Prerequisites Test

Cryptography and Security 2023

Exercise 1 Probability

Given a discrete random variable X, we recall the definition of entropy $H(X)$

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

And the associated notion of joint entropy $H(X,Y)$ and conditional entropy $H(X|Y)$

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y)$$

$$H(X|Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x|y)$$

Question 1.1

Show the following *chain rule* for entropy.

$$H(X,Y) = H(Y) + H(X|Y)$$

Proof. Recall the chain rule for probabilities : $p(x,y) = p(x|y)p(y)$. Then we have

$$\begin{aligned}
H(X,Y) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y) \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y)(\log_2 p(x|y) + \log_2 p(y)) \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x|y) - \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(y) \\
&= H(X|Y) + H(Y).
\end{aligned}$$

$\square$

Question 1.2

Consider the following symmetric encryption system : We consider a plaintext space $\mathcal{M} = \{m_1, m_2, m_3\}$, key space $\mathcal{K} = \{k_1, k_2, k_3\}$ and a ciphertext space $\mathcal{C} = \{1, 2, 3, 4, 5\}$. We denote by P, C, K the random variables for the plaintext, ciphertext and keys respectively. The encryption is given by the following matrix:

$$\begin{pmatrix} & m_1 & m_2 & m_3 \\ k_1 & 2 & 1 & 4 \\ k_2 & 1 & 5 & 3 \\ k_3 & 2 & 4 & 5 \end{pmatrix}.$$

This means that e.g. the encryption of m_1 with k_1 is 2.
Additionally we suppose that the keys are equiprobable, and that the plaintexts appear with the following probabilities :

$$p(m_1) = \frac{1}{2}, \quad p(m_2) = \frac{3}{20}, \quad p(m_3) = \frac{7}{20}.$$

Compute the following entropies, justify your computations when necessary and simplify all the expressions as much as possible :

1. $H(P)$
2. $H(K)$
3. $H(C)$
4. $H(P|C)$

Answer.

1. $$\begin{aligned} H(P) &= -\frac{1}{2}\log\frac{1}{2} - \frac{3}{20}\log\frac{3}{20} - \frac{7}{20}\log\frac{7}{20} \\ &= \text{(simplifying...)} \\ &= \frac{3}{2} + \frac{1}{2}\log 5 - \frac{3}{20}\log 3 - \frac{7}{20}\log 7 \simeq 1.44. \end{aligned}$$

2. The keys are equiprobable.
$$H(K) = -\log\frac{1}{3} = \log 3 \simeq 1.58.$$

3. Step 1: Compute $p(c_i) = \Pr(C = i)$ for each i.
$$\begin{aligned} & p(c_1) = p(m_2, k_1) + p(m_1, k_2) = \frac{13}{60} \\ & \text{similarly, } p(c_2) = \frac{1}{3},\ p(c_3) = \frac{7}{60},\ p(c_4) = \frac{1}{6},\ p(c_5) = \frac{1}{6} \end{aligned}$$
Step 2: Compute $H(C)$
$$\begin{aligned} H(C) &= \text{(use formula and replace with the values computed above, simplify)} \\ &= 1 + \log 3 + \frac{1}{3}\log 5 - \frac{7}{60}\log 7 - \frac{13}{60}\log(13) \\ &\simeq 2.23 \end{aligned}$$

4. Direct computation method. (Might use chain rule if they want to).
   Step 1: Compute $P(m_i|C=i)$.

   1.1 Rule out impossibilities:

$$
\begin{aligned}
0 &= p(m_3|C=1) = p(m_2|C=2) = p(m_3|C=2) \\
&= p(m_1|C=3) = p(m_2|C=3) = p(m_1|C=4) \\
&= p(m_1|C=5)
\end{aligned}
$$

   1.2 Observe "obvious ones":

$$
1 = p(m_1|C=2) = p(m_3|C=3). \tag{1}
$$

   1.3 Compute the combinations that are left.

$$
\begin{aligned}
p(m_1|C=1) &= \frac{p(m_1,c_1)}{p(c_1)} = \frac{p(m_1,k_2)}{p(c_1)} = \frac{10}{13} \\
p(m_2|C=1) &= \frac{3}{13} \\
p(m_2|C=4) &= \frac{3}{10} \\
p(m_3|C=4) &= \frac{7}{10} \\
p(m_2|C=5) &= \frac{3}{10} \\
p(m_3|C=5) &= \frac{7}{10}.
\end{aligned}
$$

   Step 2: Put everything together and compute $H(P|C)$.

$$
H(P|C) = \frac{1}{6} - \frac{3}{20}\log 3 + \frac{1}{6}\log 5 - \frac{7}{30}\log 7 + \frac{13}{60}\log 13 \simeq 0.46.
$$

Question 1.3

What do you think of this cryptosystem ? Do you think it is secure ? Give one concrete example/justification.

Answer. Look at equation 1 for example. We observe that if the ciphertext is 2 then we know for sure that the encrypted message is m_1 and similarly for $c=3$ and m_3. This is not good.

Exercise 2 Arithmetic

Question 2.1

Let $a = 247338$ and $b = 139776$. Using Euclid's Algorithm, compute $\gcd(a, b)$.

Answer. By reading Table 1, we have $\gcd(a, b) = 546$.

index	q_{i-1}	r_i
0		247338
1		139776
2	1	107562
3	1	32214
4	3	10920
5	2	10374
6	1	546
7	19	0

Table 1: Extended Euclidean Algorithm for $a = 247338$ and $b = 139776$. The step r_{i+1} at index $i + 1$ is defined by $r_{i+1} = r_{i-1} - q_i r_i$ with $0 \le r_{i+1} < |r_i|$ and initial values $r_0 = a$ and $r_1 = b$.

Question 2.2

Let $(F_n)_{n\ge0}$ be the Fibonacci sequence, starting from $F_0 = 0$ and $F_1 = 1$ and recursively defined by $F_{n+2} = F_n + F_{n+1}$ for $n \ge 0$.

2.2a Prove that $\gcd(F_n, F_{n-1}) = 1$ for all $n \ge 1$.

Proof. The proof goes by induction on n. Since $\gcd(F_1, F_0) = \gcd(1, 0) = 1$, the claim is verified for $n = 1$. Assume that it holds up to n and let us show that it holds for $n + 1$, namely $\gcd(F_{n+1}, F_n) = 1$. Since $\gcd(a + b, b) = \gcd(a, b)$ for every integers $a, b \in \mathbb{Z}$, we have

$$\gcd(F_{n+1}, F_n) = \gcd(F_n + F_{n-1}, F_n) = \gcd(F_n, F_{n-1}).$$

By induction hypothesis, $\gcd(F_n, F_{n-1}) = 1$, whence the result. □

2.2b Prove that $F_{m+n+1} = F_{m+1}F_{n+1} + F_m F_n$ for all $n, m \ge 0$.

Proof. The proof goes by induction on n and extends by symmetry to m. The claim is trivially verified for the base case $n = 0$. Assume that the claim holds for all $0 \le n \le k$ and let us show that it holds for $n = k + 1$. We have

$$\begin{aligned} F_{k+m+2} &\triangleq F_{k+m} + F_{k+m+1} \\ &= (F_{m+1}F_k + F_m F_{k-1}) + (F_{m+1}F_{k+1} + F_m F_k) \\ &= F_{m+1}(F_k + F_{k+1}) + F_m(F_{k-1} + F_k) \\ &\triangleq F_{m+1}F_{k+2} + F_m F_{k+1}, \end{aligned}$$

where the second equality is deduced from the induction hypothesis for $n = k-1$ and $n = k$. □

2.2c Let $m, n \geq 1$. Prove that if m divides n, then F_m divides F_n.

Proof. Let $n = km$ be a multiple of m. The proof goes by induction on k. Clearly, if $k = 1$, then $F_n = F_m$ is divisible by F_m. Assume that the result holds up to some k and let us prove that it holds for $k + 1$. By the previous question,

$$F_{(k+1)m} = F_{km+m} = F_{km+1}F_m + F_{km}F_{m-1}.$$

By induction hypothesis, F_{km} is divisible by F_m, say $F_{km} = sF_m$ for some $s \in \mathbb{Z}$. Therefore $F_{(k+1)m} = F_m(F_{km+1} + sF_{m-1})$ is divisible by F_m as well. □

2.2d Prove that $\gcd(F_m, F_n) = F_{\gcd(m,n)}$ for all $n, m \geq 1$.

Hint: Show that Euclid's algorithm can be applied to F_m and F_n and to their subscripts simultaneously.

Proof. Consider the Euclidean division $n = qm + r$. By the second claim,

$$\gcd(F_m, F_n) \triangleq \gcd(F_m, F_{qm+r}) = \gcd(F_m, F_{qm+1}F_r + F_{qm}F_{r-1}).$$

Since F_m divides F_{qm}, this implies that

$$\gcd(F_m, F_{qm+1}F_r + F_{qm}F_{r-1}) = \gcd(F_m, F_{qm+1}F_r).$$

Combining the divisibility property with the fact that $\gcd(F_{qm}, F_{qm+1}) = 1$, we deduce that $\gcd(F_m, F_{qm+1}F_r) = \gcd(F_m, F_r)$. We now note that the subscripts of F follow the same process as in Euclid's algorithm. In particular, one may apply Euclid's algorithm to F_m and F_n and their subscripts simultaneously until eventually reaching $\gcd(m, n) = \gcd(s, 0) = s$. On the Fibonacci's side, we would end up with $\gcd(F_m, F_n) = \gcd(F_s, 0) = F_s = F_{\gcd(m,n)}$. □

2.2e Let $m, n \geq 1$ and $m \neq 2$.[1] Deduce that if F_m divides F_n, then m divides n.

Proof. Since $F_m = \gcd(F_m, F_n) = F_{\gcd(m,n)}$, it follows that $m = \gcd(m, n)$. □

[1] **POST-TEST EDIT**: One should assume $m \neq 2$ since $F_2 = 1$ divides every F_n, but obviously 2 does not divide each n.

Exercise 3 Algorithms

Question 3.1 Random root finding

Given a polynomial p of degree d (i.e. $deg(p) = d$) defined over a finite field $\mathbb{F}$ and a maximum number of iterations N. Algorithm 1 tries to find a root of this polynomial. Compute the expected number of iterations of this algorithm.

Algorithm 1: FindRoot

Input: $p(x)$ of degree d over $\mathbb{F}$ and a maximum number of iterations N.

for $1 \dots N$ **do**
    $r \xleftarrow{\$} \mathbb{F}$ ▷ $\xleftarrow{\$}$: sampling uniformly at random
    if $p(r) = 0$ **then**
        return r
return $\perp$

Answer. Let n_d be the number of roots of p, we have $\Pr[p(r) = 0] = n_d/|\mathbb{F}|$. Hence the expected number of iterations is the following:

$$E(\#iterations) = \sum_{i=1}^{N-1} i \cdot (1 - n_d/|\mathbb{F}|)^{i-1} \cdot (n_d/|\mathbb{F}|) + N \cdot (1 - n_d/|\mathbb{F}|)^{N-1}.$$

Note that N is a fixed input to the algorithm. Assuming $N = \infty$ does not yield the correct answer.

Question 3.2 Asymptotic Notation

Definition 1 ($O(.)$)**.** Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = O(g(n))$ if and only if there exist constants $N > 0$ and $C > 0$ such that for all $n \geq N$, $|f(n)| \leq C|g(n)|$.

Definition 2 ($\Omega(.)$)**.** Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = \Omega(g(n))$ if and only if there exist constants $N > 0$ and $C > 0$ such that for all $n \geq N$, $|f(n)| \geq C|g(n)|$.

Definition 3 ($o(.)$)**.** Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = o(g(n))$ if and only if for all $C > 0$ there exists a constant $N > 0$ such that for all $n \geq N$, $|f(n)| < C|g(n)|$.

Definition 4 ($\omega(.)$)**.** Assume that $f(n)$ and $g(n)$ are two functions. We write $f(n) = \omega(g(n))$ if and only if for all $C > 0$ there exists a constant $N > 0$ such that for all $n \geq N$, $|f(n)| > C|g(n)|$.

3.2a Show that $3n^3 - 5n + 16 = \Omega(n^3)$.

Answer. Using the definition of $\Omega(.)$, we need to find constants N and C such that $3n^3 - 5n + 16 \geq Cn^3$. We write

$$3n^3 - 5n + 16 = n^3 + (2n^3 - 5n) + 16$$

if we choose $N = 2.$ and $C = 1$ (you need to mention that $2n^3 - 5n$ is increasing), we have

$$\begin{aligned} 3n^3 - 5n + 16 &\geq n^3 + 16 \\ &\geq 1 \cdot n^3 \end{aligned}$$

3.2b Let $f(n)$ and $g(n)$ be positive functions. Show that $f(n) = O(g(n))$ implies $g(n) = \Omega(f(n))$.

Answer. We know that $f(n) \geq 0$ and $f(n) = O(g(n))$. Hence by Definition 1, $0 \leq f(n) \leq C \cdot g(n)$ for some C. We need to show that $0 \leq C' \cdot g(n) \leq f(n)$ for some C' (i.e. $g(n) = \Omega(f(n))$). Simply by setting $C' = 1/C$ and dividing the first inequality by C, we obtain $0 \leq C' \cdot g(n) \leq f(n)$.

3.2c If $f(n) = n^2 \log(n)$ and $g(n) = n^3$. Which of the following statements are true? List all.

1. $f = O(g)$
2. $f = \Omega(g)$
3. $f = o(g)$
4. $g = \omega(f)$

Answer. 1, 3, 4.