entropy, which is properly answering to the fundamental requirements of computation. We have discussed such an approach, because it can be easily applied to another nonadditive entropy, the Kaniadakis entropy, to determine its mutual information [16]. This problem will be addressed in a following paper.

References

Figure 3: Venn diagram when Y is completely dependent on X.