Information theory: Applications

Yuval Filmus

December 6, 2017

Applications to combinatorics:

1. Shearer's lemma:[1] Let S be a distribution over subsets of $\{1,\ldots,n\}$ satisfying $\Pr[i \in S] \geq p$ for all i, let $X = X_1,\ldots,X_n$ be a random n-dimensional vector, for $s \subseteq \{1,\ldots,n\}$ let $X_s := (X_i : i \in s)$ be the projection of X to the coordinates in s, and let $X_{<i} = (X_1,\ldots,X_{i-1})$.

   (a) Using the chain rule, show that for $i_1 < \cdots < i_k$,
   $$H(X_{i_1},\ldots,X_{i_k}) \geq H(X_{i_1}|X_{<i_1}) + \cdots + H(X_{i_k}|X_{<i_k}).$$

   (b) Deduce that $\mathbb{E}_S[H(X_S)] \geq \sum_{i=1}^n \Pr[i \in S] H(X_i|X_{<i})$.

   (c) Use the chain rule to conclude $\mathbb{E}_S[H(X_S)] \geq pH(X)$.

   (d) Let $\mathcal{F}$ be a finite collection of vectors of length n, and let $\mathcal{F}_s$ be the projection of $\mathcal{F}$ to the coordinates in s. Show that
   $$|\mathcal{F}|^p \leq \prod_s |\mathcal{F}_s|^{\Pr[S=s]}.$$

2. Application to triangle-intersecting families:
   Let $\mathcal{F}$ be a collection of graphs on the vertex set $\{1,\ldots,n\}$ (for even n) such that the intersection of any two graphs in $\mathcal{F}$ contains a triangle.

   (a) Show that if s is the complement of a triangle-free graph then $\mathcal{F}_s$ (the projection of $\mathcal{F}$ to the edges of s) is an intersecting family: every two edge-sets in $\mathcal{F}_s$ intersect.

   (b) Deduce that $|\mathcal{F}_s| \leq 2^{|s|-1}$, where $|s|$ is the number of edges in s.

   (c) Let S be the complement of a random complete bipartite graph with $n/2$ vertices in both parts. Determine the parameter p in Shearer's lemma.

   (d) Apply Shearer's lemma to conclude that $|\mathcal{F}| < 2^{\binom{n}{2}-2}$.

   The tight upper bound $2^{\binom{n}{2}-3}$ is proved using Fourier analysis.

[1] Proof attributed to Jaikumar Radhakrishnan.

Large deviation bounds:

1. Method of types: Let D be a countable domain. Given a probability distribution q over D, we can define a random variable X by $\Pr[X = \alpha] = q(\alpha)$. Conversely, given n elements $x_1, \ldots, x_n \in D$, we can define the *type* or *empirical distribution* of $\vec{x}$, denoted $T(\vec{x})$, by $\Pr[T(\vec{x}) = \alpha] = |\{i : x_i = \alpha\}|/n$.
   Let $X_1, \ldots, X_n$ be i.i.d. samples of q.

   (a) Show that if $T(x_1, \ldots, x_n) = p$ then
   $$\Pr[X_1 = x_1, \ldots, X_n = x_n] = 2^{-n(H(p)+D(p\|q))}.$$

   (b) Deduce that if $T(x_1, \ldots, x_n) = q$ then
   $$\Pr[X_1 = x_1, \ldots, X_n = x_n] = 2^{-nH(q)}.$$

   (c) By considering i.i.d. samples of p, deduce that
   $$|\{(x_1, \ldots, x_n) \in D^n : T(x_1, \ldots, x_n) = p\}| \leq 2^{nH(p)}.$$

   (d) Conclude that for all types p,
   $$\Pr_{X_1,\ldots,X_n}[T(X_1, \ldots, X_n) = p] \leq 2^{-nD(p\|q)}.$$

2. Sanov's inequality: Let D be a finite domain, let q be a probability distribution over D, and let Π be a closed set of probability distributions over D.
   Let $X_1, \ldots, X_n$ be i.i.d. samples of q.

   (a) Show that there are at most $(n+1)^{|D|}$ possible types attained by $X_1, \ldots, X_n$.

   (b) Let $p^* = \arg\min_{p \in \Pi} D(p\|q)$. Show that
   $$\Pr_{X_1,\ldots,X_n}[T(X_1, \ldots, X_n) \in \Pi] \leq (n+1)^{|D|} 2^{-nD(p^*\|q)}.$$

   (c) Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with success probability q, and let $p < q$. Upper bound the following probability:
   $$\Pr\left[\frac{X_1 + \cdots + X_n}{n} \leq p\right].$$

   Optional: Compare with Chernoff's bound (which you can look up).

   In some circumstances (for example, when Π is convex), we can get rid of the factor $(n+1)^{|D|}$.

Exercise: Show that a graph having m edges contains at most $(2m)^{3/2}/6$ triangles. (Hint: use Shearer's lemma with $n = 3$.)