Semaine 9 : Série d'exercices sur les signaux et l'entropie

1 [N1] Questions-test

Pour chaque paire de mots ci-dessous, spécifier lequel des deux a la plus grande entropie, ou s'ils ont la même entropie.

a) AAAAAAHH et HAHAHAHA

b) ABBA et BEBE

c) CALC et CALCUL

d) MEDITERRANNEE et MEDETERRENNEE

e) EPFL et EEPPFFLL

2 [N2] Quelques pâtisseries

Voici des séquences de 16 lettres chacune (*inclus* les espaces, cette fois). Ordonnez-les dans l'ordre croissant (!) de leurs entropies respectives :

a) TRESSE AU BEURRE

b) PAIN AU CHOCOLAT

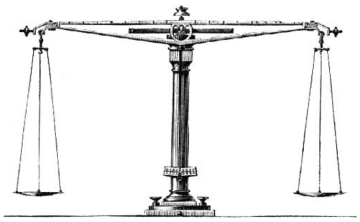
c) CROISSANT FOURRE

d) CHOUX A LA CREME

e) GATEAUX MILANAIS

3 [N3] Quelques pièces de monnaie

Le problème de détecter une fausse pièce (détectable seulement avec une balance) dans un groupe de pièces identiques permet de généraliser à un cas ternaire les notions d'entropie et de code binaire vues en cours. Le problème de trouver une lettre par des questions oui–non se transpose bien à la recherche par pesée d'une fausse pièce dans un ensemble de pièces donné. La différence principale provient de l'outil utilisé ; il ne s'agit plus de questions oui/non mais d'une pesée qui peut donner 3 réponses : « la balance penche à gauche », « la balance penche à droite » ou « la balance reste stable ».



a) Si la balance est bien utilisée, chaque pesée peut diviser par 3 le nombre de pièces à tester. En faisant le parallèle avec l'algorithme de dichotomie, quelle est la complexité d'une telle méthode de pesée pour trouver une fausse pièce plus légère parmi N ? Comment définirait-on l'entropie dans ce cas là ?

b) Supposez que vous ayez devant vous 3 pièces de monnaie, identiques en apparence, mais quelqu'un vous dit que l'une d'elles est fausse et pèse un peu moins lourd que les autres. Combien de pesées sont nécessaires pour trouver la fausse pièce ? Quel est le lien avec l'entropie ? Avec un code de Shannon-Fano ?

c) Si vous avez maintenant 9 pièces de monnaie, comprenant toujours une seule fausse pièce plus légère. Combien de pesées sont nécessaires pour trouver la fausse pièce ? Expliquez comment vous effectuez les pesées (c.-à-d. donnez un algorithme pour trouver la pièce fautive).

d) [optionnel, pour le fun] Supposons maintenant que vous ayez toujours 9 pièces dont l'une est fausse, mais sans vous dire si elle est plus légère ou au contraire plus lourde que les autres. Combien de pesées sont nécessaires pour trouver la fausse pièce ? Expliquez comment vous effectuez les pesées (c.-à-d. donnez un algorithme pour trouver la pièce fautive et dire si elle est plus lourde ou plus légère).

4 [N3] Entropie et mots de passe

Nous avons vu en cours une définition spécifique de l'entropie qui (comme indiqué transparent 15/31) :

- peut être étendue à des probabilités estimées ailleurs que sur la séquence elle-même ;
- peut être généralisée à d'autres « jeux » que celui du choix d'une lettre (par exemple le choix d'une *séquence*).

Reprenons ici ces deux points et appliquons les à la complexité des mots de passe.

a) [rappel de cours] Pour une séquence de n lettres différentes, quelles sont les bornes de l'entropie telle que définie en cours ?

b) Si l'on considère une séquence X de n lettres, quelle est l'entropie de la séquence XX (c.-à-d. la séquence répétée deux fois) ?

Par exemple, quelle est l'entropie de « YAPUKAYAPUKA » par rapport à l'entropie de « YAPUKA » ?

Cela vous semble-t-il une bonne propriété :

1. du point de vue du jeu de choisir une lettre au hasard dans le mot ?
2. du point de vue du choix d'un mot de passe ?

Il est donc important, lorsqu'on définit l'entropie, de bien savoir de quel « jeu » on parle : c'est moins l'entropie d'une séquence en tant que telle que celle de la façon dont on l'utilise.

c) Pour des séquences de longueur L écrites avec un alphabet de n lettres différentes :

1. quelles sont les bornes de l'entropie telle que définie en cours ?
2. combien y a-t-il de séquences différentes possibles ?
3. quelle serait la définition de l'entropie si l'on choisissait une *séquence* au hasard (et non plus une lettre) ?

d) Une définition possible de la « complexité » d'un mot de passe pourrait être :

$$R = \log_2\left(n^L\right) = L\,\log_2(n)$$

où n est la taille de l'alphabet utilisé et L est la longueur du mot de passe.

Exemples : si $n = 26$ et $L = 8$, alors $R = 8\log_2(26) \simeq 8 \times 4.7 = 37.6$.

1. Quand est-ce que l'entropie d'un mot de passe (selon la définition que vous avez proposé en **c.3**) est égale à la mesure de complexité proposée ci-dessus ?
2. De combien faut-il augmenter la longueur L d'un mot de passe pour doubler sa « complexité », sans changer l'alphabet utilisé ?
3. De combien faut-il augmenter la taille n de l'alphabet utilisé pour doubler la « complexité » d'un mot de passe, sans changer sa longueur ?
4. SANS LES RÉVÉLER, calculez la « complexité » des mots de passe que vous utilisez fréquemment sur Internet (tout en sachant qu'un mot de passe avec une « complexité » de 40 peut être décrypté en une minute environ par un PC standard, quand bien même cette mesure de « complexité » est une quantité *logarithmique* de l'effort réél à fournir pour craquer un mot de passe (= essayer toutes les possibilités)).

   Prenez $n = 26$ si vous n'utilisez que des lettres majuscules, $n = 52$ pour que des lettres, $n = 62$ pour des lettres et des chiffres (ou $n = 36$ si ce ne sont que des lettre majuscules et des chiffres), $n = 128$ si vous utilisez des symboles et $n = 256$ si vous utilisez des lettres accentuées (mais elles posent d'autres problèmes pratique par ailleurs !).

NB : Bien entendu, l'entropie des caractères n'est pas le seul facteur déterminant dans la sécurité d'un mot de passe. Si le mot de passe est un mot du dictionnaire ou une date de naissance, sa sécurité est bien moindre. Encore une fois, ce qu'il faut considérer ici ce sont les probabilités des *séquences* en tant que telles (en tant qu'entité entière et non plus simplement en tant que groupe de lettres individuelles). Pour plus d'infos sur le sujet, voir p. ex. `http://en.wikipedia.org/wiki/Password_strength` ou également `http://ophcrack.sourceforge.net/`. PAR CONTRE, ne saisissez *JAMAIS* vos mots de passe sur des sites tels que `http://www.howsecureismypassword.net/` !

Pour aller plus loin

5 [N3] Est-ce que l'entropie augmente ou diminue ?

Considérons une séquence de lettres le longueur N finie.

a) Est-il *toujours* vrai que si l'on ajoute à cette séquence une lettre qui fait déjà partie de la séquence, alors l'entropie [1] de la nouvelle séquence (de longueur $N + 1$) diminue ?

b) Est-il *toujours* vrai que si on ajoute à cette séquence une lettre qui ne fait pas partie de la séquence, alors l'entropie de la nouvelle séquence (de longueur $N + 1$) augmente ?

Note : Pour répondre à chacune de ces deux questions, une première étape pourrait être de vous convaincre de la réponse (oui ou non) à l'aide de plusieurs exemples. Ensuite, vous devez soit *prouver* que c'est toujours vrai, soit trouver un contre-exemple à l'affirmation énoncée.

Cours ICC : liens théorie $\longleftrightarrow$ Programmation

L'exercice 5 de la série 8 du cours de Programmation I vous propose de programmer le calcul de l'entropie d'une chaîne.

(Vous pourriez alors vous en servir pour vérifier certaines de vos réponses à cette série d'exercices.)

Retrouvez tous les exercices de programmation liés à la partie théorie du cours sous le lien « Exercices de C++ spécifiques à ICC (lien programmation - théorie) » en bas de la page Moodle du cours, dans la section « Ressources complémentaires / Références ».

1. telle que définie en cours : probabilités estimées sur la séquence elle-même