Alternative Entropy Measures and Generalized Khinchin–Shannon Inequalities

Rubem P. Mondaini * and Simão C. de Albuquerque Neto

COPPE, Centre of Technology, Federal University of Rio de Janeiro, Rio de Janeiro 21941-901, Brazil; simao.albuquerque@coppe.ufrj.br
* Correspondence: Rubem.Mondaini@ufrj.br

Abstract: The Khinchin–Shannon generalized inequalities for entropy measures in Information Theory, are a paradigm which can be used to test the Synergy of the distributions of probabilities of occurrence in physical systems. The rich algebraic structure associated with the introduction of escort probabilities seems to be essential for deriving these inequalities for the two-parameter Sharma–Mittal set of entropy measures. We also emphasize the derivation of these inequalities for the special cases of one-parameter Havrda–Charvat’s, Rényi’s and Landsberg–Vedral’s entropy measures.

Keywords: entropy measures; synergy; probabilistic distributions

1. Introduction

In the present contribution we derive the Generalized Khinchin–Shannon inequalities (GKS) [1,2] associated to entropy measures of the Sharma–Mittal (SM) set [3]. We stress that the derivations to be presented here are a tentative way of implementing the ideas of the literature on interdisciplinary topics of Statistical Mechanics and Theory of Information [4–6]. The algebraic structure of the escort probability distributions on these derivations seems to be essential, contrary to the intuitive derivation of the usual Khinchin–Shannon inequalities for the Gibbs–Shannon entropy measures. We start on Section 2 with the construction of a generic probabilistic space with their elements—the probabilities of occurrence—arranged on blocks of \( m \) rows and \( n \) columns. It then follows the introduction of the definitions of simple, joint, conditional and marginal probabilities through the use of the Bayes’ law. In Section 3, we make use of the assumption of concavity in order to unveil the Synergy of the distribution of values of Gibbs–Shannon entropy measures [2]. In Section 4, we present the same development but for the SM set of entropy measures, after the introduction of the concept of escort probabilities. We then specialize the derivations to Havrda–Charvat’s, Rényi’s and Landsberg–Vedral’s entropies [7–9]. A detailed study is then undertaken in this section to treat the eventual ordering between the probabilities of occurrence and their associated escort probabilities. This is then enough for deriving the GKS inequalities for the SM entropy measures. In Section 5, we present a proposal for Information measure associated to SM entropies and we derive its related inequalities [10]. At this point we stress once more the upsurge of the synergy effect on the comparison of the information obtained from the entropy calculated with joint probabilities of occurrence and the entropies corresponding to simple probabilities. In Section 6, we present an alternative derivation of the GKS inequalities based on Hölder inequalities [11]. These can provide, in association with Bayes’ law, the same assumptions of concavity which have been used in Sections 3 and 4 and a consequent identical derivation of the GKS inequalities given in Section 4.

2. The Probability Space. Probabilities of Occurrence

We consider that the data could be represented on two-dimensional arrays of \( m \) rows and \( n \) columns. We then have \( m \times n \) blocks of data to undertake the statistical analysis.
The joint probabilities of occurrence of a set of $t$ variables $a_1, \ldots, a_t$ in columns $j_1, \ldots, j_t$, respectively, are given by

\[ p_{j_1 \ldots j_t}(a_1, \ldots, a_t) = \frac{n_{j_1 \ldots j_t}(a_1, \ldots, a_t)}{m}, \tag{1} \]

where $m$ is the number of rows in the subarray $m \times t$ of the array $m \times n$, and $n_{j_1 \ldots j_t}(a_1, \ldots, a_t)$ is the number of occurrences of the set $a_1, \ldots, a_t$. The values assumed by the variables $j_1, \ldots, j_t$, $j_1 < j_2 < \ldots < j_t$, are respectively given by:

\[
\begin{align*}
    j_1 &= 1, 2, \ldots, n - t + 1 \\
    j_2 &= j_1 + 1, j_1 + 2, \ldots, n - t + 2 \\
    &\vdots \\
    j_{t-1} &= j_{t-2} + 1, j_{t-2} + 2, \ldots, n - 1 \\
    j_t &= j_{t-1} + 1, j_{t-1} + 2, \ldots, n,
\end{align*}
\]

or,

\[
\begin{align*}
    j_1 &= 1, 2, \ldots, n - t + 1 \\
    j_2 &= j_1 + 1, j_1 + 2, \ldots, n - t + 2 \\
    &\vdots \\
    j_{t-1} &= j_1 + t - 2, j_1 + t - 1, \ldots, n - 1 \\
    j_t &= j_1 + t - 1, j_1 + t, \ldots, n.
\end{align*}
\]

There are then \( \binom{n}{t} = \frac{n!}{t!(n-t)!} \) objects of $t$ columns each, $1 \leq t \leq n$, and if the variables $a_1, \ldots, a_t$ take on values $1, \ldots, W$, then we will have $W^t$ components for each of these objects.

Since:

\[ \sum_{a_1, \ldots, a_t=1}^W n_{j_1 \ldots j_t}(a_1, \ldots, a_t) = m, \tag{4} \]

we can write:

\[ \sum_{a_1, \ldots, a_t=1}^W p_{j_1 \ldots j_t}(a_1, \ldots, a_t) = 1. \tag{5} \]

On the study of distributions of bases of nucleotides or distributions of amino acids in proteins, the related values of $W$ are $W = 4$ and $W = 20$, respectively.

The Bayes’ law for the probabilities of occurrence of Equation (1) is written as:

\[
\begin{align*}
    p_{j_1 \ldots j_t}(a_1, \ldots, a_t) &= p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1} | a_t) \cdot p_{j_t}(a_t) \\
    &= p_{j_1 \ldots j_{t-1}}(a_t | a_1, \ldots, a_{t-1}) \cdot p_{j_t \ldots j_{t-1}}(a_1, \ldots, a_{t-1}) \\
    &= p_{j_1 \ldots j_{t-1}}(a_t, a_1, \ldots, a_{t-1}), \tag{6}
\end{align*}
\]

where $p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1} | a_t)$ stands for the conditional probability of occurrence of the values associated to the variables $a_1, \ldots, a_{t-1}$ in the columns $j_1, \ldots, j_{t-1}$, respectively, if the values associated to $a_t$ in the $j_t$th column are given a priori. This also means that:

\[ \sum_{a_1, \ldots, a_{t-1}=1}^W p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1} | a_t) = 1. \tag{7} \]
The marginal probabilities related to $p_{j_1...j_l}(a_1,\ldots,a_l)$ are then given by

$$p_{j_l}(a_l) = \sum_{a_{l-1},\ldots,a_1} p_{j_1...j_l}(a_1,\ldots,a_l)$$
$$= \sum_{a_{l-1},\ldots,a_1} p_{j_1...j_l}(a_1,\ldots,a_l-1|a_l) p_{j_l}(a_l).$$  \hspace{1cm} (8)

We then have from Equations (6) and (8):

$$1 = \sum_{a_l} p_{j_l}(a_l) = \sum_{a_{l-1},\ldots,a_1} p_{j_1...j_l}(a_1,\ldots,a_l),$$  \hspace{1cm} (9)

which is the same result of Equation (5).

3. The Assumption of Concavity and the Synergy of Gibbs–Shannon Entropy Measures

A concave function of several variables should satisfy the following inequality:

$$\sum \lambda_l f(x_l) \leq f\left(\sum \lambda_l x_l\right); \quad \lambda_l \geq 0; \quad \sum \lambda_l = 1. \hspace{1cm} (10)$$

We shall apply Equation (10) to the Gibbs–Shannon entropies:

$$S_{j_1...j_l} = - \sum_{a_{l-1},\ldots,a_1} p_{j_1...j_l}(a_1,\ldots,a_l) \log p_{j_1...j_l}(a_1,\ldots,a_l),$$  \hspace{1cm} (11)

$$S_{j_1...j_{l-1}|j_l} = - \sum_{a_{l-1},\ldots,a_1} p_{j_l}(a_l) p_{j_1...j_{l-1}|j_l}(a_1,\ldots,a_{l-1}|a_l) \log p_{j_1...j_{l-1}|j_l}(a_1,\ldots,a_{l-1}|a_l).$$  \hspace{1cm} (12)

where Equation (12) stands for the definition of Gibbs–Shannon entropy which is related to the conditional probabilities $p_{j_1...j_l}(a_1,\ldots,a_{l-1}|a_l)$. It is a measure of the uncertainty [2] on the distribution of probabilities of the columns $j_1,\ldots,j_{l-1}$, when we have previous information on the distribution of the column $j_l$.

From Bayes’ law, Equation (6) and from Equations (8), (11) and (12), we get:

$$S_{j_1...j_l} = S_{j_1...j_{l-1}|j_l} + S_{j_l}. \hspace{1cm} (13)$$

We now use the correspondences:

$$\lambda_l \leftrightarrow p_{j_l}(a_l); \quad x_l \leftrightarrow p_{j_1...j_l}(a_1,\ldots,a_{l-1}|a_l),$$

$$f(x_l) \leftrightarrow - \sum_{a_{l-1},\ldots,a_1} p_{j_1...j_l}(a_1,\ldots,a_{l-1}|a_l) \log p_{j_1...j_l}(a_1,\ldots,a_{l-1}|a_l),$$  \hspace{1cm} (14)

and we then have:

$$\sum \lambda_l x_l \leftrightarrow \sum_{a_l} p_{j_l}(a_l) p_{j_1...j_l}(a_1,\ldots,a_{l-1}|a_l) = \sum_{a_l} p_{j_1...j_l}(a_1,\ldots,a_l) = p_{j_1...j_l}(a_1,\ldots,a_{l-1}),$$  \hspace{1cm} (15)

$$\sum \lambda_l f(x_l) \leftrightarrow - \sum_{a_l,a_{l-1},\ldots,a_1} p_{j_l}(a_l) p_{j_1...j_{l-1}|j_l}(a_1,\ldots,a_{l-1}|a_l) \log p_{j_1...j_{l-1}|j_l}(a_1,\ldots,a_{l-1}|a_l),$$  \hspace{1cm} (16)

and

$$f\left(\sum \lambda_l x_l\right) \leftrightarrow - \sum_{a_l,a_{l-1},\ldots,a_1} p_{j_1...j_{l-1}}(a_1,\ldots,a_{l-1}) \log p_{j_1...j_{l-1}}(a_1,\ldots,a_{l-1}).$$  \hspace{1cm} (17)

After substituting Equations (12), (16) and (17) into Equation (10), we get:
where

\[ S_{j_{1} \ldots j_{t-1}} = - \sum_{a_{1}, \ldots, a_{t}} p_{j_{1}}(a_{1}) \cdots p_{j_{t-1}}(a_{t-1}) \log p_{j_{1}}(a_{1}) \cdots p_{j_{t-1}}(a_{t-1}) \]

or,

\[ S_{j_{1} \ldots j_{t-1}} \leq S_{j_{1} \ldots j_{t-1}}. \]

This means that the uncertainty of the distribution on the columns \( j_{1}, \ldots, j_{t-1} \) cannot be increased when we have previous information on the distribution of column \( j_{t} \).

From Equations (13) and (19), we then write:

\[ S_{j_{1} \ldots j_{t}} \leq S_{j_{1} \ldots j_{t-1}} + S_{j_{t}}, \]

and by iteration we get the Khinchin–Shannon inequality for the Gibbs–Shannon entropy measure:

\[ S_{j_{1} \ldots j_{t}} \leq \sum_{t=1}^{T} S_{j_{t}}. \]

The usual meaning given to Equation (21) is that the minimum of the information to be obtained from the analysis of the joint probabilities of a set of \( t \) columns is given by the sum of the informations associated with the \( t \) columns if considered as independent \([1,2,10]\). This is also seen as an aspect of Synergy \([12,13]\) of the distribution of probabilities of occurrence.

4. The Assumption of Concavity and the Synergy of Sharma–Mittal (SM) Entropy Measures. The GKS Inequalities

We shall now use the assumption of concavity given by Equation (10) on Sharma–Mittal (SM) entropy measures:

\[ (SM)_{j_{1} \ldots j_{t}} = \left( \frac{(a_{j_{1} \ldots j_{t}})^{\frac{1}{r}}}{1 - r} - 1 \right), \]

where,

\[ a_{j_{1} \ldots j_{t}} = \sum_{a_{1}, \ldots, a_{t}} \left[p_{j_{1}}(a_{1}) \cdots p_{j_{t-1}}(a_{t-1})\right]^{s}, \]

and \( r, s \) are non-dimensional parameters.

Analogously to Equation (12), we also introduce the “conditional entropy measure”

\[ (SM)_{j_{1} \ldots j_{t-1} | j_{t}} = \left( \frac{(\hat{\beta}_{j_{1} \ldots j_{t-1} | j_{t}})^{\frac{1}{r}}}{1 - r} - 1 \right), \]

where

\[ \hat{\beta}_{j_{1} \ldots j_{t-1} | j_{t}} = \sum_{a_{1}, \ldots, a_{t}} \hat{\beta}_{j_{t}}(a_{t}) \left[p_{j_{1}}(a_{1}) \cdots p_{j_{t-1}}(a_{t-1} | a_{t})\right]^{s}, \]

and \( \hat{\beta}_{j_{t}}(a_{t}) \) stands for the escort probability

\[ \hat{\beta}_{j_{t}}(a_{t}) = \left( \frac{p_{j_{t}}(a_{t})}{a_{j_{t}}} \right)^{s} = \left( \frac{p_{j_{t}}(a_{t})}{\sum_{a_{t}} (p_{j_{t}}(a_{t}))^{s}} \right)^{s}. \]

We have in general:

\[ \hat{\beta}_{j_{1} \ldots j_{t}}(a_{1}, \ldots, a_{t}) = \left( \frac{p_{j_{1} \ldots j_{t}}(a_{1}, \ldots, a_{t})}{a_{j_{1} \ldots j_{t}}} \right)^{s} = \left( \frac{p_{j_{1} \ldots j_{t}}(a_{1}, \ldots, a_{t})}{\sum_{a_{1}, \ldots, a_{t}} (p_{j_{1} \ldots j_{t}}(a_{1}, \ldots, a_{t}))^{s}} \right)^{s}. \]

The inverse transformations are given by:
\[ p_{j}(a_t) = \frac{ (\hat{p}_{j}(a_t))^{1/s} }{ \sum_{a_t} (\hat{p}_{j}(a_t))^{1/s} }, \quad (28) \]

\[ p_{j_{1},...j_{t}}(a_{1},...,a_{t}) = \frac{ (\hat{p}_{j_{1},...j_{t}}(a_{1},...,a_{t}))^{1/s} }{ \sum_{a_{1},...,a_{t}} (\hat{p}_{j_{1},...j_{t}}(a_{1},...,a_{t}))^{1/s} }, \quad (29) \]

with

\[ \sum_{a_{1},...,a_{t}} p_{j}(a_{t}) = 1 = \sum_{a_{1},...,a_{t}} \hat{p}_{j}(a_{t}) , \quad (30) \]

\[ \sum_{a_{1},...,a_{t}} p_{j_{1},...j_{t}}(a_{1},...,a_{t}) = 1 = \sum_{a_{1},...,a_{t}} \hat{p}_{j_{1},...j_{t}}(a_{1},...,a_{t}). \quad (31) \]

A range of variation for the parameters \( r, s \) of Sharma–Mittal entropies, Equation (22), should be derived from a requirement for strict concavity. In order to do so, let us remember that for each set of \( t \) columns \((m \times t)\) subarray there are \( m \) rows of \( t \) values each \( (t\text{-sequences}) \). We now denote these \( t \)-sequences by:

\[ (a^{q_0}_1, ..., a^{q_0}_m), \quad \mu = 1, ..., m \quad (32) \]

A sufficient requirement for strict concavity is the negative definiteness of the quadratic form associated to the Hessian matrix [14], whose elements are given by:

\[ H_{\mu\nu q_0} = \frac{ \partial^2 (SM)_{j_{1},...j_{t}} }{ \partial p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \partial p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) }, \quad \mu, \nu = 1, ..., m. \quad (33) \]

We then consider the \( m \) submatrices along the diagonal of the Hessian matrix. Their determinants should be alternately negative or positive according to whether their order is odd or even [15], respectively:

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1) = s(\alpha_{j_{1},...j_{t}})^{s-2} \left[ [p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m)]^{s-2} - 1 \right], \quad (34) \]

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1, 2) = -s(\alpha_{j_{1},...j_{t}})^{s-2} \left[ p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \cdot p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \right]^{s-2} \left\{ \frac{s(s-r)}{1-s} \cdot [\hat{p}_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) + \hat{p}_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m)] - 1 \right\}, \quad (35) \]

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1, 2, 3) = s^3(\alpha_{j_{1},...j_{t}})^{3(s-2)} \left[ p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \cdot p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \cdot p_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) \right]^{s-2} \left\{ \frac{s(s-r)}{1-s} \cdot [\hat{p}_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) + \hat{p}_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m) + \hat{p}_{j_{1},...j_{t}}(a^{q_0}_1, ..., a^{q_0}_m)] - 1 \right\}. \quad (36) \]

We then choose:

\[ 1 > r \geq s > 0, \quad (37) \]

and we have from Equations (34)–(36):

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1) < 0, \quad (38) \]

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1, 2) > 0, \quad (39) \]

\[ \det H_{\mu\nu q_0} (\mu, \nu = 1, 2, 3) < 0. \quad (40) \]
We then have generally:

\[
\det H_{\mu_0, \nu_0}, (\mu_0, \nu_0 = 1, \ldots, m) = (-1)^{m-1} \cdot \sum_{j_1, \ldots, j_m} \left[ \prod_{\mu=1}^{m} p_{j_1, \ldots, j_m, \mu} \right]^{m(\frac{1}{s} - 1)} \left\{ \frac{s(s-r)}{(1-s)^2} \sum_{\mu=1}^{m} \hat{\rho}_{j_1, \ldots, j_m, \mu} - 1 \right\}. \tag{41}
\]

This completes the proof.

From Bayes’ law, Equation (6) and from Equations (8), (22)–(25), we can write:

\[
(SM)_{j_1, \ldots, j_l} = (SM)_{j_1, \ldots, j_l-1} + (SM)_{j_1} + (1-r)(SM)_{j_1, \ldots, j_l} \cdot (SM)_{j_l}. \tag{42}
\]

We are now ready to use the concavity assumption, Equation (10) for deriving the GKS inequalities. In order to do so, we make the correspondences:

\[
\lambda_j \leftrightarrow \hat{\rho}_j(a_i); \quad x_j \leftrightarrow p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1), \quad f(x_j) \leftrightarrow \sum_{a_{t-1}} \left[ p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1) \right]^s. \tag{43}
\]

We can then write:

\[
\sum_j \lambda_j x_j \leftrightarrow \sum_{a_i} \hat{\rho}_j(a_i) p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1), \quad \sum_j \lambda_j f(x_j) \leftrightarrow \sum_{a_i, a_{t-1}} \hat{\rho}_j(a_i) \left[ p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1) \right]^s, \tag{44}
\]

and

\[
f\left( \sum_j \lambda_j x_j \right) \leftrightarrow \sum_{a_i, a_{t-1}} \left[ \sum_{a_i} \hat{\rho}_j(a_i) p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1) \right]^s. \tag{45}
\]

With the correspondences above, Equation (10) will turn into:

\[
\sum_{a_1, \ldots, a_{t-1}} \left[ \sum_{a_i} \hat{\rho}_j(a_i) p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1) \right]^s \geq \sum_{a_1, \ldots, a_{t-1}} \left[ \sum_{a_i} \hat{\rho}_j(a_i) p_{j_1, \ldots, j_l | a_1, \ldots, a_{t-1}}(a_1) \right]^s. \tag{46}
\]

An additional information should be taken into consideration before we derive the GKS inequalities:

On each \( j_i \) column, \( t = 1, \ldots, n \) of a \( m \times n \) block, there will be values \( a_i', a_i'' \) of \( a_i \) such that

\[
p_{j_i}(a_i') \geq \hat{\rho}_j(a_i'), \quad p_{j_i}(a_i'') \leq \hat{\rho}_j(a_i''). \tag{48}
\]

After multiplying inequalities (48) and (49) by \( p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i') \) and \( p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i'') \), respectively, and summing up in \( a_i' \) and \( a_i'' \), respectively, we get:

\[
\sum_{a_i'} p_{j_i}(a_i') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i') \geq \sum_{a_i'} \hat{\rho}_j(a_i') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i'), \tag{50}
\]

and

\[
\sum_{a_i''} p_{j_i}(a_i'') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i'') \leq \sum_{a_i''} \hat{\rho}_j(a_i'') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i''). \tag{51}
\]

From Equations (48) and (49), any sum over the \( a_i \) values can be partitioned as sums over the sets of values \( a_i' \) and \( a_i'' \):

\[
\sum_{a_i} p_{j_i}(a_i) p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i) = \sum_{a_i'} p_{j_i}(a_i') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i') + \sum_{a_i''} p_{j_i}(a_i'') p_{j_1, \ldots, j_l}(a_1, \ldots, a_{t-1} | a_i''). \tag{52}
\]
Substituting Equation (52) into Equations (50) and (51), we have:

\[
\sum_{a_t} p_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) - \sum_{a_t'} p_j(a_t') p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') \\
\geq \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) - \sum_{a_t'} \hat{p}_j(a_t') p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') , \tag{53}
\]

and

\[
\sum_{a_t} p_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) - \sum_{a_t'} p_j(a_t') p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') \\
\leq \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) - \sum_{a_t'} \hat{p}_j(a_t') p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') , \tag{54}
\]

respectively.

After applying the Bayes’ law, Equation (6), to the first term on the left hand side of Equations (53) and (54), we get:

\[
p_{j \rightarrow t-1}(a_1, \ldots, a_{t-1}) \geq \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) + B'' , \tag{55}
\]

\[
p_{j \rightarrow t-1}(a_1, \ldots, a_{t-1}) \leq \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) + B' , \tag{56}
\]

where

\[
B'' = \sum_{a_t'} (p_j(a_t') - \hat{p}_j(a_t')) \cdot p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') , \tag{57}
\]

\[
B' = \sum_{a_t'} (p_j(a_t') - \hat{p}_j(a_t')) \cdot p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t') , \tag{58}
\]

and we have, \( B'' \leq 0, B' \geq 0 \), according to Equations (48) and (49), respectively.

After taking the \( s \)-power in Equations (55) and (56) and summing up in \( a_1, \ldots, a_{t-1} \), we have:

\[
\sum_{a_1, \ldots, a_{t-1}} \left[ p_{j \rightarrow t-1}(a_1, \ldots, a_{t-1}) \right]^s \geq \sum_{a_1, \ldots, a_{t-1}} \left[ \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) + B'' \right]^s , \tag{59}
\]

\[
\sum_{a_1, \ldots, a_{t-1}} \left[ p_{j \rightarrow t-1}(a_1, \ldots, a_{t-1}) \right]^s \leq \sum_{a_1, \ldots, a_{t-1}} \left[ \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) + B' \right]^s . \tag{60}
\]

We now write, the concavity assumption, Equation (10), as:

\[
C \geq D , \tag{61}
\]

where

\[
C = \sum_{a_1, \ldots, a_{t-1}} \left[ \sum_{a_t} \hat{p}_j(a_t) p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) \right]^s , \tag{62}
\]

and

\[
D = \sum_{a_1, \ldots, a_{t-1}} \hat{p}_j(a_t) \left[ p_{j \rightarrow t}(a_1, \ldots, a_{t-1} | a_t) \right]^s . \tag{63}
\]

Equations (59) and (60) are now written as:

\[
A \geq C'' , \tag{64}
\]

\[
A \leq C' , \tag{65}
\]

where
\[ A \equiv \sum_{a_1, \ldots, a_t} \left[ p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}) \right]^s, \quad (66) \]

and

\[ C'' \equiv \sum_{a_1, \ldots, a_t} \left[ \sum_{a_t} \hat{p}_j(a_t) p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}|a_t) + B'' \right]^s, \quad (67) \]

\[ C' \equiv \sum_{a_1, \ldots, a_t} \left[ \sum_{a_t} \hat{p}_j(a_t) p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}|a_t) + B' \right]^s. \quad (68) \]

Since \( B'' \leq 0 \) and \( B' \geq 0 \), we have trivially that:

\[ C'' \leq C, \quad (69) \]

and

\[ C' \geq C. \quad (70) \]

The set of inequalities, Equations (61), (64) and (69), or

\[ C \geq D; \quad A \geq C''; \quad C'' \leq C, \quad (71) \]

and the set of inequalities, Equations (61), (65) and (70), or

\[ C \geq D; \quad A \leq C'; \quad C' \geq C, \quad (72) \]

can be arranged as the chains of inequalities

\[ C \geq A \geq D \geq C'', \quad (73) \]

and

\[ C' \geq A \geq C \geq D, \quad (74) \]

respectively.

The inequality \( A \geq D \) is common to the two chains above and it can be written as:

\[ \sum_{a_1, \ldots, a_t} \left[ p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}) \right]^s \geq \sum_{a_1, \ldots, a_t} \hat{p}_j(a_t) \left[ p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}|a_t) \right]^s. \quad (75) \]

From the definition of the escort probabilities, Equations (26) and (27), we can write the right-hand side of Equation (75), as:

\[ \sum_{a_1, \ldots, a_t} \hat{p}_j(a_t) \left[ p_{j_1 \ldots j_t}(a_1, \ldots, a_{t-1}|a_t) \right]^s = \frac{\sum_{a_1, \ldots, a_t} \left[ p_{j_1 \ldots j_t}(a_1, \ldots, a_t) \right]^s}{\sum_{a_t} \left[ p_j(a_t) \right]^s}. \quad (76) \]

From Equations (75) and (76) and the definition of the \( \alpha \)-symbols, Equation (23), we have:

\[ \alpha_{j_1 \ldots j_t} \leq \alpha_{j_1 \ldots j_{t-1}} \cdot \alpha_{j_t}. \quad (77) \]

We then get by iteration,

\[ \alpha_{j_1 \ldots j_t} \leq \prod_{i=1}^t a_{j_i}. \quad (78) \]

Equation (78) do correspond to the Generalized Khinchin–Shannon Inequalities (GKS) here derived for Sharma–Mittal entropies.

From Equations (22) and (37) we can also write for the GKS inequalities:

\[ (SM)_{j_1 \ldots j_t} \leq \prod_{i=1}^t \left[ 1 + (1 - r)(SM)_{j_i} \right] - 1 \quad \frac{1}{1 - r}. \quad (79) \]
The same words which have been written after Equation (21), could be written also here for the Sharma–Mittal entropy measures as the aspect of Synergy is concerned. We will introduce a proposal for information measure to stress this aspect on the next section.

For \( t = 2, 3 \), we can write from Equation (79):

\[
(SM)_{h_1 h_2} \leq (SM)_{j_1} + (SM)_{j_2} + (1 - r)(SM)_{j_1} (SM)_{j_2},
\]

(80)

\[
(SM)_{j_1 h_2 h_3} \leq (SM)_{j_1} + (SM)_{j_2} + (SM)_{j_3} + (1 - r)\left[(SM)_{j_1} (SM)_{j_2} + (SM)_{j_2} (SM)_{j_3} + (SM)_{j_1} (SM)_{j_3} \right] + (1 - r)^2(SM)_{j_1} (SM)_{j_2} (SM)_{j_3}.
\]

(81)

The Havrda–Charvat’s, Rényi’s and Landsberg–Vedral’s entropies are easily obtained by taking the convenient limits in Equation (24):

\[
\lim_{r \to 1} (SM)_{j_1 \ldots j_t} = (HC)_{j_1 \ldots j_t} = \frac{\alpha_{j_1 \ldots j_t} - 1}{1 - s},
\]

(82)

\[
\lim_{r \to 1} (SM)_{j_1 \ldots j_t} = R_{j_1 \ldots j_t} = \frac{\log(\alpha_{j_1 \ldots j_t})}{1 - s},
\]

(83)

\[
\lim_{r \to 2 - s} (SM)_{j_1 \ldots j_t} = (LV)_{j_1 \ldots j_t} = \frac{\alpha_{j_1 \ldots j_t} - 1}{(1 - s)\alpha_{j_1 \ldots j_t}}.
\]

(84)

The Gibbs–Shannon entropy measure \( S_{j_1 \ldots j_t} \), Equation (11), is included in all these entropies through:

\[
\lim_{s \to 1} (HC)_{j_1 \ldots j_t} = \lim_{s \to 1} R_{j_1 \ldots j_t} = \lim_{s \to 1} (LV)_{j_1 \ldots j_t} = S_{j_1 \ldots j_t}.
\]

(85)

Equations (83) and (85) have been derived via the l’Hôpital theorem. For \( t = 2, 3 \), we write from Equations (82)–(84):

\[
(HC)_{h_1 h_2} \leq (HC)_{j_1} + (HC)_{j_2} + (1 - s)(HC)_{j_1} (HC)_{j_2},
\]

(86)

\[
(HC)_{j_1 h_2 h_3} \leq (HC)_{j_1} + (HC)_{j_2} + (HC)_{j_3} + (1 - s)\left[(HC)_{j_1} (HC)_{j_2} + (HC)_{j_2} (HC)_{j_3} + (HC)_{j_1} (HC)_{j_3} \right] + (1 - s)^2(HC)_{j_1} (HC)_{j_2} (HC)_{j_3},
\]

(87)

\[
R_{h_1 h_2} \leq R_{j_1} + R_{j_2},
\]

(88)

\[
R_{h_1 h_2 h_3} \leq R_{j_1} + R_{j_2} + R_{j_3},
\]

(89)

\[
(LV)_{h_1 h_2 h_3} \leq (LV)_{j_1} + (LV)_{j_2} + (LV)_{j_3} - (1 - s)(LV)_{j_1} (LV)_{j_2}.
\]

(90)

\[
(LV)_{h_1 h_2 h_3} \leq (LV)_{j_1} + (LV)_{j_2} + (LV)_{j_3} - (1 - s)[(LV)_{j_1} (LV)_{j_2} + (LV)_{j_2} (LV)_{j_3} + (LV)_{j_1} (LV)_{j_3}] + (1 - s)^2(LV)_{j_1} (LV)_{j_2} (LV)_{j_3}.
\]

(91)

As a last result of this section, we note that Equation (79) could be also derived from Equation (75), since this equation could be also written as:

\[
(SM)_{j_1 \ldots j_{t-1}} \leq (SM)_{j_1 \ldots j_t},
\]

(92)

where we have used Equations (22)–(25).

After the comparison of Equation (42) and (92), we get the result of Equation (79) again.
5. An Information Measure Proposal Associated to Sharma–Mittal Entropy Measures

We are looking for a proposal of information measure which can fulfill a requirement of clear interpretation of the upsurge of Synergy in a probabilistic distribution and is supported by the usual idea of entropy as a measure of uncertainty.

For the Sharma–Mittal set of entropy measures the proposal for the associated information measure would be:

\[ I_{j_1 \ldots j_t} = -\frac{(SM)_{j_1 \ldots j_t}}{(a_{j_1 \ldots j_t})^{\frac{1}{1+r}}} , \quad 1 > r \geq s > 0 , \]  

(93)

where \((SM)_{j_1 \ldots j_t}\) and \(a_{j_1 \ldots j_t}\) are given by Equations (22) and (23). We then have from Equation (93)

\[ (SM)_{j_1 \ldots j_t} = - \frac{I_{j_1 \ldots j_t}}{1 + (1 - r)I_{j_1 \ldots j_t}} . \]  

(94)

From the GKS inequalities, Equation (79), and from Equation (94), we get:

\[ I_{j_1 \ldots j_t} \geq \frac{\prod_{l=1}^{t} \left[ 1 + (1 - r) I_{j_l} \right] - 1}{1 - r} \geq \sum_{l=1}^{t} I_{j_l} . \]  

(95)

The meaning of Equation (95) is that the minimum of information associated with \(t\) columns of probabilities of occurrence is given by the sum of informations associated to each column. This corresponds to the expression of Synergy of the distribution of probabilities of occurrences which we have derived on the previous section.

The inequalities (95), for \(t = 2, 3\) are written as:

\[ I_{j_1 j_2} \geq I_{j_1} + I_{j_2} + (1 - r) I_{j_1} I_{j_2} \geq I_{j_1} + I_{j_2} , \]  

(96)

\[ I_{j_1 j_2 j_3} \geq I_{j_1} + I_{j_2} + I_{j_3} + (1 - r) \left[ I_{j_1} I_{j_2} + I_{j_2} I_{j_3} + I_{j_1} I_{j_3} \right] + (1 - r)^2 I_{j_1} I_{j_2} I_{j_3} \] \[ \geq I_{j_1} + I_{j_2} + I_{j_3} , \]  

(97)

It seems worthwhile to derive yet another result which unveils once more the fundamental aspect of synergy of the distribution of probabilities of occurrence. From Equation (93), we have:

\[ dI_{j_1 \ldots j_t} = -\frac{d(SM)_{j_1 \ldots j_t}}{(a_{j_1 \ldots j_t})^{2\left(\frac{1}{1+s}\right)}} , \]  

(98)

and we then write from the GKS inequalities, Equation (78):

\[ \frac{1}{(a_{j_1 \ldots j_t})^{2\left(\frac{1}{1+s}\right)}} \geq \frac{1}{\prod_{l=1}^{t} (a_{j_l})^{2\left(\frac{1}{1+s}\right)}} = \prod_{l=1}^{t} \frac{1}{(a_{j_l})^{2\left(\frac{1}{1+s}\right)}} . \]  

(99)

We then get:

\[ -\frac{dI_{j_1 \ldots j_t}}{d(SM)_{j_1 \ldots j_t}} \geq \prod_{l=1}^{t} \left( -\frac{dI_{j_l}}{d(SM)_{j_l}} \right) . \]  

(100)

Equation (100) do correspond to another result which originates from the Synergy of the distribution of probabilities of occurrence. It can be written as: The minimum of the rate of information increase with decreasing entropy in probability distribution for sets of \(t\) columns, is given by the product of the rates of information increase pertaining to each of the \(t\) columns.
6. The Use of Hölder’s Inequality for an Alternative Derivation of the GKS Inequalities

We firstly note that:

\[
\sum_{a_t} \rho_j(a_t) p_j(a_1, \ldots, a_t) = \frac{\sum_{a_t} p_j(a_1, \ldots, a_t) (p_j(a_t))^{s-1}}{\sum_{a_t} (p_j(a_t))^s}, \quad (101)
\]

and we now introduce the Hölder’s inequality: Ref. [11]

\[
\left( \sum_{l} x_l y_l \right)^{ab} \leq \left( \sum_{l} (x_l)^a \right)^{1/b} \cdot \left( \sum_{l} (y_l)^b \right)^{1/a}, \quad a \neq 1. \quad (102)
\]

We can also write:

\[
\sum_{l} x_l y_l \geq \left( \sum_{l} (x_l)^a \right)^{1/a} \cdot \left( \sum_{l} (y_l)^b \right)^{1/b}, \quad 0 \leq a < 1, \quad 0 \leq b \leq 0. \quad (103)
\]

or

\[
\left( \sum_{l} x_l y_l \right)^a \geq \sum_{l} (x_l)^a \cdot \left( \sum_{l} (y_l)^b \right)^{a/b}, \quad 0 \leq a < 1, \quad b \leq 0. \quad (104)
\]

We now make the correspondences:

\[
x_l \leftrightarrow p_j(a_1, \ldots, a_t), \quad y_l \leftrightarrow (p_j(a_t))^{s-1}. \quad (105)
\]

We take the \(s\)-power of the sides of Equation (101) and after using Equations (104) and (105) with \(a = s\) and \(b = \frac{s}{s-1}\), we get:

\[
\left( \sum_{a_t} \rho_j(a_t) p_j(a_1, \ldots, a_t) \right)^s = \left( \frac{\sum_{a_t} p_j(a_1, \ldots, a_t) (p_j(a_t))^{s-1}}{\sum_{a_t} (p_j(a_t))^s} \right)^s \geq \left( \frac{\sum_{a_t} (p_j(a_t))^{s-1}}{\sum_{a_t} (p_j(a_t))^s} \right)^s \geq \sum_{a_t} (p_j(a_1, \ldots, a_t))^s \quad (106)
\]

After summing up in \(a_1, \ldots, a_{t-1}\), we then have:

\[
\sum_{a_1, \ldots, a_{t-1}} \left( \sum_{a_t} \rho_j(a_t) p_j(a_1, \ldots, a_t) \right)^s \geq \frac{\sum_{a_1, \ldots, a_t} (p_j(a_1, \ldots, a_t))^s}{\sum_{a_t} (p_j(a_t))^s}. \quad (107)
\]

From the application of Bayes’ law to the right-hand side of Equation (107), we get Equation (47) again, i.e., the correspondence of the assumption of concavity has been derived once more. All the subsequent development of Section 4 will then follows accordingly.

7. Concluding Remarks

It should be stressed that the introduction of escort probabilities has been efficient on the construction of generalized entropy measures. These can be used for the classification of databases in terms of their clustering as driven by their intrinsic synergy and the resulting formation of more complex structures like families and clans.
A fundamental aim would be the derivation of a dynamical theory which would be able to describe the process of formation of these structures. A theory based on the evolution of the entropy values on databases which we hope that could be realized by methods introduced by the exhaustive study of Fokker–Planck equations.

Some introductory results on this promising line of research have been already published [16] and a forthcoming publication of a comprehensive review will summarize all of them.

**Author Contributions:** Conceptualization, R.P.M. and S.C.d.A.N.; methodology, R.P.M. and S.C.d.A.N.; formal analysis, R.P.M. and S.C.d.A.N.; writing—original draft preparation, R.P.M.; writing—review and editing, R.P.M. and S.C.d.A.N.; visualization, R.P.M. and S.C.d.A.N.; supervision, R.P.M.; project administration, R.P.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

**Abbreviations**

The following abbreviations are used in this manuscript:

| Abbreviation | Description |
|--------------|-------------|
| GKS          | Generalized Khinchin–Shannon |
| SM           | Sharma–Mittal |

**References**

1. Mondaini, R.P.; de Albuquerque Neto, S.C. Khinchin–Shannon Generalized Inequalities for “Non-additive” Entropy Measures. In *Trends in Biomathematics* 2; Mondaini, R.P., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 177–190.
2. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover Publications: New York, NY, USA, 1957.
3. Sharma, B.D.; Mittal, D.P. New Non-additive Measures of Entropy for Discrete Probability Distributions. *J. Math. Sci.* **1975**, *10*, 28–40.
4. Valkenstein, M.V. *Entropy and Information*; Birkhäuser: Basel, Switzerland, 2009.
5. Beck, C. Generalized Information and Entropy Measures in Physics. *Contemp. Phys.* **2009**, *50*, 495–510. [CrossRef]
6. Lavenda, B.H. *A New Perspective on Thermodynamics*; Springer Science+Business Media: New York, NY, USA, 2010.
7. Havrda, J.; Charvat F. Quantification Method of Classification Processes. Concept of Structural $\alpha$-entropy. *Kybernetica* **1967**, *3*, 30–35.
8. Rényi, A. On Measures of Entropy and Information. In *Contributions to the Theory of Statistics, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 20 June–30 July 1960; Neyman, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
9. Landsberg, P.T.; Vedral, V. Distributions and Channel Capacities in Generalized Statistical Mechanics. *Phys. Lett. A* **1997**, *224*, 326–330. [CrossRef]
10. Mondaini, R.P.; de Albuquerque Neto, S.C. The Statistical Analysis of Protein Domain Family Distributions via Jaccard Entropy Measures. In *Trends in Biomathematics* 3; Mondaini, R.P., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 169–207.
11. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*; Cambridge University Press: London, UK, 1934.
12. Ay, N.; Olbrich, E.; Bertschinger, N.; Jost, J. A Geometric Approach to Complexity. *Chaos* **2011**, *21*, 037103. [CrossRef] [PubMed]
13. Olbrich, E.; Bertschinger, N.; Rauh, J. Information Decomposition and Synergy. *Entropy* **2015**, *17*, 3501–3517. [CrossRef]
14. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: New York, NY, USA, 2009.
15. Marsden, J.E.; Tromba, A. *Vector Calculus*; W. H. Freeman and Company Publishers: New York, NY, USA, 2012.
16. Mondaini, R.P.; de Albuquerque Neto, S.C. A Jaccard-like Symbol and its Usefulness in the Derivation of Amino Acid Distributions in Protein Domain Families. In *Trends in Biomathematics* 4; Mondaini, R.P., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 201–220.