CS 229r Information Theory in Computer Science	Feb 9, 2016
Lecture 5	
Lecturer: Vasileios Nakos and Manolis Zampetakis	*Scribe: Colin Lu*

1 Subadditivity of Entropy

Subadditivity of entropy is a simple but very useful result. It states that for a random vector $(X_1, ..., X_n)$,

$$H(X_1, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$$

Proof: By definition,

$$H(X_1, ..., X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + ... + H(X_n|X_1, ..., X_{n-1})$$

Term-wise, $H(X_i|X_{i-1}, ..., X_1) \leq H(X_i)$, so

$$H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + ... + H(X_n|X_1, ..., X_{n-1}) \leq \sum_{i=1}^{n} H(X_i)$$

giving the desired result. ■

1.1 Review of Shearer's Lemma

Shearer's Lemma may be stated as follows: Let $\mathcal{F}$ be a family of subsets of $[n]$, such that each $i \in [n]$ is included in at least t members of $\mathcal{F}$. Then for random variables $X_1, ..., X_n$,

$$H(X_1, ..., X_n) \leq \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F)$$

where $X_F = (X_{i_1}, ..., X_{i_{|F|}})$ where $F = (i_1, ..., i_{|F|})$, such that $i_1 < i_2 < ... < i_{|F|}$.

Proof of Shearer's lemma: By definition, $H(X_F) = H(X_{i_1}, ..., X_{i_{|F|}}) = H(X_{i_1}) + H(X_{i_2}|X_{i_1}) + ... + X(X_{i_{|F|}}|X_{i_1}, ..., X_{i_{|F|-1}})$

Considering this expression term-by-term, we get the inequalities:

$$H(X_{i_j}|X_{i_{j-1}}, ..., X_{i_1}) \geq H(X_{i_j}|X_{i_j-1}, ..., X_1)$$

since the right hand side simply conditions on more information.

Then by summing, we get that

$$\sum_{F \in \mathcal{F}} H(X_F) \geq \sum_{i=1}^{n} tH(X_i|X_{i-1}, ..., X_1) \geq tH(X)$$

since each term $H(X_i|X_{i-1}, ..., X_1)$ appears at least t times over $F \in \mathcal{F}$. ■

2 Sums of binomial coefficients

Subadditivity of entropy can be applied to prove inequalities in combinatorial contexts.

Recall Stirling's approximation: $n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$.

Applying this to a binomial coefficient of the form $\binom{n}{\alpha n}$ with $\alpha \in [0, 1]$ gives

$$\binom{n}{\alpha n} = \frac{n!}{(\alpha n)!(n - \alpha n)!} \approx \frac{2^{H(\alpha)n}}{\sqrt{2\pi n\alpha(1 - \alpha)}}$$

where $H(\alpha)$ represents the entropy of a bernoulli random variable with probability of success α, satisfying $H(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha)$. This suggests a connection between entropy and binomial coefficients.

Theorem 1: When $\alpha \leq \frac{1}{2}$, $\sum_{i \leq \alpha n} \binom{n}{i} \leq 2^{H(\alpha)n}$

This can be proved using the subadditivity of entropy:

Consider a collection of sets $\mathcal{C} = \{C : C \subset [n], |C| \leq \alpha n\}$.

Let X be a random variable, chosen uniformly at random from $\mathcal{C}$. Then $H(X) = \log |\mathcal{C}| = \log \left(\sum_{i \leq \alpha n} \binom{n}{i} \right)$, so it is sufficient to show that $H(X) \leq H(\alpha)n$.

Now suppose $X = \{X_1, ..., X_n\}$ with $X_i = 1$ if $i \in X$ and 0 otherwise. Then $H(X) = H(X_1, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$ by subadditivity of entropy. Then it is sufficient to show that $H(X_i) \leq H(\alpha)$, since all of the X_i are symmetric.

$$P(i \in X) = \sum_{l=0}^{\alpha n} P(i \in X || X| = l) P(|X| = l) \leq \alpha \sum_{l=0}^{\alpha n} P(|X| = l) = \alpha$$

since for all l, $P(i \in X||X| = l) = \frac{l}{n} \leq \frac{\alpha n}{n} = \alpha$.

In the case that $|X| = \alpha n$, $P(i \in X) = \alpha$, and otherwise $P(i \in X) < \alpha$, so in general $P(i \in X) \leq \alpha$. Since $\alpha \leq \frac{1}{2}$, this gives us that $H(X_i) \leq H(\alpha)$, as desired. ■

An example application of this result:

Theorem 2: For $X \sim Binom\left(n, \frac{1}{2}\right)$, $\sigma = \frac{\sqrt{n}}{2}$

$$P\left(|X - \frac{n}{2}| \geq c\sigma\right) \leq 2^{1 - \frac{c^2}{2}}$$

$\forall c \geq 0$.

Proof: By symmetry, $P\left(|X - \frac{n}{2}| \geq c\sigma\right) = 2P\left(X \leq \frac{n}{2} - c\sigma\right)$. Clearly, $P(X = i) = \binom{n}{i} 2^{-n}$, so we have that $\frac{1}{2} P(|X - \frac{n}{2}| \geq c\sigma) = \sum_{i=0}^{\frac{n}{2} - c\sigma} \binom{n}{i} 2^{-n}$. Applying theorem 1 gives that $\sum_{i=0}^{\frac{n}{2} - c\sigma} \binom{n}{i} 2^{-n} \leq 2^{H(\frac{1}{2} - \frac{c\sigma}{n})}$. Then we get the theorem by noting that $H(\frac{1}{2} - \epsilon) \leq 1 - \frac{1}{2} \log(1 - \epsilon^2)$ to get that $2^{H(\frac{1}{2} - \frac{c\sigma}{n})} \leq 2^{\frac{-c^2}{2}}$.

3 The coin-weighing problem

Suppose there are n coins, indexing by $C = [n]$. There is a subset $B \subseteq C$ of "fake" coins with known different weights (all of the fake coins have the same weight, and all of the real coins have the same weight). We proceed by selecting subsets $D_i \subseteq C$ and weighing them, which tells us the number of fake coins in D_i, $|D_i \cap B|$. We wish to determine the identity of the fake coins, with the subsets D_i all selected in advance, before any weighings have been done. Clearly we can choose the n singleton sets $D_i = \{i\}$, and this would be enough to determine B, but we can make tighter upper and lower bounds.

Note that determining B is equivalent to selecting enough subsets D_i such that for any subsets $B, B' \subseteq C$, there exists an i such that $|D_i \cap B| \neq |D_i \cap B'|$. Then we can represent the information in B by $(|D_1 \cap B|, |D_2 \cap B|, ..., |D_l \cap B|)$ since these values uniquely determine B.

Denote the minimal size of l by $f(n)$. It can be shown by a combinatorial argument that

$$f(n) \leq \frac{2n}{\log n} \left(1 + O\left(\frac{\log \log n}{\log n}\right)\right)$$

See [2] for a proof of this upper bound.

We will use an information theoretic argument to show that

$$f(n) \geq \frac{2n}{\log n} \left(1 + \Omega\left(\frac{1}{\log n}\right)\right)$$

Suppose that B is picked uniformly at random from C (each element has a $\frac{1}{2}$ chance of inclusion). By the subadditivity of entropy, we have that

$$n = H(B) = H\left(|D_1 \cap B|, |D_2 \cap B|, ..., |D_l \cap B|\right) \leq \sum_{i=1}^{l} H\left(|D_i \cap B|\right) \leq \sum_{i=1}^{l} \log(n+1) \leq l \log(n+1) \quad (1)$$

giving us a weak bound, $f(n) \geq \frac{n}{\log(n+1)}$.

However, we can do better. Since the elements of B are picked at random and don't depend on the D_i, we have that $|D_i \cap B| \sim Y_i$, where $Y_i \sim B(d_i, \frac{1}{2})$ is a binomial random variable, with $d_i = |D_i|$. Then

$$H\left(|D_i \cap B|\right) = H(Y_i) = \sum_{j=0}^{d_i} \binom{d_i}{j} 2^{-j} \log\left(\frac{2^{d_i}}{\binom{d_i}{j}}\right)$$

Using Theorem 2 here then gives us that

$$H(Y_i) \leq \frac{1}{2} \log d_i + \epsilon(c) \log d_i$$

where $\epsilon(c)$ denotes a factor that can be made arbitrarily small by choosing c appropriately. Substituting this back into (1) gives us the missing factor of 2 in our weaker bound provided above. ■

4 Bregman's Theorem

Bregman's Theorem states that for a bipartite graph G on color classes $\mathcal{E} = \{v_1, ..., v_n\}$ and $\mathcal{O} = \{w_1, ..., w_n\}$ with each $v_i \in$ having degree d_i, then

$$|\mathcal{M}_{perf}(G)| \leq \prod_{i=1}^{n} (d_i!)^{\frac{1}{d_i}}$$

where $\mathcal{M}_{perf}(G)$ refers to the number of perfect matchings.

A proof of Bregman's Theorem will be presented in the following class. See [3, 1] for more on these topics.

References

[1] S. Ajesh Babu and Jaikumar Radhakrishnan, An entropy based proof of the Moore bound for irregular graphs, 2010.

[2] D. Cantor and W. Mills, Determination of a subset from certain combinatorial properties, *Can. J. Math*, 1966.

[3] David Galvin, Three tutorial lectures on entropy and counting, 2014.