IERG 6154: Network Information Theory
Homework 1

Due: Jan 17, 2019

1. *Inequalities*: Show that
   (a) $H(X|Z) \leq H(X|Y) + H(Y|Z)$
   (b) $I(X_1, X_2; Y_1, Y_2) \leq I(X_1; Y_1) + I(X_2; Y_2)$ if $p(x_1, x_2, y_1, y_2) = p(x_1, x_2)p(y_1|x_1)p(y_2|x_2)$
   (c) $I(X_1, X_2; Y_1, Y_2) \geq I(X_1; Y_1) + I(X_2; Y_2)$ if $p(x_1, x_2, y_1, y_2) = p(x_1)p(x_2)p(y_1, y_2|x_1, x_2)$
   (d) $h(X + aY) \geq h(X + Y)$ when $Y \sim \mathrm{N}(0, 1), a \geq 1$. Assume that X and Y are independent.

2. Prove or give a counterexample to the following inequalities:
   (a) $H(X_1, X_2, X_3) + H(X_1, X_2, X_4) + H(X_2, X_3, X_4) + H(X_1, X_3, X_4) \leq \frac{3}{2}[H(X_1, X_2) + H(X_1, X_4) + H(X_2, X_3) + H(X_3, X_4)]$.
   (b) $H(X_1, X_2, X_3) + H(X_1, X_2, X_4) + H(X_2, X_3, X_4) + H(X_1, X_3, X_4) \leq 3[H(X_1, X_2) + H(X_3, X_4)]$.

3. (Korner-Marton Identity) Show the following equality for any pair of sequences Y^n, Z^n

$$\sum_{i=1}^{n} I(Y^{i-1}; Z_i|Z_{i+1}^n) = \sum_{i=1}^{n} I(Z_{i+1}^n; Y_i|Y^{i-1})$$

   Note: Y^m and Y_n^m denote $Y_1, Y_2, ...Y_m$ and $Y_n, Y_{n+1}, ...Y_m$ respectively. There is an abuse of notation at the edges. Take Y^0 and Z_{n+1}^n as trivial random variables.

   Remark: This equality was discovered by Marton and was first used in a joint paper with Korner (1977). This was conveyed to me by Janos Korner. Unfortunately, it has been mistakenly called Csiszar-sum Lemma by the community after attributing the result to a later paper by Csiszar and Korner.

4. Prove that when $f : \mathbb{R} \to [0, \frac{1}{2}]$ such that $H(f(u))$ is convex and f is twice differentiable, then $H(f(u) * p)$ is convex in u for $p \in [0, 1]$. (Fan Cheng's extension of Mrs. Gerber's lemma). In the above, for $0 \leq x, y \leq 1$ define $x * y = x(1 - y) + y(1 - x)$. As a corollary note that $H(H^{-1}(u) * p)$ is convex in u for $p \in [0, 1]$ where $H^{-1}(u)$ as a mapping from $[0, 1]$ to $[0, \frac{1}{2}]$.

5. (Pinsker's Inequality)

   - Let $x, y \in (-1, 1)$. Show that (assume natural logarithms)

   $$\frac{1+x}{2}\log\left(\frac{1+x}{1+y}\right)+\frac{1-x}{2}\log\left(\frac{1-x}{1-y}\right)=\sum_{k=1}^{\infty}\left(\frac{x^{2k}+(2k-1)y^{2k}-2kxy^{2k-1}}{2k(2k-1)}\right),$$

   and in particular that all the terms on the right-hand-side are non-negative for each $k \geq 1$. Hence conclude that

   $$\frac{1+x}{2}\log\left(\frac{1+x}{1+y}\right)+\frac{1-x}{2}\log\left(\frac{1-x}{1-y}\right)\geq\frac{1}{2}(x-y)^2.$$

   - Given two distributions p and q, the total variation distance between them is defined as

   $$d_{TV}(p,q)=\frac{1}{2}\sum_x |p(x)-q(x)|.$$

   Show that

   $$D(p\|q)\geq 2d_{TV}^2(p,q).$$

   (Hint: use data-processing inequality of $D(p\|q)$.)

6. (Han's inequality or Shearer's lemma)

   - Show that

   $$H(X^n)\leq\frac{1}{n-1}\sum_{i=1}^{n}H(X^{n\setminus i})$$

   - More generally, show that given any collection $A_i \subseteq [1:n]$, then

   $$a_* H(X^n)\leq\sum_i H(X_{A_i}),$$

   where $a_* = \min_k \left(\sum_i 1_{k\in A_i}\right)$.

   Remark: The second generalization appears in a paper by Madiman and Tetali.

7. (Strong data processing inequalities for relative entropy) Given a channel $W(y|x)$, define

   $$\eta_W=\sup_{p,q:q\ll p}\frac{D(Wq\|Wp)}{D(q\|p)}.$$

   Show the following:

   - When W is $BEC\,(\epsilon)$ then $\eta_W = 1-\epsilon$
   - When W is $BSC\left(\frac{1+\rho}{2}\right)$ then $\eta_W = \rho^2$
   - When W is a Z-channel with $W(Y=0|X=0)=1$ and $W(y=0|x=1)=z$, then $\eta_W = 1-z$.