

INFO910 : cryptologie

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 (t_i) est défini comme la probabilité que deux caractères t_i et t_j ($i \neq j$) pris au hasard dans t soient égaux.

Question 1.

- Quel est l'indice de coïncidence d'un long texte aléatoire utilisant uniquement les 26 lettres de l'alphabet ?
- Quel est l'indice de coïncidence d'un long texte aléatoire utilisant n symboles différents ?
- Quel est l'indice de coïncidence de `abcdefghijklmnopqrstuvwxyzz` ?
- Et pour `aabbccddeeffgghhiijjkkllmmnnooppqqrrssttuuvvwwxxyyzz` ?

Question 2.

- Que pouvez-vous dire de l'indice de coïncidence d'une *permutation* de t ?
- Que pouvez-vous dire de l'indice de coïncidence d'un chiffrement monoalphabétique de t ?

Question 3. On se donne un texte (t_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 l'indice de coïncidence de t .

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* ?

Pour toute la suite du TD, on note

- \mathcal{M} pour l'ensemble des textes clairs,
- \mathcal{K} pour l'ensemble des clés,
- \mathcal{C} pour l'ensemble des textes chiffrés.

On suppose données les distributions discrètes

- " $P(M = m)$ " pour "la probabilité a priori que le texte clair soit égal à m ",
- " $P(K = k)$ " pour "la probabilité a priori que la clé soit égale à k ".

On note " $P(C = c)$ " pour "la probabilité a priori que le texte chiffré soit égal à c ".

Exercice 2 : recherche exhaustive

Question 1. On suppose que le cardinal de \mathcal{K} est N , et que les clés sont générées uniformément ($P(K = k) = 1/N$).

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.

Exercice 3 : chiffre de Vernam et secret parfait

Rappels

- On note $P(A|B)$ 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 $P(B) > 0$.
- On dit que A et B sont *indépendants* lorsque $P(A|B) = P(A)$.
- Le théorème de Bayes affirme que, si $P(B) > 0$, alors

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

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.

On suppose données des distributions de probabilités pour $P(M = m)$ et $P(K = k)$.

Question 1. Donnez une expression pour $P(C = c|M = m)$ en fonction de la distribution $P(K = k)$.

Question 2. Donnez une expression pour $P(C = c)$ en fonction des distributions $P(M = m)$ et $P(K = k)$.

(Indice : sous quelles conditions sur le message clair et la clé est-ce que le message chiffré est c ?)

Question 3. En utilisant la formule de Bayes et les questions précédentes, montrez que

$$P(M = m|C = c) = P(M = m) \iff P(K = m \oplus c) = \sum_{m' \in \mathcal{M}} P(M = m')P(K = m' \oplus c)$$

Question 4. En posant $c = 0^n$, déduisez en que la condition implique que $P(K = m) = 2^{-n}$.

Question 5. En reprenant l'expression de la question 3, montrez que $P(K = m) = 2^{-n}$ implique le secret parfait.

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 $t = 2$?