Problem Set 1

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

2.1 *Coin flips.* 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: if $|r| < 1$, then

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad \sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2}.$$

(b) A random variable X is drawn according to this distribution. 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.

2.3 *Minimum entropy.* What is the minimum value of $H(p_1, \ldots, p_n)$ as $(p_1, \ldots, p_n)$ ranges over the set of n-dimensional probability vectors? Find all choices of $(p_1, \ldots, p_n)$ that achieve this minimum.

2.4 *Entropy of functions of a random variable.* Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X by justifying the following steps:

$$\begin{aligned}
H(X, g(X)) &\stackrel{(a)}{=} H(X) + H(g(X) \mid X) \\
&\stackrel{(b)}{=} H(X) \\
H(X, g(X)) &\stackrel{(c)}{=} H(g(X)) + H(X \mid g(X)) \\
&\stackrel{(d)}{\geq} H(g(X))
\end{aligned}$$

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

2.5 *Zero conditional entropy.* Show that if $H(Y \mid X) = 0$, then Y is a function of X, i.e., for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$.

2.8 *Drawing with and without replacement.* An urn contains r red, w white, and b black balls. Which has higher entropy, drawing $k \geq 2$ balls from the urn with replacement or without replacement? Set it up and show why. (There is both a difficult way and a relatively simple way to do this.)

2.10 *Entropy of a disjoint mixture.* Let X_1 and X_2 be discrete random variables draw according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over respective alphabets $\mathcal{X}_1 = \{1, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$. Let

$$X = \begin{cases} X_1 & \text{with probability } \alpha, \\ X_2 & \text{with probability } 1-\alpha. \end{cases}$$

(a) Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$ and α.

(b) Maximize over α to show that that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

2.11 *Measure of correlation.* 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 $\rho = \frac{I(X_1;X_2)}{H(X_1)}$. (Here $I(X_1; X_2)$ is the *mutual information* between X_1 and X_2, given as $I(X_1; X_2) = H(X_1) - H(X_1 \mid X_2) = H(X_2) - H(X_2 \mid X_1)$.)

(b) Show that $0 \leq \rho \leq 1$.

(c) When is $\rho = 0$?

(d) When is $\rho = 1$?

3.1 *Markov's inequality and Chebyshev's inequality*

(a) *(Markov's inequality)* For any nonnegative random variable X and any $t > 0$, show that

$$P[X \geq t] \leq \frac{E(X)}{t}.$$

Exhibit a random variable that achieves this inequality with equality.

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

$$P[|Y - \mu| \geq \epsilon] \leq \frac{\sigma^2}{\epsilon^2}.$$

(c) *(Weak law of large numbers)* 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 $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ be the sample mean. Show that

$$P[|\bar{Z}_n - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}.$$

Thus $P[|\bar{Z}_n - \mu| \geq \epsilon] \to 0$ as $n \to \infty$. This is known as the *weak law of large numbers*.