Topics in Database Theory – Homework 3

1 Proof of the Shannon Inequalities

1. (0 points)

   The entropy of a random variable X with N outcomes and probabilities $p_1, \ldots, p_N$ is defined as the following quantity:

   $$h(X) \stackrel{\text{def}}{=} -\sum_{i=1,N} p_i \log p_i \tag{1}$$

   As usual, this definition extends to joint random variables. For example, if X, Y are joint random variables, and p_{ij} is the probability of the outcome $X = i \wedge Y = j$, then the entropies $h(XY)$, $h(X)$ are derived from (1) as:

   $$h(XY) = -\sum_{i,j} p_{ij} \log p_{ij}$$
   $$h(X) = -\sum_{i} p_{i*} \log p_{i*} \qquad \text{where } p_{i*} = \sum_j p_{ij}$$

   Prove the statements below by using directly the definition (1). Recall that any concave function $\varphi(x)$ satisfies *Jensen's inequality*: for all $x_1, \ldots, x_n$ in the domain of φ and for all $p_i \geq 0$ s.t. $\sum_i p_i = 1$:

   $$\sum_i p_i \cdot \varphi(x_i) \leq \varphi\left(\sum_i p_i \cdot x_i\right)$$

   If φ is not a linear function, then equality holds iff all values x_i corresponding to non-zero p_i's are equal, i.e. $p_i, p_j > 0$ implies $x_i = x_j$.

   (a) $h(X) \geq 0$

   (b) If X has N outcomes, then $h(X) \leq \log N$. **Hint:** $\varphi(x) = \log x$ is concave.

   (c) Define $h(Y|X) \stackrel{\text{def}}{=} h(XY) - h(X)$. Prove that $h(Y|X) = \mathbb{E}_x[h(Y|X = x)]$, where $h(Y|X = x)$ is the entropy of the random variable Y conditioned on the outcome $X = x$. **Hint:** direct calculation.

   (d) $h(XY) \geq h(X)$. **Hint:** the solution has one line.

(e) Prove that $h(X) + h(Y) \geq h(XY)$. Moreover, equality holds iff X, Y are independent random variables (denoted as $X \perp Y$), which means that $p(X = x \wedge Y = y) = p(X = x) \cdot p(Y = y)$ for all outcomes (x, y). **Hint**: show that $h(X)+h(Y)-h(XY) \geq 0$ by using Jensen's inequality for $\log x$.

(f) Prove submodularity: $h(XZ) + h(YZ) \geq h(XYZ) + h(Z)$. **Hint**: 1-2 lines.

2 Polymatroids and Modular Functions

2. (0 points)

A function $h : 2^{[n]} \to \mathbb{R}_+$ (or, equivalently, a vector $h \in \mathbb{R}_+^{2^{[n]}}$) is called a *polymatroid* if it satisfies Shannon's basic inequalities:

$$h(\emptyset) = 0 \tag{3}$$

$$h(\boldsymbol{XY}) \geq h(\boldsymbol{X}) \qquad \text{here } \boldsymbol{XY} \text{ denotes } \boldsymbol{X} \cup \boldsymbol{Y} \tag{4}$$

$$h(\boldsymbol{X}) + h(\boldsymbol{Y}) \geq h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y}) \tag{5}$$

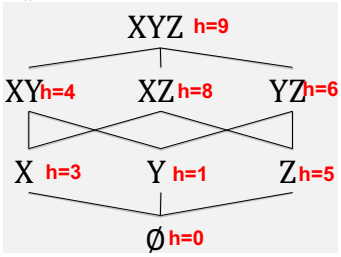
(Note that every entropic vector h is a polymatroid, but the converse is in general not true: this is a break-through result from [1].)

The function is called *modular* if $h(\boldsymbol{X}) + h(\boldsymbol{Y}) = h(\boldsymbol{XY})$ whenever $\boldsymbol{X} \cap \boldsymbol{Y} = \emptyset$. Equivalently, a modular function is uniquely defined by its values on single variables, $h(X_i)$, such that:

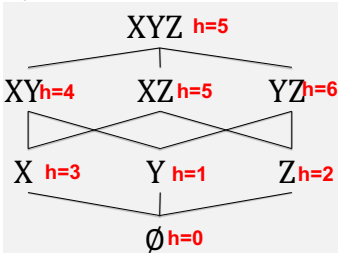
$$h(X_{i_1} X_{i_2} \cdots X_{i_k}) = h(X_{i_1}) + \cdots + h(X_{i_k}) \tag{6}$$

(a) For each function below indicate whether it is modular, or a polymatroid, or neither.

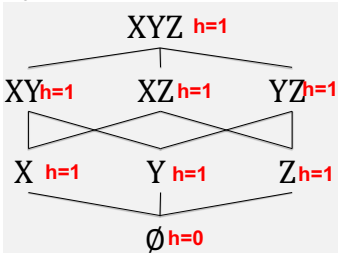
h_a :



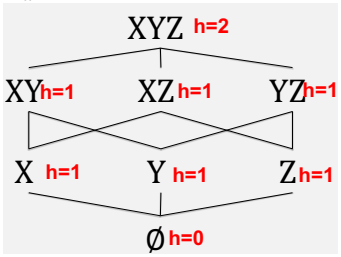
h_b :



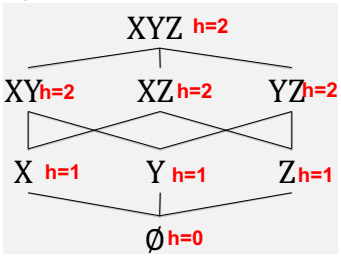
h_c :



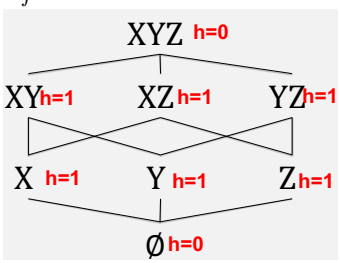
h_d :



h_e :



h_f :



(b) Prove that every modular function h is a polymatroid. (Note that $h \geq 0$ by definition).

(c) Let h be any polymatroid. Fix an order of the n variables, e.g. $X_1, X_2, \ldots, X_n$. If $i \leq n$ then $\boldsymbol{X}_{[i]}$ denotes the set of variables $X_1 X_2 \ldots X_i$. Define the following function $h_0 : 2^{[n]} \to \mathbb{R}_+$, which we call the *modularization of* h:

$$h_0(X_{i_1} \cdots X_{i_k}) \stackrel{\text{def}}{=} \sum_{j=1,k} h(X_{i_j} | X_{[i_j - 1]})$$

(For example, $h_0(X_2 X_5) = h(X_2|X_1) + h(X_5|X_1 X_2 X_3 X_4)$.) Prove the following items (which say that h_0 is a modular function, no greater than h, and equal to h on the set of all variables $\boldsymbol{X}_{[n]}$):

i. h_0 is modular.

ii. For every set $\boldsymbol{U}$, $h_0(\boldsymbol{U}) \leq h(\boldsymbol{U})$. **Hint**: use the chain rule, which says $h(XYZU) = h(X) + h(Y|X) + h(Z|XY) + h(U|XYZ)$. Also use submodularity: $h(Z|X) \geq h(Z|XY)$.

iii. $h_0(\boldsymbol{X}_{[n]}) = h(\boldsymbol{X}_{[n]})$, where $X_{[n]}$ is the set of all n variables.

(d) For each $i = 1, n$, define the following function $h^{(X_i)} : 2^{[n]} \to \mathbb{R}_+$:

$$h^{(X_i)}(\boldsymbol{U}) = \begin{cases} 1 & \text{if } X_i \in \boldsymbol{U} \\ 0 & \text{otherwise} \end{cases}$$

i. Prove that $h^{(X_i)}$ is a modular function. We call it a *basic modular function.*

ii. Prove that every modular function h is a positive linear combination of the n basic modular functions, i.e. there exists $a_1, \ldots, a_n \geq 0$ such that $h = \sum_i a_i h^{(X_i)}$. For example, the function h_a in Question (a) above can be written as $h_a = 3h^{(X)} + 1h^{(Y)} + 5h^{(Z)}$. (Yes, h_a is modular!)

(e) Consider an inequality of the following form:

$$\sum_{j=1,m} w_j h(\boldsymbol{U}_j) \geq h(\boldsymbol{X}) \tag{7}$$

where $\boldsymbol{X}$ is the set of all variables (same as $\boldsymbol{X}_{[n]}$), and for each j, $w_j \geq 0$ and $\boldsymbol{U}_j \subseteq \boldsymbol{X}$. Prove that the following are equivalent:

1. The coefficients $w_1, \ldots, w_m$ form a fractional edge cover of the hypergraph with hyperedges $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$. In other words, for all $i = 1, n$, $\sum_{j : X_i \in \boldsymbol{U}_j} w_j \geq 1$.
2. Inequality (7) holds for all modular functions.
3. Inequality (7) holds for all polymatroids.

The inequality (7) corresponds to an AGM bound. What the problem above asks you to do is to show that in order to prove the AGM bound it suffices to check the inequality only on modular functions. This explains why the AGM bound is easier to compute than the general case.

Hint: Prove Item 1 $\Rightarrow$ Item 2 $\Rightarrow$ Item 3 $\Rightarrow$ Item 1 (three proofs). For the first proof, show Inequality (7) holds for every basic modular function $h^{(X_i)}$ (then it holds for

any modular function; why?); for the second step, use the modularization function; for the third step, use the basic modular functions $h^{(X_i)}$ again.

Note: in class we proved that Item 1 implies Item 3. Your solution to this question provides an alternative, arguably simpler proof, by relying on modular functions.

References

[1] Z. Zhang and R. W. Yeung. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory*, 44(4):1440–1452, 1998.