Problem Set 2

These problems have been selected from the Course Textbook by Cover and Thomas.

2.14 *Entropy of a sum.* Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.

(a) Show that $H(Z \mid X) = H(Y \mid X)$. Argue that if X, Y are independent, then $H(Z) \geq H(X)$ and $H(Z) \geq H(Y)$. Thus, the addition of *independent* random variables adds uncertainty.

(b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) Under what conditions does $H(Z) = H(X) + H(Y)$?

2.16 *Bottleneck.* Suppose that a (nonstationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus, $X_1 \to X_2 \to X_3$, that is, $p(x_1, x_2, x_3) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)$, for all $x_1 \in \{1, 2, \ldots, n\}$, $x_2 \in \{1, 2, \ldots, k\}$, $x_3 \in \{1, 2, \ldots, m\}$.

(a) Show that the dependence of X_1 and X_3 is limited by the bottleneck, by proving that $I(X_1; X_3) \leq \log k$.

(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

2.17 *Pure randomness and bent coins.* Let $X_1, X_2, \ldots, X_n$ denote the outcomes of independent flips of a bent coin. Thus, $Pr(Xi = 1) = p$, $Pr(Xi = 0) = 1 - p$, where p is unknown. We wish to obtain a sequence $Z_1, Z_2, \ldots, Z_K$ of fair coin flips from $X_1, X_2, \ldots, X_n$. Toward this end, let $f : \mathcal{X}^n \to \{0, 1\}^*$ (where $\{0, 1\}^* = \{\epsilon, 0, 1, 00, 01, ...\}$ is the set of all finite-length binary strings, including the empty string ϵ) be a mapping so that $(Z_1, Z_2, \ldots, Z_K) = f(X_1, X2, \ldots, X_n)$ where $Z_i \sim \text{Bernoulli}(\frac{1}{2})$ (and the Z_i are independent) and K may depend on $(X_1, \ldots, X_n)$. In order that the sequence $Z_1, Z_2, \ldots$ appear to be fair coin flips, the map f from bent coin flips to fair flips must have the property that all 2^k sequences $(Z_1, Z_2, \ldots, Z_k)$ of a given length k have equal probability (possibly 0), for $k = 1, 2, \ldots$. For example, for $n = 2$, the map $f(01) = 0$, $f(10) = 1$, $f(00) = f(11) = \epsilon$ (the empty string) has the property that $Pr(Z_1 = 1 \mid K = 1) = Pr(Z_1 = 0 \mid K = 1) = \frac{1}{2}$. Give reasons for the following inequalities:

$$\begin{aligned} nH(p) &\overset{(a)}{=} H(X_1, \ldots, X_n) \\ &\overset{(b)}{\geq} H(Z_1, Z_2, \ldots, Z_K, K) \\ &\overset{(c)}{=} H(K) + H(Z_1, \ldots, Z_K \mid K) \\ &\overset{(d)}{=} H(K) + E(K) \\ &\overset{(e)}{\geq} E(K). \end{aligned}$$

Thus, no more than $nH(p)$ fair coin tosses can be derived from $(X_1, \ldots, X_n)$, on the average. Exhibit a good map f on sequences of length $n = 4$.

2.19 *Infinite entropy.* This problem shows that the entropy of a discrete random variable can be infinite. Let $A = \sum_{n=2}^{\infty} (n \log^2 n)^{-1}$. [It is easy to show that A is finite by bounding the infinite sum by the integral of $(x \log^2 x)^{-1}$.] Show that the integer-valued random variable X defined by $Pr(X = n) = (An \log^2 n)^{-1}$ for $n = 2, 3, \ldots,$ has $H(X) = +\infty$.

2.21 *Markov's inequality for probabilities.* Let $p(x)$ be a probability mass function. Prove, for all $d > 0$, that

$$Pr(p(X) \leq d) \log \frac{1}{d} \leq H(X).$$

2.27 *Grouping rule for entropy.* Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements (i.e., $p_i \geq 0$ and $\sum_{i=1}^{m} p_i = 1$). Define a new distribution $\mathbf{q}$ on $m - 1$ elements as $q_1 = p_1$, $q_2 = p_2$, $\ldots$, $q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$. In other words, the distribution $\mathbf{q}$ is the same as $\mathbf{p}$ on $\{1, \ldots, m-2\}$, while the last mass q_{m-1} of $\mathbf{q}$ is the sum of the last two masses p_{m-1} and p_m of $\mathbf{p}$. Show that

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

3.4 *AEP*. Let X_i be i.i.d. $\sim p(x)$, $x \in \mathcal{X}$. Let $\mu = E(X)$ and $H = -\sum p(x) \log p(x)$. Let

$$A^n = \left\{ x^n \in \mathcal{X}^n \colon \left| -\frac{1}{n} \log p(x^n) - H \right| \leq \epsilon \right\},$$

and let

$$B^n = \left\{ x^n \in \mathcal{X}^n \colon \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \leq \epsilon \right\}.$$

(a) Does $P(X^n \in A^n) \to 1$ as $n \to \infty$?

(b) Does $P(X^n \in A^n \cap B^n) \to 1$ as $n \to \infty$?

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

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

3.5 *Sets defined by probabilities.* Let $X_1, X_2, \ldots$ be an i.i.d. sequence of discrete random variables over $\mathcal{X}$ with entropy $H(X)$. Let

$$C_n(t) = \{x^n \in \mathcal{X}^n \colon p(x^n) \geq 2^{-nt}\}.$$

(a) Show that $|C_n(t)| \leq 2^{nt}$.

(b) For which values of t does $P(X^n \in C_n(t)) \to 1$ as $n \to \infty$?

3.7 *AEP and source coding.* A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.

(a) Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.

(b) Calculate the probability of observing a source sequence for which no codeword has been assigned.

(c) Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

3.13 *Calculation of typical set.* To clarify the notion of a typical set $A_\epsilon^{(n)}$ and the smallest set of high probability $B_\delta^{(n)}$, we will calculate these sets for a simple example. Consider a sequence of i.i.d. binary random variables, $X_1, X_2, \ldots, X_n$, where the probability that $X_i = 1$ is 0.6 (and therefore the probability that $X_i = 0$ is 0.4).

(a) Calculate $H(X)$.

(b) With $n = 25$ and $\epsilon = 0.1$, which sequences fall in the typical set $A_\epsilon^{(n)}$? What is the probability of the typical set? How many elements are there in the typical set? (This involves computation of a table of probabilities for sequences with k ones, $0 \leq k \leq 25$, and finding those sequences that are in the typical set.

(c) How many elements are there in the smallest set that has probability at least 0.9?

(d) How many elements are there in the intersection of the sets in parts (b) and (c)? What is the probability of this intersection?