Piecewise linear conditional information inequality

František Matúš

***Abstract*— A new information inequality of non-Shannon type is proved for three discrete random variables under conditional independence constraints, using framework of entropy functions and polymatroids. Tightness of the inequality is described via quasigroups.**

***Index Terms*— Conditional independence, entropy function, information inequality, polymatroid, quasigroup, Shannon entropy.**

I. Introduction

THE *entropy function* of a random vector $(\xi_i)_{i\in N}$ indexed by a finite set N assigns to each subset I of N the Shannon entropy of the subvector $(\xi_i)_{i\in I}$. The notion depends on the base of logarithms occurring in the entropy; our preference is given to the natural ones.

By basic properties of the Shannon entropy, an entropy function g satisfies $g(\emptyset)=0$, is nondecreasing, thus $g(I)\leqslant g(J)$ for $I\subseteq J$, and submodular, $g(I)+g(J)\geqslant g(I\cup J)+g(I\cap J)$ for $I,J\subseteq N$. A pair (N,g) where g is a real function on the power set $\mathcal{P}(N)$ of N that satisfies these three requirements is called a *polymatroid* with the ground set N and rank function g [3], [2], [8], making usually no difference between a polymatroid and its rank function. Thus, real-valued entropy functions are polymatroids [1]. Let us call a polymatroid *entropic* if it is equal to the entropy function of a random vector taking a finite number of values.

Theorem 1 below solves the problem which of the polymatroids ag_1+bg_2, $a,b\geqslant 0$, is entropic for two very special polymatroids g_1, g_2 with a common three-element ground set.

A real function ϕ on $\mathbb{R}^{\mathcal{P}(N)}$, the Euclidean space containing entropy functions, generates an *information inequality* if $\phi(g)\geqslant 0$ holds for the entropic polymatroids. If the nonnegativity takes place even for all polymatroids then the inequality is of *Shannon type*. First non-Shannon type information inequalities appeared in [12], see also [10], [7]. If $\phi(g)\geqslant 0$ holds only for the entropic polymatroids satisfying additional constraints, usually on conditional independences among subvectors, then the inequality is *conditional*. For non-Shannon type conditional information inequalities see [11], [5].

One implication of Theorem 1 interprets as a non-Shannon type, conditional, piecewise linear information inequality, see Remarks 1 and 2. Tightness is discussed in Remark 3.

Information inequalities, entropy functions of random vectors and related problems are summarized on introductory level in [9, Chapters 12-14].

Manuscript submitted July 5, 2005; revised September 22, 2005. This work was supported by Grant Agency of Academy of Sciences of the Czech Republic under Grant A1075104, and by Grant Agency of the Czech Republic under Grant 201/01/1482. F. Matúš is with Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic (e-mail: matus@utia.cas.cz).

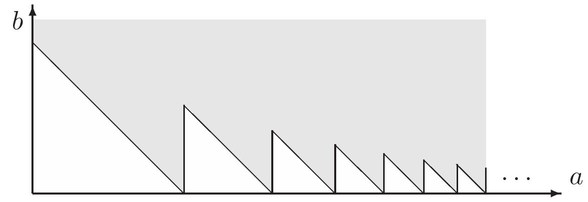

Fig. 1. The region where $a+b\geqslant \ln\lceil e^a\rceil$.

II. The main result

Theorem 1: If $N=\{1,2,3\}$, $J=\{2,3\}$ and $a,b\geqslant 0$ then the polymatroid given by

$$g_{a,b}(I)=a\min\{2,|I|\}+b\min\{1,|I\cap J|\}\,,\quad I\subseteq N\,,$$

is entropic if and only if $a+b\geqslant \ln\lceil e^a\rceil$.

Remark 1: Denoting the difference of $g(I)+g(J)$ and $g(I\cup J)+g(I\cap J)$ by $\Delta_{I,J}\,g$, it is not difficult to prove that a polymatroid g equals $g_{a,b}$ for some $a,b\geqslant 0$ if and only if $\Delta_{N\setminus J,\{j\}}\,g=0$ for $j\in J$, and $g(N)=g(N\setminus\{i\})$ for $i\in N$. If g is equal to the entropy function of a random vector then each of the five linear constraints is equivalent to a conditional independence among subvectors of the vector. Hence, one implication of Theorem 1 is the conditional information inequality

$$g(\{2\})\geqslant \ln\lceil e^{g(\{1\})}\rceil\,.$$

Remark 2: The region of pairs $(a,b)\in\mathbb{R}^2$ for which the inequality holds is outlined in Fig. 1. It consists of the half-line given by $a=0$ and $b\geqslant 0$, and the strips given by

$$\ln(n-1)<a\leqslant \ln n \;\text{ and }\; a+b\geqslant \ln n$$

for $n\geqslant 2$. Thus, the inequality of Theorem 1 is piecewise linear.

Proof: Suppose $g_{a,b}$ is the entropy function of a random vector (ξ_1,ξ_2,ξ_3) taking a finite number of values. This implies that ξ_1 is independent of ξ_2, ξ_1 is independent of ξ_3, and each individual variable is a function of the remaining two ones. There is no loss of generality in assuming that ξ_i ranges in a set X_i such that it attains each x_i in X_i with a positive probability, $i\in N$; this is referred to as the positivity assumption. Consider the bipartite graph between X_2 and X_3 that has for its edges those pairs (x_2,x_3) which (ξ_2,ξ_3) attains with positive probabilities. Since ξ_1 is a function of this subvector each edge (x_2,x_3) can be colored by a unique element x_1 of X_1 such that (x_1,x_2,x_3) is attained with a positive probability. The independence of ξ_1 and ξ_2 implies that each x_2 is incident to m edges where m is the cardinality of X_1, using the positivity assumptions for X_1 and X_2. Since ξ_3 is a function of (ξ_1,ξ_2) two edges incident to any x_2 get different colors. Thus, each x_2 is incident to m edges of m

colors. The same assertion is valid for each x_3, by symmetry. It follows that if C is a connected component of the bipartite graph then C has the same number of vertices in X_2 and in X_3, say n_C. Since each vertex of the graph is incident to m others $n_C \geqslant m$.

If an edge (x_2, x_3) is colored by x_1 then (x_1, x_2) is attained with the same positive probability as (x_1, x_3) because ξ_3 is a function of (ξ_1, ξ_2) and ξ_2 is a function of (ξ_1, ξ_3). Then, the two pairwise independences imply that x_2 is attained with the same probability as x_3. This probability does not change when x_2 and x_3 range in a connected component C of the graph and can be denoted by q_C. Summing over the components, $\Sigma\, n_C q_C = 1$. This implies that the Shannon entropy $-\Sigma n_C q_C \ln q_C$ of ξ_2 is lower bounded by $\Sigma n_C q_C \ln n_C$. Since the entropy equals $g_{a,b}(\{2\}) = a + b$ and $n_C \geqslant m$ it follows that $a + b \geqslant \ln m$. On the other hand, $\ln m$ is a trivial upper bound on the entropy of ξ_1, equal to $g_{a,b}(\{1\}) = a$. Hence, $e^{a+b} \geqslant m \geqslant e^a$, and thus $a + b \geqslant \ln\lceil e^a \rceil$ holds.

To prove that $g_{a,b}$ is entropic if the inequality holds recall the simple fact that the sum of entropic polymatroids is entropic. Since $g_{0,b}$, $b \geqslant 0$, are obviously entropic it suffices to prove that $g_{a,\ln n - a}$ is entropic if $n \geqslant 2$ and $0 \leqslant a \leqslant \ln n$. To this end, consider any binary quasigroup operation $\circ$ on a set of n elements and three random variables ranging in the set such that the entropy of ξ_1 equals a, ξ_2 is uniform and independent of ξ_1, and $\xi_3 = \xi_1 \circ \xi_2$. ∎

Remark 3: Suppose $g_{a,b}$ is the entropy function of a random vector as in the above proof and $a + b$ equals $\ln\lceil e^a \rceil$. It is not difficult to see that the bipartite graph is connected and $e^{a+b} = n = m = \lceil e^a \rceil$ where n is the cardinality of X_2, equal to the cardinality of X_3. This implies $b = \ln n - a$ and $a > \ln(n-1)$ provided $a \neq 0$. Since all random variables of the vector take n values they can be assumed to range in a common set of n elements. Hence, $\xi_3 = \xi_1 \circ \xi_2$ for an operation $\circ$ on the set. It follows from the properties of the graph that $\circ$ is a quasigroup. This explains tightness in the information inequality of Theorem 1 and the use of quasigroups in the second part of its proof.

Remark 4: Recalling that a matroid on a set N is an integer-valued polymatroid bounded by the cardinality function defined on $\mathcal{P}(N)$, $g_{a,0}$ is a multiple of the matroid $g_{1,0}$. By Theorem 1, the polymatroid $g_{a,0}$ is entropic if and only if a equals $\ln n$ for some integer $n \geqslant 1$, as mentioned in [9, Figure 14.2]. It can be shown that the necessity of this condition holds for positive multiples of the rank functions of nontrivial matroids [4, Theorem, p.190], even more, if such a multiple equals the entropy function of a random vector then all subvectors have uniform distributions, with functional dependences governed by quasigroups of higher arities. More on this and related problems of matroid representation theory, and relations to ideal secret sharing schemes, can be found in [6].

Remark 5: Since the polymatroids on a set N of cardinality three form a polyhedral cone and the constraints of Remark 1 are equalities arising from defining inequalities of the cone the set of polymatroids $g_{a,b}$, $a, b \geqslant 0$, is a two-dimensional face of the cone. Theorem 1 solves which polymatroid of this face is entropic. For more dimensional faces of the cone this problem is far from being understood.

Remark 6: It is not difficult to prove that each polymatroid with a three-element ground set is a limit of the entropic polymatroids [12, Theorem 1], and thus so are the polymatroids $g_{a,b}$ of Theorem 1.

References

[1] S. Fujishige (1978) Polymatroidal dependence structure of a set of random variables. *Inform. and Control* **39** 55–72.

[2] S. Fujishige (1991) *Submodular functions and optimization.* North-Holland, Amsterdam.

[3] L. Lovász (1982) Submodular functions and convexity. *Mathematical Programming – The State of the Art* (A. Bachem, M. Grötchel and B. Korte, eds.), Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 235–257.

[4] F. Matúš (1994) Probabilistic conditional independence structures and matroid theory: background. *Int. J. of General Systems* **22** 185–196.

[5] F. Matúš (1999) Conditional independences among four random variables III: final conclusion. *Combinatorics, Probab. and Comput.* **8** 269–276.

[6] F. Matúš (1999) Matroid representations by partitions. *Discrete Mathematics* **203** 169–194.

[7] F. Matúš (2005) Inequalities for Shannon entropies and adhesivity of polymatroids. *Proceedings of the 9-th Canadian Workshop on Information Theory*, McGill University, Montréal, Canada, 28–31.

[8] H. Narayan (1997) *Submodular Functions and Electrical Networks.* Elsevier, Amsterdam.

[9] R.W. Yeung (2002) *A First Course in Information Theory*. Kluwer Academic/Plenum Publishers, New York.

[10] Z. Zhang (2003) On a new non-Shannon type information inequality. *Communications in Information and Systems* **3** 47–60.

[11] Z. Zhang and R.W. Yeung (1997) A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Information Theory* **43** 1982–1986.

[12] Z. Zhang and R.W. Yeung (1998) On characterization of entropy function via information inequalities. *IEEE Trans. Information Theory* **44** 1440–1452.