

Projet n°2 (jeu du pendu)

Le jeu du pendu est un jeu à deux joueurs (A et B), où A choisit un mot (dont il précise le nombre de lettres) et B tente de le deviner en un minimum de coups. A chaque coup, le joueur B propose une lettre, et le joueur A lui donne la position de toutes les occurrences de cette lettre dans ce mot. Pour gagner, le joueur B doit deviner le mot choisi par A en moins de n coups (n étant un nombre fixé à l'avance). L'objectif de ce projet est de concevoir un programme qui permet au joueur B de gagner systématiquement pour la valeur de n la plus petite possible.

1. **Donnée de type dictionnaire.** Soit $\mathcal{M} = \{m_1, m_2, \dots, m_N\}$ l'ensemble des mots possibles, chaque mot m s'écrivant sous la forme $m = x_1x_2 \dots x_k$ (on ne considère que les mots de k lettres, puisque k est connu à l'avance). Soit x la lettre proposée par le joueur B au premier coup, montrer que l'entropie de la variable aléatoire "réponse du joueur A", est, en supposant tous les mots de \mathcal{M} équiprobables,

$$h(x) = \sum_{I \subset \{1,2,\dots,k\}} -p_x(I) \log p_x(I),$$

où $p_x(I)$ désigne la proportion de mots de \mathcal{M} qui ont la lettre x exactement aux indices donnés par l'ensemble I , c'est-à-dire

$$p_x(I) = \frac{1}{N} \#\{m = x_1x_2 \dots x_k \in \mathcal{M}, \{i; x_i = x\} = I\}.$$

À partir de la fonction h , comment le joueur B doit-il choisir la première lettre à proposer ? Décrire les étapes suivantes du processus de décision (choix de la deuxième lettre et des suivantes), puis concevoir un programme qui se substitue au joueur B. Tester ce programme sur la base de mots `mots.txt` (voir le site web du cours) ou sur une autre base de votre choix. Calculer numériquement (à l'aide d'un programme) le nombre $n(k)$ de questions nécessaires dans le cas le pire en fonction du nombre k de lettres du mot à deviner. Donner l'un des mots qui nécessite le plus grand nombre de questions pour B. Calculer numériquement (à l'aide d'un programme) le nombre $n'(k)$ de questions nécessaires en moyenne en fonction du nombre k de lettres du mot à deviner. Quel est la longueur k la plus défavorable pour B en moyenne ?

2. **Donnée de type dictionnaire et fréquences.** On suppose maintenant que l'objectif pour le joueur B est de répondre juste en posant, en moyenne, le moins de questions possible. On suppose donc que le joueur A choisit le mot $m \in \mathcal{M}$

selon une certaine distribution de probabilité $p(m)$, connue de B. Quelle lettre le joueur B doit-il proposer en premier pour maximiser l'information moyenne ? Décrire et implémenter un programme qui se substitue au joueur B , et le tester sur la base (avec fréquences) `frequence_mots.txt` ou sur une autre base de votre choix.