Discrete Entropy
Exercise Sheet 3

Given two random variables X and Y over the same set $\mathcal{V}$ the *statistical distance* between X and Y is defined as

$$\mathrm{SD}(X,Y) = \frac{1}{2}\sum_{v\in\mathcal{V}} |\mathbb{P}(X=v) - \mathbb{P}(Y=v)|.$$

A classical encryption scheme K, C for M with an encryption function e is ϵ-*statistically secure* if for all $m, m' \in \mathcal{M}$

$$SD(e(m,K), e(m',K)) \leq \epsilon.$$

Over the next few questions we will show that if K and M are uniformly distributed and the encryption scheme is ϵ-statistically secure then

$$|\mathcal{K}| \geq (1-\epsilon)|\mathcal{M}|$$

Question 1. Suppose K, C is a classical encryption scheme for M with an encrpytion function e, with K and M uniformly and independently distributed. Show that M and C are independent if and only if for every $m, m' \in \mathcal{M}$ the distributions of $e(m,K)$ and $e(m',K)$ are identical.

Question 2. Given $c \in \mathcal{C}$ let

$$\mathcal{D}(c) = \{m \in \mathcal{M} \colon \text{ there exists } k \in K \text{ such that } e(m,k) = c\}.$$

Show that $|\mathcal{D}(c)| \leq |\mathcal{K}|$. Hence show that there exists $m, m' \in \mathcal{M}$ such that

$$\mathbb{P}(m' \in \mathcal{D}(e(m,K)) \leq \frac{|\mathcal{K}|}{|\mathcal{M}|}.$$

Question 3. You may assume without proof that

$$\mathrm{SD}(X,Y) = \max_{f:\mathcal{V}\to\{0,1\}} |\mathbb{P}(f(X)=1) - \mathbb{P}(f(Y)=1)|.$$

Show that

$$SD(e(m,K), e(m',K)) \geq 1 - \frac{|\mathcal{K}|}{|\mathcal{M}|}.$$

Deduce the result claimed before Question 1.

Question 4 (Han's inequality)**.** Let $X = (X_1, X_2, \ldots, X_m)$ be a random variable and for $I \subset [m]$ let $X_I = (X_i \colon i \in I)$. Recall that $\mathbb{H}(X) \leq \sum_i H(X_i)$, show that

$$\sum_i H(X_i) \geq \frac{1}{m-1}\sum_{I\in[m]^{(2)}} \mathbb{H}(X_I) \geq \mathbb{H}(X).$$

(* Can you generalise the above to subsets of size 3? What about general $1 \leq k \leq m$?)

Question 5. A code $C : \mathcal{X} \to \{0,1\}^*$ is *fix-free* there is no $x, x' \in \mathcal{X}$ such that $C(x)$ is a prefix of $C(x')$ or $C(x)$ is a suffix of $C(x')$. Show that if

$$\sum_{c\in\mathcal{C}} \frac{1}{2^{||c||}} \leq \frac{1}{2}$$

then there is a fix-free code C with range $\mathcal{C}$. (* Does the converse hold?)