CSE533: Information Theory in Computer Science October 4, 2010

Lecture 2

Lecturer: Anup Rao *Scribe: Kevin Zatloukal*

In the last lecture, we introduced entropy $H(X)$, and conditional entry $H(X|Y)$, and showed how they are related via the chain rule. We also proved the inequality $H(X_1, \ldots, X_n) \leq H(X_1) + \cdots + H(X_n)$. In this lecture, we will derive two more useful inequalities and then give some examples where they are useful.

1 More Inequalities

Our first inequality shows that conditioning can only reduce the entropy. This is in contrast to general probabilities (where conditioning can either decrease or increase) as well as some of the quantities we look look at in the next lecture such as the mutual information.

Lemma 1. *$H(Y|X) \leq H(Y)$.*

Proof Using our inequality from last time, we have $H(X,Y) \leq H(X) + H(Y) \Rightarrow H(X) + H(Y|X) = H(X,Y) \leq H(X) + H(Y)$. Canceling out the common $H(X)$ term gives the desired result. ■

Our next inequality shows that the uniform distribution has the highest entropy. In fact, we will see that a distribution has the maximum entropy if and only if it is uniform.

Lemma 2. *Let $\mathcal{X}$ be the support of X. Then $H(X) \leq \log |\mathcal{X}|$.*

Proof We write $H(X)$ as an expectation and then apply Jensen's inequality (from Lecture 1) to the convex function $\log(1/t)$:

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log(1/p(x)) \leq \log \left(\sum_{x \in \mathcal{X}} p(x)(1/p(x)) \right) = \log \left(\sum_{x \in \mathcal{X}} 1 \right) = \log |\mathcal{X}|$$

■

2 Applications

2.1 Bounding the Binomial Tail

Suppose $2^n + 1$ people have each watched a subset of n movies. Since there are only 2^n possible subsets of these movies, there must be two people that have watched exactly the same subset by the pigeonhole principle.

We shall show how to argue something similar in less trivial cases where the pigeonhole principle does not apply. This first example will show one way to do that using information theory.

Suppose 2^n people have each watched a subset of $2n$ movies and every person has watched at least 90% of the movies. If the number of possible subsets meeting this constraint is less than 2^n, then we must have two people who have watched exactly the same subset, of movies as before. The following result will give us what we need.

Lemma 3. *If $k \leq n/2$, then $\sum_{i=0}^{k} \binom{n}{i} \leq 2^{nH(k/n)}$.*

We would like to compute $\sum_{i=0.9(2n)}^{2n} \binom{2n}{i}$. Since $\binom{2n}{i} = \binom{2n}{2n-i}$, this sum is equal to $\sum_{i=0}^{0.1(2n)} \binom{2n}{i} \leq 2^{2nH(1/10)} < 2^n$ since we can compute $H(0.1) < 0.469 < 0.5$.

It remains only to prove the lemma.

Proof Let $X_1X_2\cdots X_n$ be a uniformly random string sampled from the set of n-bit strings of weight at most k. Thus, $H(X_1X_2\cdots X_n) = \log\left(\sum_{i=0}^{k}\binom{n}{k}\right)$. Further, we have that $\Pr[X_i = 1] = \mathbb{E}[X_i]$. By symmetry, this probability is equal to $\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[X_j] = \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n} X_j\right] \leq \frac{k}{n}$, where we have used linearity of expectation. We can relate this to the entropy by using the fact that $p \leq H(p)$ for $0 \leq p \leq \frac{1}{2}$. Finally, applying our inequality from last time, we see that $H(X_1X_2\cdots X_n) \leq H(X_1) + \cdots + H(X_n) \leq nH(k/n)$. Thus, we have shown that $\log\left(\sum_{i=0}^{k}\binom{n}{i}\right) \leq nH(k/n)$, proving the lemma. ■

2.2 Triangles and Vees [KR10]

Let $G = (V, E)$ be a directed graph. We say that vertices (x, y, z) (not necessarily distinct) form a triangle if $\{(x, y), (y, z), (z, x)\} \subset E$. Similarly, we say they form a vee if $\{(x, y), (x, z)\} \subset E$. Let T be the number of triangles in the graph, and V be the number of vees.

We are interested in the relationship between V and T. From any particular triangle, we can get one vee from each edge, say (x, y), by repeating the second vertex: (x, y, y) is a vee. If the vertices of a triangle are distinct, the number of vees in the triangle is equal to the number of triangles contributed by the vertices of the triangle, since the three cyclic permutations of (x, y, z) are distinct triangles. However, the same edge could be used in many different triangles, so that this simple counting argument does not tell us anything about the relationship between V and T. We shall use an information theory based argument to show:

Lemma 4. *In any directed graph, $V \geq T$.*

Proof Let (X, Y, Z) be a uniformly random triangle. Then by the chain rule, $\log(T) = H(X, Y, Z) = H(X) + H(Y|X) + H(Z|X, Y)$. We will construct a distribution on vees with at least $\log T$ entropy, which together with Lemma 2, would imply that $\log V \geq \log T \Rightarrow V \geq T$.

Since conditioning can only reduce entropy, we have that $H(X, Y, Z) \leq H(X) + H(Y|X) + H(Z|Y)$. Now observe that by symmetry, the joint distribution of Y, Z is exactly the same as that of X, Y. Thus we can simply the bound to $\log T \leq H(X) + 2H(Y|X)$.

Sample a random vee, (A, B, C) with the distribution

$$q(a, b, c) = \Pr[X = a] \cdot \Pr[Y = b|X = a] \cdot \Pr[Y = c|X = a].$$

In words, we sample the first vertex with the same distribution as X, and then sample two independent copies of Y to use as the second and third vertices. Observe that if $q(a, b, c) > 0$, then (a, b, c) must be a vee, so this is a distribution on vees. On the other hand, the entropy of this distribution is $H(A) + H(B|A) + H(C|AB) = H(A) + H(B|A) + H(C|A) = H(X) + 2H(Y|X)$, which is at least $\log T$ as required. ■

The lemma is tight. Consider a graph with $3n$ vertices partitioned into three sets A, B, C with $|A| = |B| = |C| = n$. Suppose that the edges are $\{(a, b) \mid a \in A, b \in B\}$ and similarly for B, C and C, A. For each triple (a, b, c), we get three triangles – (a, b, c), (b, c, a), and (c, a, b) – so $T = 3n^3$. On the other hand, each vertex $a \in A$ is involved in n^2 vees of the form (a, b_1, b_2), and similarly for B, C. So $V = 3n^3$.

2.3 Counting Perfect Matchings [Rad97]

Suppose that we have a bipartite graph $G = (A \cup B, E)$, with $|A| = |B| = n$. Let $A = [n]$. A *perfect matching* is a subset of E that is incident to every vertex exactly once. Hence, it is a bijection between the sets A and B. How many possible perfect matchings can there be in a given graph? If we let d_v be the degree of vertex v, then a trivial bound on the number of perfect matchings is $\prod_{i \in A} d_i$. A tighter bound was proved by Brégman:

Theorem 5 (Brégman)**.** *The number of perfect matchings is at most* $\prod_{i\in A}(d_i!)^{1/d_i}$.

To see that this is tight, consider the complete bipartite graph. Any bijection can be chosen, and the number of bijections is the number of permutations of n letters, which is $n!$. In this case, the bound of the theorem is $\prod_{i=1}^{n}(n!)^{1/n} = (n!)^{(1/n)\cdot n} = n!$.

We give a simple proof of this theorem using information theory, due to Radhakrishnan [Rad97].

Proof Let ρ be a uniformly random perfect matching, and for simplicity, assume that $A = 1, 2, \ldots, n$. Write $\rho(i)$ for the neighbor of i under the matching ρ. Given a permutation τ, we write $\overline{\tau(i)}$ to denote the concatenation $\tau(i), \tau(i-1), \ldots, \tau(1)$.

Then, using the chain rule, the fact that conditioning only reduces entropy, and the fact that the entropy of a variable is at most the logarithm of the size of its support,

$$H(\rho) = \sum_{i=1}^{n} H(\rho(i)|\overline{\rho(i-1)}) \leq \sum_{i=1}^{n} H(\rho(i)) \leq \sum_{i=1}^{n} \log d_i = \log \prod_{i=1}^{n} d_i$$

This proves that the number of matchings is at most $\prod_{i=1}^{n} d_i$. Can we improve the proof? In computing $H(\rho(i)|\ldots)$, we are losing too much by throwing away all the previous values.

To improve the bound, let us start by symmetrizing over the order in which we condition the individual values of ρ. If π is any permutation, then conditioning in the order of $\pi(1), \pi(2), \ldots, \pi(n)$ shows that $H(\rho) = \sum_{i=1}^{n} H(\rho\pi(i)|\overline{\rho\pi(i-1)})$, where here $\rho\pi(i)$ denotes $\rho(\pi(i))$. Since this is true for any choice of π, we can take the expectation over a uniformly random choice of π without changing the value. Let L be a uniformly random index in $\{1, 2, \ldots, n\}$. Then,

$$H(\rho\pi(L)|L, \pi, \overline{\rho\pi(L-1)}) = \sum_{i=1}^{n} \frac{1}{n} H(\rho\pi(i)|\pi, \overline{\rho\pi(i-1)}) = \frac{1}{n} H(\rho).$$

Now we rewrite this quantity according to the contribution of each vertex in A:

$$\begin{aligned} H(\rho\pi(L)|L, \pi, \overline{\rho\pi(L-1)}) &= \sum_{i=1}^{n} \Pr[\pi(L) = i] \mathop{\mathbb{E}}_{\pi, L \text{ s.t. } \pi(L)=i} \left[H(\rho\pi(L)|L, \pi, \overline{\rho\pi(L-1)}) \right] \\ &= (1/n) \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\pi, L \text{ s.t. } \pi(L)=i} \left[H(\rho\pi(L)|L, \pi, \overline{\rho\pi(L-1)}) \right] \end{aligned}$$

Consider any fixed perfect matching ρ. We are interested in the number of possible choices for $\rho\pi(L)$ conditioned on $\pi(L) = i$, after $\rho\pi(1), \ldots, \rho\pi(L-1)$ have been revealed. Let $a_1, a_2, \ldots, a_{d_i-1}$ be such that the set of neighbors of i in the graph is exactly $\{\rho(a_1), \rho(a_2), \ldots, \rho(a_{d_i-1}), \rho(i)\}$. π, L in the expectation can be sampled by first sampling a uniformly random permutation π and then setting L so that $\pi(L) = i$. Thus, the ordering of $\{a_1, a_2, \ldots, a_{d_i-1}, i\}$ induced by π is uniformly random, and

$$|\{\rho(a_1), \rho(a_2), \ldots, \rho(a_{d_i-1})\} \cap \{\rho\pi(L-1), \ldots, \rho\pi(1)\}| = |\{a_1, a_2, \ldots, a_{d_i-1}\} \cap \{\pi(L-1), \ldots, \pi(1)\}|$$

is equally likely to be $0, 1, 2, \ldots, d_i - 1$. The number of available choices for $\rho(\pi(L))$ is equally likely to be bounded by $1, 2, \ldots, d_i$. This allows us to bound

$$\begin{aligned} (1/n)H(\rho) &= (1/n) \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\pi, L \text{s.t.} \pi(L)=i} \left[H(\rho\pi(L)|L, \pi, \overline{\rho\pi(L-1)}) \right] \\ &\leq (1/n) \sum_{i=1}^{n} \sum_{j=1}^{d_i} (1/d_i) \log(j) = (1/n) \sum_{i=1}^{n} \log \left((d_i!)^{1/d_i} \right) = (1/n) \log \left(\prod_{i\in A} (d_i!)^{1/d_i} \right), \end{aligned}$$

which proves that the number of perfect matchings is at most $\prod_{i\in A}(d_i!)^{1/d_i}$. ■

References

[KR10] Swastik Kopparty and Benjamin Rossman. The homomorphism domination exponent. Technical report, ArXiv, April 14 2010.

[Rad97] Jaikumar Radhakrishnan. An entropy proof of bregman's theorem. *J. Comb. Theory, Ser. A*, 77(1):161–164, 1997.