**Information theory: assignment 2**

1. Describe a (probabilistic) algorithm that gets as input a $k$-CNF formula $f = \bigwedge_{i\in[s]} C_i$ so that $|\Gamma(C_i)| \leq 2^{k-3}$ for all $i$. The algorithm should output, with probability at least $2/3$, a satisfying assignment to $f$. Explain how to change the algorithm to get probability of success at least $1-\delta$ for a given $\delta > 0$. How does this affect the running time?

2. For two independent random variables $X, Y$ taking values in a finite abelian group $G$, define (the entropy version of) Ruzsa's distance as

$$d(X,Y) = H(X-Y) - (H(X)+H(Y))/2.$$

   Prove

$$d(X,X)/2 \leq d(X,-X) \leq 2d(X,X),$$

   where the two input to $d$ are "assumed to be independent" (e.g., $d(X,-X) := d(X,Y)$ with $Y$ independent of $X$ and distributed like $-X$).

3. Let $p, q$ be two distributions on $X \times Y$.

   Prove

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + \mathbb{E}_{x\sim p} D(p(y|x)||q(y|x)),$$

   where $p(x), q(x)$ are marginals of $x$, and $p(y|x), q(y|x)$ are conditional probabilities of $y$ given $x$.

4. Let $P_n$ be the families of probability distributions on $[n]$. Let $R_n(k)$ be the subset of $P_n$ of distributions $(p_i)_{i\in[n]}$ with expectation $k$, that is, $\sum_{i\in[n]} p_i i = k$.

   What is the distribution with largest entropy in $R_n(k)$?

5. A family $F$ of sets is *$k$-intersection-unique* if all sets in $F$ are of size exactly $k$ and for every $A \neq B$ in $F$, the set $A \cap B$ is not contained in any other sets in $F$, that is, $A \cap B \not\subset C$ for all $C \in F \setminus \{A, B\}$.

   Prove that the size of a $k$-intersection-unique family $F$ is at most $2^{0.9k+1}$. (Hint: use subadditivity, concavity, and reduce to problem of "constant size.")