Joint Range of Rényi Entropies

Peter Harremoës

Abstract—The exact range of the joined values of several Rényi entropies is determined. The method is based on topology with special emphasis on the orientation of the objects studied.

Index Terms—Rényi entropies, Shannon entropy, information diagram.

I. INTRODUCTION

Let $P = (p_1, p_2, ..., p_n)$ be a probability vector. For $\alpha \in [0; \infty \setminus \{1\}$ the Rényi entropy of $P$ of order $\alpha$ is defined by the equation

$$H_{\alpha} (P) = \frac{1}{1 - \alpha} \log \left( \sum_i p_i^\alpha \right).$$

This definition is extended by continuity so that

$$H_0 (P) = \log(\text{number of } p_i \neq 0),$$

$$H_1 = -\sum_i p_i \log p_i,$$

$$H_\infty (P) = \max p_i.$$

The Rényi entropy $H_0$ is essentially the Hartley entropy, and was one among other sources of inspiration to Shannon’s information theory. The Shannon entropy $H_1$ has proved its fundamental importance in information theory and related areas and has the nice property of being additive on product measures. The relation between $H_0$ and $H_1$ is given by the simple inequality

$$H_1 (P) \leq H_0 (P).$$

The Rényi entropy of order $\infty$ is essentially related to the “probability of error”. The relation between $H_1$ and $H_\infty$ was described independently in [7], [8], [2], [4] and [3]. The Rényi entropy $H_2$ is related to index of coincidence and Tsallis entropy and other quantities used for special purposes in crypto analysis, physics etc. (see [6] or [1] for references). The Shannon entropy and $H_2$ has been studied in [5] and in more detail in [6]. In [6] also the precise range of the joint values of any 2 Rényi entropies is described.

In source coding in finite systems one wants to avoid very long code words in variable length coding. In such systems the Rényi entropy of some order $\alpha < 1$ (depending on the memory of the system) determines how much the source can be compressed. In some cases joint values of $H_2 (P)$ and $H_3 (P)$ are measurable or computable and one is interested in bounds on $H_1$ (see [9] for references).

In order to get bounds on $H_1$ we are interested in the exact range of the mapping

$$\Phi : P \rightarrow (H_1 (P), H_2 (P), H_3 (P)).$$

In this paper the methods developed in [6] will be refined in order to be able to describe the joint range of any number of Rényi entropies. Only the range of $\Phi$ is described in detail, but the method can easily be generalized to any number of Rényi entropies.

II. REDUCTION TO MIXTURES OF UNIFORM DISTRIBUTIONS

The probability vector $P$ is parametrized as $(p_1, p_2, ..., p_n)$ where $p_j \geq 0$ and $\sum p_j = 1$. It will be assumed that $n$ is a fixed positive integer. In order to study the range of $P \subset (H_1 (P), H_2 (P), H_3 (P))$ we may assume that $p_1 \geq p_2 \geq ... \geq p_n$. The functional matrix of

$$P \rightarrow \left( \begin{array}{cccc} -1 & -1 & \cdots & -1 \\ -\frac{p_1}{\sum p_j} & -\frac{p_2}{\sum p_j} & \cdots & -\frac{p_n}{\sum p_j} \\ -\frac{3 p_1^2}{\sum p_j^2} & -\frac{3 p_2^2}{\sum p_j^2} & \cdots & -\frac{3 p_n^2}{\sum p_j^2} \end{array} \right).$$

Assume that $P$ has four different point probabilities, say $p_1 > p_2 > p_3 > p_4 > 0$. Put $x = p_1, y = p_2, z = p_3$ and $w = p_4$. Then

$$\begin{array}{cccc} -1 & 1 & 1 & 1 \\ -\frac{x^2}{2} & -\frac{y^2}{2} & -\frac{z^2}{2} & -\frac{w^2}{2} \\ -\frac{3 x^2}{\sum p_j^2} & -\frac{3 y^2}{\sum p_j^2} & -\frac{3 z^2}{\sum p_j^2} & -\frac{3 w^2}{\sum p_j^2} \end{array}$$

(1)

$$\begin{array}{cccc} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{x}{2} & \frac{y}{2} & \frac{z}{2} & \frac{w}{2} \\ \frac{x^2}{4} & \frac{y^2}{4} & \frac{z^2}{4} & \frac{w^2}{4} \end{array}$$

$$\begin{array}{cccc} -3 & 0 & 0 & 1 \\ -\frac{x y}{2} & -\frac{y z}{2} & -\frac{z w}{2} & -\frac{w x}{2} \\ -\frac{x^2 y}{4} & -\frac{y^2 z}{4} & -\frac{z^2 w}{4} & -\frac{w^2 x}{4} \end{array}$$

$$\begin{array}{cccc} 3 w^3 & -\frac{x}{w} & -\frac{y}{w} & -\frac{z}{w} \\ \frac{x^2}{w^2} & \frac{y^2}{w^2} & \frac{z^2}{w^2} & \frac{w^2}{w^2} \end{array}$$

$$\begin{array}{cccc} -\frac{3 w^2 - 1}{w} & -\frac{3 w^2 - 1}{w} & -\frac{3 w^2 - 1}{w} \\ \frac{x}{w^2} & \frac{y}{w^2} & \frac{z}{w^2} & \frac{w}{w^2} \end{array}$$
Put \( r = \frac{z}{w}, s = \frac{y}{w}, t = \frac{x}{w} \). Then \( r > s > t > 1 \) and the sign of

\[
\begin{vmatrix}
\log r & \log s & \log t \\
 r & s & t \\
 r^2 & s^2 & t^2 \\
\end{vmatrix}
\]

should be determined. For \( t = 1 \) the determinant is zero so it is sufficient to show that the derivative

\[
\begin{vmatrix}
\frac{1}{r} & \frac{1}{s} & \frac{1}{t} \\
1 & 1 & 1 \\
2r & 2s & 2t \\
\end{vmatrix}
\]

is negative for \( r, s, t > 1 \). The derivative is zero for \( s = 1 \) so we take the derivative with respect to \( s \) and get

\[
\begin{vmatrix}
\frac{1}{r} & \frac{1}{s} & \frac{1}{t} \\
1 & 1 & 1 \\
2r & 2s & 2t \\
\end{vmatrix} = \frac{2(t-s)(r-t)(r-s)}{rst} < 0.
\]

Thus probability vectors with four different point probabilities are mapped into interior points of the range. Therefore points on the boundary of the range are images of probability vectors with only 3 different point probabilities, i.e. vectors which are mixtures of just 3 different uniform distributions.

III. THE ORIENTATION AND BOUNDARY OF SIMPLICES

The next step is to determine the relative position of surfaces which are the images of mixtures of 3 different uniform distributions. To do this it is important to be aware of the orientation of the objects under consideration. Let \( U_k \) denote the uniform distribution \( \left( \frac{1}{k}, \frac{1}{k}, \frac{1}{k}, 0, 0, \ldots, 0 \right) \). Let \( U_j, U_k, U_l, U_m \) be uniform distributions such that \( j < k < l < m \). Then the mixtures of the four uniform distributions will form a simplex. This simplex will be oriented such that the sequence \( (U_j, U_k, U_l, U_m) \) is positive, and with this orientation the simplex will be denoted by \( \Delta_{j,k,l,m} \). Then the boundary with orientation is given by

\[
\partial \Delta_{j,k,l,m} = \Delta_{k,l,m} - \Delta_{j,l,m} + \Delta_{j,k,m} - \Delta_{j,k,l},
\]

where \( \Delta_{j,k,l} \) denotes the simplex of mixtures of \( U_j, U_k \) and \( U_l \) such that the orientation of the sequence \( (U_j, U_k, U_l) \) is positive. By the same computation of the functional determinant as in the previous section one sees that \( \Phi \) reverses the orientation.

Let \( \overline{P} \) denote the projection of \( \mathbb{R}^3 \) into the first 2 coordinates. Let \( \overline{\Phi} \) denote the mapping \( \overline{P} \circ \Phi \), i.e. the mapping \( P \rightarrow (H_1(P), H_2(P)) \). According to [6], \( \overline{\Phi} \) is a homeomorphism on each of the simplices \( \Delta_{j,k,l}, \Delta_{j,k,m}, \Delta_{j,l,m}, \Delta_{k,l,m} \).

Now \( \Phi \) maps interior points in \( \Delta_{j,k,l,m} \) into interior points in the range of \( \Phi \). Therefore points on the boundary of the range of \( \Phi \) are elements in the range of the restriction of \( \Phi \) to \( \Delta_{j,k,l} \cup \Delta_{j,k,m} \cup \Delta_{j,l,m} \cup \Delta_{k,l,m} \). For an element in \( \bar{v} \in \mathbb{R}^2 \) there are 3 possibilities:

1) \( \bar{v} \) is not in the range of \( \Phi \).
2) \( \bar{v} \) is on the boundary of \( \left( \Phi \circ \Psi \right)(\Delta) \) and there exists a unique mixture of 2 uniform distributions which is mapped into \( \bar{v} \). Therefore there exists 1 point in the range of \( \Phi \) which projects into \( \bar{v} \).
3) \( \bar{v} \) is an interior point in \( \left( \Phi \circ \Psi \right)(\Delta) \). Then there exists a unique point \( P_1 \) in \( \Delta_{j,k,l,m} \cup \Delta_{j,k,m} \) which maps into \( \bar{v} \) and a unique point \( P_2 \) in \( \Delta_{j,l,m} \cup \Delta_{j,k,l} \) which maps into \( \bar{v} \). The elements in \( \left( \Phi \circ \Psi \right)(\Delta) \) which projects into \( v \) are the elements on the line segment \( \Phi \circ \Psi(P_1) \) to \( \Phi \circ \Psi(P_2) \).

Therefore \( \left( \Phi \circ \Psi \right)(\Delta_{j,k,l} \cup \Delta_{j,l,m} \cup \Delta_{k,l,m}) \) is “above” \( \left( \Phi \circ \Psi \right)(\Delta_{j,k,m} \cup \Delta_{k,l,m}) \) in the sense that the value of \( H_3 \) is higher on the first surface than on the second for a fixed value of \( \bar{v} \).

IV. RANGE OF \( \Phi \)

First the lower bound of \( H_3 \) will be determined for fixed values of \( H_1 \) and \( H_2 \). Assume that \( j < k < l \) and \( j \neq 1 \).
Then the surface \( \Phi (\Delta_{1,j,k}) \cup \Phi (\Delta_{1,k,l}) \) is below the surface \( \Phi (\Delta_{j,k,l}) \). Therefore the surfaces giving the lower bound are of the form \( \Phi (\Delta_{i,j,k}) \). Assume that \( 1 < j \) and \( j + 1 < k \). Then the surface \( \Phi (\Delta_{1,j,j+1}) \cup \Phi (\Delta_{1,j+1,k}) \) is below the surface \( \Phi (\Delta_{1,j,k}) \). Therefore the lower bounding surfaces are of the form \( \Phi (\Delta_{1,j,j+1}) \). For fixed values of \( H_1 \) and \( H_2 \) the number \( j \) is uniquely determined and can be found from the \( H_1/H_2 \) diagram.

Then the upper bound of \( H_3 \) will be determined for fixed values of \( H_1 \) and \( H_2 \). Assume that \( j < k < l \) and \( l \neq n \). Then the surface \( \Phi (\Delta_{j,k,n}) \cup \Phi (\Delta_{k,l,n}) \) is above the surface \( \Phi (\Delta_{j,k,l}) \). Therefore the surfaces giving the upper bound are of the form \( \Phi (\Delta_{j,k,n}) \). Assume that \( j + 1 < k \). Then the surface \( \Phi (\Delta_{j,j+1,k}) \cup \Phi (\Delta_{j+1,k,n}) \) is above the surface \( \Phi (\Delta_{j,k,n}) \). Therefore the upper bounding surfaces are of the form \( \Phi (\Delta_{j,j+1,k}) \). For fixed values of \( H_1 \) and \( H_2 \) the number \( j \) is uniquely determined and can be found from the \( H_1/H_2 \) diagram.

Now the range of \( \Phi \) is described by the surfaces on the boundary, and this result can be used to give exact upper and lower bounds for \( H_1 \) for given values of \( H_2 \) and \( H_3 \). These are given by the same surfaces in the following way:

- For given values of \( H_2 \) and \( H_3 \) the lower bound of \( H_1 \) is assumed on the surface of the form \( \Phi (\Delta_{1,j,j+1}) \).
- For given values of \( H_2 \) and \( H_3 \) the upper bound of \( H_1 \) is assumed on the surface of the form \( \Phi (\Delta_{1,j,j+1,n}) \).

If no bound on \( n \) is given only a lower bound on \( H_1 \) can be given. On an infinite alphabet let \( P_\alpha \) be a probability distribution with point probability \( p_j \) proportional to \( \frac{1}{j^{(\log j)^\alpha}} \), \( \alpha > 1 \). such distributions have finite Rényi entropies \( H_\beta (P_\alpha) \) for \( \alpha > 1 \) and \( \beta > 1 \). The Shannon entropy \( H_1 (P_\alpha) \) is finite if and only if \( \alpha > 2 \). Thus given values of \( H_2 \) and \( H_3 \) will only give an upper bound on \( H_1 \) if \( H_0 \) is upper bounded.

The lower bound on \( H_1 \) depends on \( n \), but for fixed values of \( H_2 \) and \( H_3 \) there exists a unique value of \( j \) such that the joint values of \( H_2 \) and \( H_3 \) are in the range of \( \Delta_{j,j+1,n} \) and this value will depend on \( n \). Actually this \( j \) is increasing as a function of \( n \), but what is more interesting is the inequality

\[
H_3 \geq H_\infty \geq \log j
\]

Therefore \( j \) will eventually be constant and take as value \( \lfloor \exp (H_3) \rfloor \).

V. DISCUSSION

The result can be seen as a generalization of the result in [6], which was used in order to prove this result. Therefore the construction can be iterated to get the joint range of \( H_1, H_2, H_3 \) and \( H_4 \). The essential step in the whole construction is the positivity of the determinant (1). Therefore also Rényi entropies of positive but non-integer order can be taken into consideration.

Acknowledgement I: I thank Karol Zyczkowski for useful discussions and for seeing the paper [9] prior to publication. This paper was the main inspiration for this article. He and Flemming Topsoe have also contributed with several important remarks to this paper.

REFERENCES