

## INFO002 : cryptologie, 2021–2022

### TD 2 : un peu de probabilité – indice de coïncidence et secret parfait

Pierre Hyvernat

Laboratoire de mathématiques de l'université de Savoie

bâtiment Chablais, bureau 17, poste : 94 22

email : Pierre.Hyvernat@univ-smb.fr

www : <http://www.lama.univ-smb.fr/~hyvernat/>

#### Exercice 1 : indice de coïncidence

L'indice de coïncidence d'un texte ( $c_i$ ) est défini comme la probabilité que deux caractères  $c_i$  et  $c_j$  ( $i \neq j$ ) pris au hasard soient égaux.

Question 1.

- Quel est l'indice de coïncidence de abcdefghijklmnopqrstuvwxyz ?
- Et pour aabbccddeeffgghhiiijkkllmmnnooppqrrssttuuvvwwxxyyzz ?
- Quel est l'indice de coïncidence d'un long texte aléatoire utilisant  $n$  symboles différents ?

Question 2.

- Que pouvez-vous dire de l'indice de coïncidence d'une *permutation* d'un texte ?
- Que pouvez-vous dire de l'indice de coïncidence d'un chiffrement monoalphabétique d'un texte ?

Question 3. On se donne un texte ( $c_i$ ) de longueur  $n$ , et on note  $n_a$  le nombre d'occurrences du caractère **a**,  $n_b$  le nombre d'occurrences du caractère **b**, etc.

Donnez une formule qui permet de calculer son indice de coïncidence.

Question 4. Pourquoi l'indice de coïncidence d'un texte clair est-il plus élevé que  $1/26 \approx 0.038$  ?

Question 5. Comment peut-on utiliser l'indice de coïncidence pour essayer de deviner la taille de la clé d'un message chiffré *polyalphabétiquement* ?

#### Exercice 2 : recherche exhaustive

Question 1. On suppose que le cardinal des clés est  $N$ , et que les clés sont générées uniformément (chaque clé est donc tirée avec probabilité  $1/N$ ).

On suppose connu un texte clair  $m$  et son chiffré  $c$ . On recherche une clé  $k$  telle que  $D_k(c) = m$ .

Quelle est l'espérance du nombre de déchiffrements à effectuer avant de trouver une telle clé lors d'une recherche exhaustive ?

*Indice* : les événements "la bonne clé est la première", "la bonne clé est la deuxième", "la bonne clé est la troisième", ... sont tous équiprobables.

#### Exercice 3 : chiffre de Vernam et secret parfait

Notation :

- $\mathcal{M}$  dénote l'ensemble des textes clairs,
- $\mathcal{K}$  dénote l'ensemble des clés,
- $\mathcal{C}$  dénote l'ensemble des textes chiffrés,
- pour  $m \in \mathcal{M}$ ,  $P(M = m)$  dénote "la probabilité a priori que le texte clair soit égal à  $m$ ",
- pour  $k \in \mathcal{K}$ ,  $P(K = k)$  dénote "la probabilité a priori que la clé soit égale à  $k$ ",
- pour  $c \in \mathcal{C}$ ,  $P(C = c)$  dénote "la probabilité a priori que le texte chiffré soit égal à  $c$ ".
- $P(A|B)$  dénote la probabilité conditionnelle de l'évènement  $A$  sachant  $B$ . Elle est égale à  $P(A \cap B)/P(B)$  et n'est donc définie que lorsque  $P(B) > 0$ .

L'objectif est de montrer que le chiffre de Vernam, aussi appelé "bloc-note à usage unique" ("one time pad" en anglais) a la propriété du secret parfait, c'est à dire que

$$\forall m \in \mathcal{M}, \forall c \in \mathcal{C}, \quad P(M = m | C = c) = P(M = m)$$

(En français : la connaissance du texte chiffré ne donne pas d'information sur le texte clair.)

Pour rappel, le chiffre de Vernam utilise  $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0, 1\}^n$ , et la fonction de chiffrement est le XOR (noté  $\oplus$ ) entre le message et la clé. Les clés sont choisies aléatoirement avec une distribution uniforme :  $P(K = k) = 2^{-n}$ .

*Question 1.* On suppose que les messages sont uniformément répartis :  $P(M = m) = 2^{-n}$ .

Calculez :

- $P(M = m \cap C = c)$  en regardant les clés qui permettent d'avoir  $M = m$  et  $C = c$ ,
- $P(C = c)$  en considérant toutes les manières possibles d'avoir un message chiffré égal à  $c$ .

Déduisez en que si les messages clairs sont uniformément répartis, le chiffre de Vernam a la propriété du secret parfait.

*Question 2.* Malheureusement, la distribution des messages clairs *n'est pas* uniforme.

Adaptez le calcul précédent au cas où la distribution pour  $P(M = m)$  est quelconque.

#### Exercice 4 : attaques sur le "bloc note à usage multiple"

Le "one time pad" a la propriété du secret parfait. Le "two times pad", où la clé est réutilisée (une ou plusieurs fois) n'est plus sûr. Voici un exemple d'attaque sur le " $t$  times pad" ( $t \geq 2$ ).

Dans la suite, on suppose que

$$\mathcal{M} = \mathcal{K} = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, \sqcup\}^n$$

( $\sqcup$  représente l'espace)

Le chiffrement s'effectue par le XOR bit à bit sur la suite des codes ASCII ( $a = 01100001$  (97),  $b = 01100001$  (98), ...  $z = 01111010$  (122), et  $\sqcup = 00100000$  (32))

*Question 1.* Que pouvez-vous dire du XOR entre 2 lettres par rapport au XOR entre 1 lettre et  $\sqcup$  ?

*Question 2.* Décrivez une attaque possible pour retrouver la clé si vous disposez de  $t > 2$  textes chiffrés (de taille  $n$ ) avec la même clé (elle aussi de taille  $n$ ).

*Question 3.* Que pensez-vous du cas  $t = 2$  ?