EE376A/Stats376A: Homework #2

Due on Thursday January 26, 5pm

You can hand in the homework either after class or deposit it, before 5 PM, in the EE376A drawer of the class file cabinet on the second floor of the Packard Building.

1. **Entropy of functions of a random variable.**

   Let X be a discrete random variable.

   (a) Show that the entropy of a function of X is less than or equal to the entropy of X by justifying the following steps:

$$H(X, g(X)) \stackrel{(a)}{=} H(X) + H(g(X)|X)$$
$$\stackrel{(b)}{=} H(X).$$
$$H(X, g(X)) \stackrel{(c)}{=} H(g(X)) + H(X|g(X))$$
$$\stackrel{(d)}{\geq} H(g(X)).$$

   Thus $H(g(X)) \leq H(X)$.

   (b) A Data-Processing Inequality: Show that if $Z = g(Y)$ then $H(X|Y) \leq H(X|Z)$. Interpret this in words.

2. **Entropy of time to first success.**

   A fair coin is flipped until the first head occurs. Let X denote the number of flips required.

   (a) Find the entropy $H(X)$ in bits. The following expressions may be useful:

$$\sum_{n=1}^{\infty} r^n = r/(1-r), \quad \sum_{n=1}^{\infty} nr^n = r/(1-r)^2.$$

   (b) Find an "efficient" sequence of yes-no questions of the form, "Is X contained in the set S?". Compare $H(X)$ to the expected number of questions required to determine X.

   (c) Let Y denote the number of flips until the second head appears. Thus, for example, $Y = 5$ if the second head appears on the 5th flip. Argue that $H(Y) = H(X_1 + X_2) < H(X_1, X_2) = 2H(X)$, and interpret in words.

3. **Concavity of entropy**

(a) Show that $H(X)$ is concave in the pmf of X, that is for any two random variables X_1, X_2 with pmfs p_1, p_2 and real number $0 \leq \lambda \leq 1$

$$H(X) \geq \lambda H(X_1) + (1-\lambda)H(X_2)$$

where X is distributed according to $\lambda p_1 + (1-\lambda)p_2$. (Hint: interpret the right hand side of the above inequality as $H(X|Y)$ for some appropriate chosen random variable Y.)

(b) Use the concavity and label-invariance properties of entropy to give another proof that the uniform distribution maximizes entropy. (Hint: think about simple case of the entropy of a Bernouilli random variable first.)

4. **Law of Large Numbers for i.i.d. random variables**

Problem 3.1, Cover and Thomas, 2nd Edition.

5. **Law of Large Numbers for Markov chains**

Let $Z_1, Z_2, \ldots$ be a sequence of states of the stationary Mickey-mouse Markov chain with symmetric transition probabilities (i.e. $\alpha = \beta$.) For this problem we will label the states to be $+1$ and -1 instead of 0 and 1 to simplify a bit of the calculations.

(a) What is the stationary distribution of this chain?

(b) Give an expression for $\Pr[X_k = j | X_1 = i]$ for all i, j, k. What happens to the random variables X_1, X_k as $k \to \infty$?

(c) Using part (b) or otherwise, compute the variance of $\bar{Z}_n = 1/n \sum_{i=1}^{n} Z_i$.

(d) Using part(c) or otherwise, state and prove a law of large numbers for this chain.

(e) Conjecture what happens to

$$\log \frac{1}{p(Z_1, Z_2, \ldots, Z_n)}$$

for large n. Explain why you think your conjecture is true, but a detailed proof is not necessary.

6. **The Typical Set vs. High Probability Sets**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables each defined over space $\mathcal{X}$. We define the typical set $A_\epsilon^{(n)}$ as the set of all sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

(a) Consider the special case when $X_i \sim \text{Bern}(p)$. For small ϵ, can you say roughly how many 1's are in a typical sequence?

(b) Show that in general $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$ where $|A|$ denotes the number of elements in A.

(c) Show that for any $\eta > 0$ if n is sufficiently large $\Pr(A_\epsilon^{(n)}) > 1 - \eta$, and hence

$$|A_\epsilon^{(n)}| \geq (1-\eta)2^{n(H(X)-\epsilon)}.$$

(d) Fix $0 < \delta < \frac{1}{2}$. Let $B_\delta^{(n)} \subset \mathcal{X}^n$ such that $\Pr(B_\delta^{(n)}) > 1 - \delta$. Prove that for n sufficiently large

$$\Pr\big(B_\delta^{(n)} \cap A_\epsilon^{(n)}\big) > 1 - \eta - \delta.$$

(e) Prove that for n sufficiently large for any $B_\delta^{(n)} \subset \mathcal{X}^n$ such that $\Pr(B_\delta^{(n)}) > 1 - \delta$ we have

$$|B_\delta^{(n)}| > (1-\eta-\delta)2^{n(H(X)-\epsilon)}.$$

Hence, any high probability set have at least roughly $2^{nH(X)}$ sequences, not only the typical set.

(f) Using part (d) or otherwise, prove Theorem 4 of Shannon's paper. Are the sequences considered in this theorem the same as typical sequences? Explain.