Discrete Entropy
Exercise Sheet 9

Question 1. Let $X = (X_1, \dots, X_n)$ be a discrete random variable. Show that there are non-negative constants $h_1, \dots, h_n$ such that $\mathbb{H}(X) = \sum_i h_i$ and

$$\sum_{i \in I} h_i \leq \mathbb{H}(X_I)$$

for every $I \subseteq I$.

(Hint : Consider the proof of the Bollobás-Thomason Box Theorem in the notes).

Question 2. Suppose X is distributed on some subset of $\mathbb{N}$ and has $\mathbb{E}(X) = \mu$ and let Y be a *geometric* random variable distributed as

$$\mathbb{P}(Y = k) = p(1-p)^k$$

where $\frac{1-p}{p} = \mu$. Show that $\mathbb{E}(Y) = \mathbb{E}(X)$ and by considering $D_{KL}(X||Y)$ show that $\mathbb{H}(X) \leq \mathbb{H}(Y)$.

Question 3. Let Y be a uniform random variable taking values in $[n]$ and let $Y = p_1^{X_1} p_2^{X_2} \dots p_k^{X_k}$ where $k = \pi(n)$, the number of primes less than $[n]$.

By considering $H(Y)$, show that $\pi(n) \to \infty$. Using the second question show that

$$\sum_{p \leq n} \frac{\log p}{p} \approx \log n$$

where the sum ranges over p prime.

Question 4. Suppose $\mathcal{C} \subset 2^{[n]}$ and for each $i \in [n]$ let p_i be the proportion of $C \in \mathcal{C}$ with $i \in C$. Show that $\log |\mathcal{C}| \leq \sum h(p_i)$ (See Sheet 5).

Suppose $\mathcal{C} = \{C_1, \dots, C_m\} \subseteq [n]^{(k)}$ is a collection of k element sets such that $C_i \cap C_j \not\subseteq C_k$ for any distinct i, j, k. Show that $m = O(2^k)$. Suppose that $p_i \leq 1/2$ for all i, show that there is some $\alpha < 2$ such that $m = O(\alpha^k)$.

(Hint: Show that there is no $\{i, k\} \neq \{k, \ell\}$ such that $C_i \cap C_j = C_k \cap C_\ell$. Consider the set $\mathcal{C}' = \{C_i \cap C_j : i, j \leq m\}$. Also, $f(p) = \frac{h(p^2)}{p}$ is concave on $[0, 1]$ and is < 2)

Question 5. A *path separator* of a graph G is a set of paths $\mathcal{P} = \{P_1, \dots, P_t\}$ such that for every pair of edge $e, f \in E(G)$ there are paths P_e and P_f with $e \in P_e \not\ni f$ and $f \in P_f \not\ni e$. The *path separation number* of G, denoted $\mathrm{psn}(G)$ is the smallest number of paths in a path separator.

Suppose $|V(G)| = n$ and $e(G) = m \geq 2(n-1)$, and let $p = \frac{n-1}{m}$ Show that

$$\frac{\log(m)}{\mathbb{H}(p)} \leq \mathrm{psn}(G).$$