Properties of the binary entropy vector region

John MacLaren Walsh and Steven Weber (Drexel University, Dept. of ECE, Philadelphia PA, 19104)

The entropy vector region (EVR) is the set of all entropic vectors for joint distributions with finite support.

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be n discrete random variables with finite support.
- Let $h(\mathbf{X}_\mathcal{A})$ be the entropy of the subset of rvs $\mathbf{X}_\mathcal{A} = (X_i, i \in \mathcal{A})$ for some non-empty subset $\mathcal{A} \subseteq \{1, \ldots, n\} \equiv [n]$.
- Let $\mathbf{h} = (h(\mathbf{X}_\mathcal{A}), \ \mathcal{A} \subseteq [n])$ be the vector of entropies of each non-empty subset $\mathcal{A} \subseteq [n]$. Note $\mathbf{h}$ has $2^n - 1$ entries.
  - Example: for $n = 3$, $\mathbf{h} = (h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123})$.
- A vector $\mathbf{h} \in \mathbb{R}^{2^n-1}$ is called entropic if its elements are the entropies for some joint distribution $\mathbf{p}$ on the n rvs $\mathbf{X}$.
- The entropy vector region (EVR) Γ_n^* is the set of all entropic vectors.
- Normalize by the number of bits for the support m: $\tilde{\mathbf{h}}_n = \mathbf{h}/\log_2 m$, and define Ω_n^* as the set of normalized entropy vectors (Hassibi and Shadbackt 2007).

Importance of EVR: many multi-terminal capacity regions are linear maps from the entropy region.

Theorem 1. Degraded broadcast channel (Cover 1972). *The capacity region of the DBC with $X \to Y_1 \to Y_2$ is the convex hull of the closure of all (R_1, R_2) satisfying*

$$\begin{array}{lllllllll} R_1 & \le & I(X;Y_1|U) & = & -H(U) & + & H(X) & + & H(U,Y_1) - H(X,Y_1) \\ R_2 & \le & I(U;Y_2) & = & H(U) & + & H(Y_2) & - & H(U,Y_2) \end{array}$$

The capacity region is the image under a linear map of the entropy vector region of the rvs (U, X, Y_1, Y_2), i.e., $\mathbf{r} \le \mathbf{M}\mathbf{h}$:

$$\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \le \begin{bmatrix} 1 & 0 & 1 & 0 & -1 & 0 & 0 & \mathbf{0} \\ -1 & 1 & 0 & 1 & 0 & 1 & 0 & \mathbf{0} \end{bmatrix} \begin{bmatrix} h_u \\ h_x \\ h_{y_2} \\ h_{u,y_1} \\ h_{u,y_2} \\ h_{x,y_1} \\ \mathbf{h}_{\text{other}} \end{bmatrix}$$

Shannon inequalities give a polymatroid (polyhedral) outer bound on the EVR (Yeung 1997).

Shannon inequalities (non-negativity of conditional mutual information) are equivalent to the polymatroidal axioms, i.e., $h(\emptyset) = 0$ and for all $\mathcal{A}, \mathcal{B} \subseteq [n]$:

$$\begin{array}{rcl} \mathcal{A} \subseteq \mathcal{B} & \Rightarrow & h(\mathcal{A}) \le h(\mathcal{B}), \\ h(\mathcal{A}) + h(\mathcal{B}) & \ge & h(\mathcal{A} \cup \mathcal{B}) + h(\mathcal{A} \cap \mathcal{B}). \end{array}$$

Define the polyhedron of points in entropy space satisfying the Shannon inequalities

$$\Gamma_n = \left\{ \mathbf{h} \in \mathbb{R}^{2^n-1} : \begin{array}{lll} \mathcal{A} \subseteq \mathcal{B} & \Rightarrow & h(\mathcal{A}) \le h(\mathcal{B}), \\ h(\mathcal{A}) + h(\mathcal{B}) & \ge & h(\mathcal{A} \cup \mathcal{B}) + h(\mathcal{A} \cap \mathcal{B}) \end{array}, \ \mathcal{A}, \mathcal{B} \subseteq [n] \right\}.$$

The Shannon inequalities provide an outer bound on the EVR: $\Gamma_n^* \subseteq \Gamma_n$.

Just as Ω_n^* is Γ_n^* normalized by cardinality, we can define the normalized Shannon polytope outer bound Ω_n s.t. $\Omega_n^* \subseteq \Omega_n$.

Properties of the EVR Γ_n^*: insufficiency of Shannon type inequalities and curvature for $n \ge 4$

- The closure of the EVR, $\bar{\Gamma}_n^*$ is a convex cone.
- For $n = 2, 3$ it is known that the Shannon inequalities are tight

$$\bar{\Gamma}_2^* = \Gamma_2, \quad \bar{\Gamma}_3^* = \Gamma_3.$$

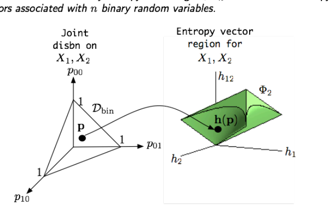
- For $n \ge 4$ it is known that the Shannon inequalities are not tight:
  - **Theorem** (Zhang and Yeung 1997): $\bar{\Gamma}_4^* \ne \Gamma_4$. In particular, for any four rvs $(X_1, \ldots, X_4)$ the following is a non-Shannon type inequality:

$$2I(X_3;X_4) \le I(X_1;X_2) + I(X_1;X_3,X_4) + 3I(X_3;X_4|X_1) + I(X_3;X_4|X_2).$$

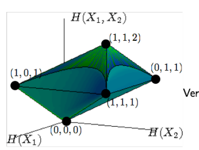
  - **Theorem** (Matúš 2007): $\bar{\Gamma}_n^*$ is not polyhedral for $n \ge 4$ (it is curved).

Our focus: restrict attention to binary rvs and the associated binary entropy vector region

Definition 1. *The binary entropy vector region, Φ_n is the set of all entropy vectors associated with n binary random variables.*



The Shannon outerbound is tight for Φ_n for $n = 2, 3$ (but not for $n \ge 4$)



Halfspace rep. of $\text{conv}(\Phi_2^*) = \bar{\Omega}_2^*$:

$$\begin{array}{lcl} h_1 & \le & 1 \\ h_2 & \le & 1 \\ h_1 & \le & h_{12} \\ h_2 & \le & h_{12} \\ h_{12} & \le & h_1 + h_2 \end{array}$$

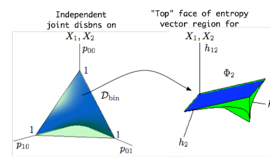
Vertex rep. of $\text{conv}(\Phi_2^*) = \bar{\Omega}_2^*$:

$$\begin{array}{ll} (0,0,0) & X_1 = X_2 = c \\ (0,1,1) & X_1 = c, \ X_2 \sim U \\ (1,0,1) & X_1 \sim U, \ X_2 = c \\ (1,1,1) & X_1 = X_2 = U \\ (1,1,2) & X_1 = U \perp X_2 = U \end{array}$$

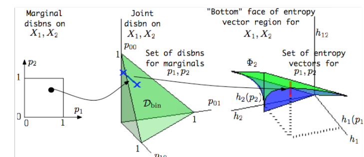
The independent distributions map to the "top face" of the binary entropy region.

Independent distributions on two bits are the subset in distribution space satisfying:

$$\mathcal{D}_{\text{bin}}^{\text{ind}} = \{\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11}) : \mathbf{p} = (p_1 p_2, p_1 \bar{p}_2, \bar{p}_1 p_2, \bar{p}_1 \bar{p}_2)\}, \ 0 \le p_1, p_2 \le 1.$$

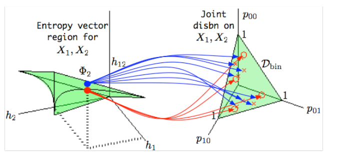


Fixed marginal distributions map to a "vertical" line of feasible joint entropies



Fixed marginals $p_1, p_2 \in [0,1]^2$ map to a line in $\mathcal{D}_{\text{bin}}$ which "punctures" the simplex at two points. One of these two puncture points is the joint distribution with minimum joint entropy for fixed marginal entropies h_1, h_2.

The inverse map from entropy space to distribution space is one to (finitely) many



For fixed marginal entropies h_1, h_2, the inverse mapping from $\mathbf{h}$ to $\mathbf{p}$ is one to finitely many. There is a critical h_{12} for each (h_1, h_2) such that

$$|\{\mathbf{p} \in \mathcal{D}_{\text{bin}} : \mathbf{h}(\mathbf{p}) = \mathbf{h}\}| = \begin{cases} 4, & h_{12} < \hat{h}_{12}(h_1, h_2) \\ 8, & h_{12} \ge \hat{h}_{12}(h_1, h_2) \end{cases}$$

Main result: a characterization of Φ_n for arbitrary n

Theorem 2. Walsh and Weber (ITW 2009, submitted). *For any $\mathbf{h} \in \mathbb{R}^{2^n-1}$, there is an algorithm that determines whether or not $\mathbf{h} \in \Phi_n$. The algorithm terminates in a finite number of steps. Further, if $\mathbf{h} \in \Phi_n$ then the algorithm returns the (finite) set of joint distributions with entropy $\mathbf{h}$.*

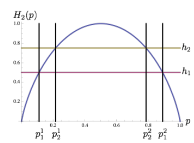
Main idea: Given a collection of $k-1$-way marginal distributions, the k-way distributions having these marginals have a single free parameter. The k-way entropies can each be written as a function of these free parameters. The free parameters can then be solved for by line search. Continue from $k = 1$ through $k = n$ to find all possible feasible joint distributions (if any).

Example: Next three slides present intuition for $n = 2$.

Intuition for inverse map for $n = 2$. Marginal entropies have unique marginal prob. up to two-fold ambiguity.

Given $\mathbf{h} = (h_1, h_2, h_{12})$ we wish to determine whether or not $\mathbf{h} \in \Phi_2$. But h_1, h_2 dictate the marginal probabilities up to a two-fold ambiguity.

$$\begin{array}{lcl} h_1 & \Rightarrow & p_1 \equiv p_{00} + p_{01} \in \{p_1^1(h_1), p_1^2(h_1)\} \\ h_2 & \Rightarrow & p_2 \equiv p_{00} + p_{10} \in \{p_2^1(h_2), p_2^2(h_2)\}. \end{array}$$



Intuition for inverse map for $n = 2$. Joint entropy can be expressed in terms of the marginal probabilities

Express elements of joint distribution in terms of marginal distribution:

$$\begin{array}{lclclcl} p_1 & = & p_{00} + p_{01} & \Rightarrow & p_{01} & = & p_1 - p_{00} \\ p_2 & = & p_{00} + p_{10} & \Rightarrow & p_{10} & = & p_2 - p_{00} \\ 1 & = & p_{00} + p_{01} + p_{10} + p_{11} & \Rightarrow & p_{11} & = & 1 + p_{00} - p_1 - p_2 \end{array}$$

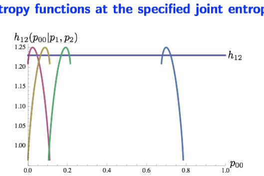
Substitute into expression for joint entropy

$$\begin{array}{rcl} h_{12} & = & -p_{00} \log p_{00} - p_{01} \log p_{01} - p_{10} \log p_{10} - p_{11} \log p_{11} \\ & = & -p_{00} \log p_{00} - (p_1 - p_{00}) \log(p_1 - p_{00}) - (p_2 - p_{00}) \log(p_2 - p_{00}) - \\ & & (1 + p_{00} - p_1 - p_2) \log(1 + p_{00} - p_1 - p_2) \end{array}$$

Define four joint entropy functions, one for each possible pair of marginals:

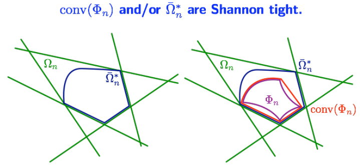
$$h_{12}(p_{00}|p_1^1, p_2^1), \quad h_{12}(p_{00}|p_1^1, p_2^2), \quad h_{12}(p_{00}|p_1^2, p_2^1), \quad h_{12}(p_{00}|p_1^2, p_2^2).$$

Intuition for inverse map for $n = 2$. Invert the four joint entropy functions at the specified joint entropy



If $h_{12} < \min_{p_{00}} h_{12}(p_{00}|p_1, p_2)$ or $h_{12} > \max_{p_{00}} h_{12}(p_{00}|p_1, p_2)$ then those marginals cannot support the required joint entropy. If this is true for all four pairs of marginals then $\mathbf{h} \notin \Phi_2$. Else, return the valid joint distributions.

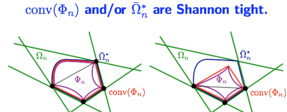
Open question: characterization of which faces of $\text{conv}(\Phi_n)$ and/or $\bar{\Omega}_n^*$ are Shannon tight.



Γ_n is a polytope while $\bar{\Omega}_n^* \subset \Gamma_n$ is non-polytopic (but convex) subset that is "tight" on some faces of Γ_n but "loose" on others.

It might be that $\text{conv}(\Phi_n) \ne \bar{\Omega}_n^*$ and that $\text{conv}(\Phi_n)$ is not a polytope (as pictured on the right).

Open question: characterization of which faces of $\text{conv}(\Phi_n)$ and/or $\bar{\Omega}_n^*$ are Shannon tight.



It might be the case that $\text{conv}(\Phi_n) = \bar{\Omega}_n^*$ (left), or that $\text{conv}(\Phi_n)$ is a polytope (right).

Regardless, it is straightforward to find the vertices of Ω_n, and the subset of those vertices in Φ_n (black circles). The polytope formed by those vertices gives an inner bound to $\bar{\Omega}_n^*$ (pictured in light black lines).

This inner bound is further going to be tight on faces of $\bar{\Omega}_n^*$ that are tight with the outer bound Ω_n.

References

R. Yeung, "A framework for linear information inequalities," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1924–1934, November 1997.

Z. Zhang and R. Yeung, "A non–Shannon–type conditional inequality of information quantities," *IEEE Trans actions on Information Theory*, vol. 43, no. 6, pp. 1982–1986, November 1997.

F. Matúš, "Infinitely many information inequalities," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 41–44.

B. Hassibi and S. Shadbakht, "Normalized entropy vectors, network information theory and convex optimization," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, Bergen, Norway, July 2007.

T. Chan and A. Grant, "Entropy vectors and network codes," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 1586–1590.

J.M. Walsh and S. Weber "Some properties of the binary entropy function", submitted to the *IEEE Information Theory Workshop (ITW)*, Volos, Greece, June, 2009.