Exercises lecture 1/JL Roch - Entropy

1. Prove that the entropy $H(S)$ is maximum when S is a discrete source with uniform probability distribution.
Hint : use the following Gibbs's lemma (note that $\forall t > 0 \log_e t \leq (t-1)$).

Lemma 1.1 Gibb's lemma. *Let $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$ be two probability distributions on n elementary events. Then $\sum_i p_i \log \frac{1}{p_i} \leq \sum_i p_i \log \frac{1}{q_i}$.*

2. When flipping C_i, the probability of obtaining a head is p_i, and a tail $(1-p_i)$. Define the random variable X_i be the output of the coin tossing : *head* or *tail*. What is the information $I(X_i = \text{head})$? What is the entropy $H(X_i)$?
Complete the following table where $p_1 = \frac{1}{2}$, $p-2 = \frac{1}{4}$; $p_3 = \frac{1}{2^{-10}}$.

i	p_i	$I(X_i = \text{head})$	$I(X_i = \text{tail})$	$H(X_i)$
1	$\frac{1}{2}$			
2	$\frac{1}{4}$			
1	$\frac{1}{2^{10}}$			

3. Oscar is looking for a mysterious file on Professor John's computer disk ; he has no information on it, so he is performing a uniform random search.
Oscar knows that there are N files on the computer disk.
- n_1 files are in the directory `COURS` ;
- n_2 files are named `exam.tex` ;
- n_3 files are named `exam.tex` in the directory `COURS`.

Give the amount of information brought by each of the following hint :
- (a) "The file is in the directory `COURS`".
- (b) "The file is named `exam.tex`".
- (c) "The file is named `exam.tex` and is in the directory `COURS`".

Application : $N = 65536$; $n_1 = 1024$; $n_2 = 256$; $n_3 = 16$. Compute the corresponding values. Verify that $I(c) = I(a) + I(c|a) = I(b) + I(c|b)$.

4. Prove that, for any cryptosystem, $H(K|C) \geq H(P|C)$; i.e. the uncertainty on the key is at least as large as the entropy on the plaintext.

5. Home exercise. Proof of Shannon's theorem on perfect secrecy.

1. Let A, B, C be three random variables ; (A, B) denotes the random variable of the couple A and B. Prove that :
   (a) $H(A) \leq H((A,B)) = H((B,A))$
   (b) $H((A,B)) = H(A) + H((B|A)) = H(B) + H(A|B)$.
   (c) $H(A|C) \leq H((A,B)|C)$
   (d) $H((A,B)|C) = H(A|C) + H(B|(A,C)) = H(B|C) + H(A|(B,C))$
   (e) A and B are independent iff $H((A,B)|C) = H(A|C) + H(B|C)$
2. In a symmetric cryptosystem, let P, C, K denote respectively the discrete random variables corresponding to the plaintext source, the ciphertext and the secret key source. If the cryptosystem provides perfect secrecy (or unconditionnal security) – i.e. $H(P|C) = H(P)$ –, then prove that the entropy of the secret key source K is larger than the one of the plaintext source P. In other word, prove :

$$[H(P|C) = H(P)] \implies [H(K) \geq H(P)]$$

   which is Shannon's theorem on perfect secrecy.
   Hint : Note that $H(P) = H(P|C) \leq H((P,K)|C)$ and conclude using previous properties on entropy.

6. Indistinguishability and perfect secrecy.
Let k be a key of length n uniformly chosen in $\{0,1\}^n$; let (E_k, D_k) be an encryption scheme for messages of length m :

$$\forall k \in \{0,1\}^n, \forall x \in \{0,1\}^m : \quad D_k(E_k(x)) = x.$$

Besides, let U_n denote the uniform distribution over $\{0,1\}^n$.

1. In this question only, $n = m$ and $E_k = E_k^{OTP} : E_k^{OTP}(x) = x \oplus k$ where $\oplus$ denotes the bitwise XOR. What is $D_k^{OTP}(x)$?
   For any $x, x' \in \{0,1\}^m$, show that the distribution $E_{U_n}^{OTP}(x)$ is the same as $E_{U_n}^{OTP}(x')$.
2. For any (E_k, D_k) : if $n < m$, show that there exist two messages $x, x' \in \{0,1\}^m$ such that $E_{U_n}(x)$ is not the same distribution as $E_{U_n}(x')$.
3. If $n \geq m$, we consider (E_k, D_k) such that $\forall x, x' : E_{U_n}(x)$ is the same distribution as $E_{U_n}(x')$. Show that E is then unconditionally secure *(hint : use Bayes theorem)*.

7. Additional exercises. CLRS, 2nd edition, Appendix C, Counting and probability.
- Probability : C.2-2–9 p 1105–1106.
- Discrete random variable : C.3-1 –7 pp 1110-1111.