EE6340: Information Theory
Problem Set 4

1. *Shuffles increase entropy.* Argue that for any distribution on shuffles T and any distribution on card positions X that

$$\begin{aligned} H(TX) &\geq H(TX|T) && (1) \\ &= H(T^{-1}TX|T) && (2) \\ &= H(X|T) && (3) \\ &= H(X), && (4) \end{aligned}$$

if X and T are independent.

2. *Monotonic convergence of the empirical distribution.* Let $\hat{p}_n$ denote the empirical probability mass function corresponding to $X_1, X_2, ...X_n$ i.i.d. $\sim p(x)$, $x \in \mathcal{X}$. Specifically,

$$\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i = x) \tag{5}$$

is the proportion of times that $X_i = x$ in the first n samples, where I is an indicator function

   (a) Show for $\mathcal{X}$ binary that

$$ED(\hat{p}_{2n}||p) \leq ED(\hat{p}_n||p). \tag{6}$$

   Thus the expected relative entropy "distance" from the empirical distribution to the true distribution decreases with sample size. *Hint:* Write $\hat{p}_{2n} = \frac{1}{2}\hat{p}_n + \frac{1}{2}\hat{p}_n^{'}$ and use the convexity of D.

   (b) Show for an arbitrary discrete $\mathcal{X}$ that

$$ED(\hat{p}_n||p) \leq ED(\hat{p}_{n-1}||p). \tag{7}$$

   Hint: Write $\hat{p}_n$ as the average of n empirical mass functions with each of the n samples deleted in turn.

3. *Random box size.* An n-dimensional rectangular box with sides $X_1, X_2, ..., X_n$ is to be constructed. The volume is $V_n = \prod_{i=1}^{n} X_i$. The edge length l of a n-cube with the same volume as the random box is $l = V_n^{1/n}$. Let $X_1, X_2, ..., X_n$ be i.i.d. uniform random variables over the unit interval [0,1]. Find $\lim_{n\to\infty} V_n^{\frac{1}{n}}$, and compare to $(EV_n)^{1/n}$. Clearly the expected edge length does not capture the idea of the volume of the box.

4. *Monotonicity of entropy per element.* For a stationary stochastic process $X_1, X_2, ..., X_n$, show that,

(a)

$$\frac{H(X_1, X_2, ..., X_n)}{n} \leq \frac{H(X_1, X_2, ..., X_{n-1})}{n-1}. \tag{8}$$

(b)

$$\frac{H(X_1, X_2, ..., X_n)}{n} \geq H(X_n|X_{n-1}, ...X_1) \tag{9}$$

5. *Doubly stochastic matrices.* An $n \times n$ matrix $P = [P_{ij}]$ is said to be *doubly stochastic* if $P_{ij} \geq 0$ and $\sum_j P_{ij} = 1$ for all i and $\sum_i P_{ij} = 1$ for all j. An $n \times n$ matrix P is said to be a *permutation* matrix if it is doubly stochastic and there is precisely one $P_{ij} = 1$ in each row and each column.

   It can be shown that every doubly stochastic matrix can be written as the convex combination of permutation matrices.

   (a) Let $\mathbf{a^t}$=$(a_1, a_2....a_n), a_i \geq 0, \sum a_i = 1$, be a probability vector. Let $\mathbf{b} = \mathbf{a}\, P$, where P is doubly stochastic. Show that $\mathbf{b}$ is a probability vector and that $H(b_1, b_2, ...b_n) \geq H(a_1, a_2, ...a_n)$. Thus stochastic mixing increases entropy.

   (b) Show that a stationary distribution μ for a doubly stochastic matrix P is the uniform distribution.

   (c) Conversely, prove that if the uniform distribution is a stationary distribution for a Markov transition matrix P, then P is doubly stochastic.

6. *The entropy rate of a dog looking for a bone.* A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:
   $(X_0, X_1, ...) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, ...)$

   (a) Find $H(X_0, X_1, X_2, ..., X_n)$.

   (b) Find the entropy rate of this browsing dog.

   (c) What is the expected number of steps the dog takes before reversing direction?