Information Theory and Coding Technique – Homework I

1. Let X be a random variable taking on a finite number of values. What is the (general) inequality relationship of $H(X)$ and $H(Y)$ if
   (a) $Y = 2^X$?
   (b) $Y = \cos X$?
2. Show that if $H(Y|X) = 0$, then Y is a function of X.
3. A function $\rho(x,y)$ is a metric if for all x, y,
   - $\rho(x,y) \geqq 0$.
   - $\rho(x,y) = \rho(y,x)$.
   - $\rho(x,y) = 0$ if and only if $x = y$.
   - $\rho(x,y) + \rho(y,z) \geqq \rho(x,z)$

   (a) Show that $\rho(X,Y) = H(X|Y) + H(Y|X)$ satisfies the first, second, and fourth properties above. If we say that $X = Y$ if there is a one-to-one function mapping from X to Y, the third property is also satisfied, and $\rho(X,Y)$ is a metric.

   (b) Verify that $\rho(X,Y)$ can also be expressed as
   $$\begin{aligned}\rho(X,Y) &= H(X)+H(Y)-2I(X;Y)\\ &= H(X,Y)-I(X;Y)\\ &= 2H(X,Y)-H(X)-H(Y).\end{aligned}$$
4. Let X_1 and X_2 be identically distributed but not necessarily independent. Let
   $$\rho = 1-\frac{H(X_2 \mid X_1)}{H(X_1)}.$$
   (a) Show that $\rho = \dfrac{I(X_1;X_2)}{H(X_1)}$.
   (b) Show that $0 \leqq \rho \leqq 1$.
   (c) When is $\rho = 0$?
   (d) When is $\rho = 1$?
5. Let $p(x, y)$ be given by:

X \ Y	0	1
0	1/3	1/3
1	0	1/3

   Find
   (a) $H(X)$, $H(Y)$
   (b) $H(X|Y)$, $H(Y|X)$
   (c) $H(X, Y)$
   (d) $H(Y) - H(Y|X)$
   (e) $I(X, Y)$

6. Let $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ form a Markov chain in this order:
   That is, let $p(x_1, x_2, \ldots, x_n) = p(x_1)\, p(x_2|x_1) \ldots p(x_n|x_{n-1})$.
   Reduce $I(X_1; X_2, X_3, \ldots, X_n)$ to its simplest form.
7. The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X|Y)$.
8. Let $X_1, X_2, \ldots, X_n$ be (possibly dependent) binary random variables. Suppose that are calculates the RUN LENGTHS $R = (R_1, R_2, \ldots)$ of this sequence (in order as they occur). For example, the sequence $X = 0001100100$ yields run lengths $R = (3,2,2,1,2)$.
   Compare $H(X_1, X_2, X_3, \ldots, X_n)$, $H(R)$, and $H(X_n, R)$.
   Show all equalities and inequalities, and bound all the differences.
9. Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements (i.e., $p_i \geqq 0$) and $\sum_{i=1}^{m} p_i = 1$). Define a new distribution $\mathbf{q}$ on $m - 1$ elements as $q_1 = p_1$, $q_2 = p_2$, …, $q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$ [i.e., the distribution $\mathbf{q}$ is the same as $\mathbf{p}$ on $\{1, 2, \ldots, m-2\}$, and the probability of the last element in $\mathbf{q}$ is the sum of the last two probabilities of $\mathbf{p}$]. Show that

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m)\, H\left(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\right).$$

10. Let X, Y, Z be three random variables with a joint probability mess function $p(x,y,z)$. The relative entropy between the joint distribution and the product of the marginals is

$$D(p(x,y,z) \,\|\, p(x)p(y)p(z)) = E\left[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}\right]$$

    Expand this in terms of entropies. When is this quantity zero?
11.
    (a) For any non-negative random variable X and any $t > 0$, show that

$$\Pr\{X \ge t\} \le \frac{EX}{t} \qquad \text{(Markov's inequality)}$$

    Exhibit a random variable that achieves this inequality with equality.

    (b) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that for any $\varepsilon > 0$,

$$\Pr\{|Y - \mu| > \varepsilon\} \le \frac{\sigma^2}{\varepsilon^2} \qquad \text{(Chebyshev's inequality)}$$

(c) Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $\overline{Z_n} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ be the sample mean. Show that

$$\Pr\{|\overline{Z_n} - \mu| > \varepsilon\} \le \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, $\Pr\{|\overline{Z_n} - \mu| > \varepsilon\} \to \infty$ as $n \to \infty$. (This is known as the weak low of large numbers.)

12. Let X_i be i.i.d. ~ p(x), $x \in \{1, 2, \ldots, m\}$.

Let $\mu = EX$ and $H = -\sum p(x) \log p(x)$.

Let $A^n = \{x^n \in \aleph^n: \ |-\frac{1}{n}\log p(x^n) - H| \le \varepsilon\}$, and

$B^n = \{x^n \in \aleph^n: \ |\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \le \varepsilon\}$.

(a) Does $\Pr\{x^n \in A^n\} \to 1$?

(b) Does $\Pr\{x^n \in A^n \cap \boldsymbol{B}^n\} \to 1$?

(c) Show that $|A^n \cap \boldsymbol{B}^n| \leqq 2^{n(H+\varepsilon)}$ for all n.

(d) Show that $|A^n \cap \boldsymbol{B}^n| \geqq (\frac{1}{2})2^{n(H-\varepsilon)}$ for n sufficiently large.

Remarks: (Convergence of random variables)

1. Given a sequence of random variables $X_1, X_2, \ldots$. We say that the sequence $X_1, X_2, \ldots$ converges to a random variable X:
   (1) In probability if for every $\varepsilon > 0$, $\Pr\{|X_n - X| > \varepsilon\} \to 0$.
   (2) In mean square if $E(X_n - X)^2 \to 0$.
   (3) With probability 1 (also called almost surely) if $\Pr\{\lim_{n\to\infty} X_n = X\} = 1$.

2. The notation $a_n \doteq b_n$ means

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$$

Thus, $a_n \doteq b_n$ means that a_n and b_n are equal to the first order in the exponent.