TD 5 : Entropie

Correction

1 Echauffement

Question 1 : Correction : On propose l'algorithme suivant :

```
tirage_parfait() =
  faire
    x ← tirage_biaisé()
    y ← tirage_biaisé()
  tant que x = y
  retourner x
```

Les valeurs des variables x et y sont clairement indépendantes à chaque tour de boucle. Lorsque la boucle termine, on a donc :

$$\begin{aligned}\mathbb{P}(x = 1 \mid x \neq y) &= \frac{\mathbb{P}(x = 1 \wedge x \neq y)}{\mathbb{P}(x \neq y)} = \frac{\mathbb{P}(x = 1 \wedge y = 0)}{\mathbb{P}(x = 1 \wedge y = 0) + \mathbb{P}(x = 0 \wedge y = 1)} \\ &= \frac{\mathbb{P}(x = 1)\mathbb{P}(y = 0)}{\mathbb{P}(x = 1)\mathbb{P}(y = 0) + \mathbb{P}(x = 0)\mathbb{P}(y = 1)} = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2}\end{aligned}$$

ce qui montre que l'algorithme simule une pièce parfaite.

Pour calculer le nombre de lancers en moyenne, il faut calculer le nombre de tours de boucle en moyenne. La probabilité de sortir de la boucle est $\mathbb{P}(x \neq y) = 2p(1-p)$. La probabilité de faire k tours de boucle est :

$$\begin{aligned}R_1 &= \mathbb{P}(x \neq y) = 2p(1-p) \\ R_2 &= \mathbb{P}(x = y)\mathbb{P}(x \neq y) = (1 - 2p(1-p))2p(1-p) \\ R_k &= (1 - 2p(1-p))^{k-1}2p(1-p)\end{aligned}$$

Le nombre moyen de tours de boucle est :

$$\begin{aligned}\sum_{k=1}^{\infty} kR_k &= 2p(1-p)\sum_{k=1}^{\infty} k(1 - 2p(1-p))^{k-1} \\ &= 2p(1-p)\frac{1}{(1 - 1 + 2p(1-p))^2} = \frac{1}{2p(1-p)}\end{aligned}$$

Comme chaque tour comporte deux lancers, le nombre moyens de lancers est $\frac{1}{p(1-p)}$.

Question 2 : Correction : (1) On considère l'algorithme suivant :

```
tirage_parfait(n) =
k ← ⌈log n⌉
faire
  x ← 0
  pour i de 1 à k:
    y ← tirage()
    si y = 1 alors
      x ← x + 2^{i−1}
tant que x ≥ n
retourner x + 1
```

Invariant de boucle interne : $x \sim \mathcal{U}([0, 2^{i-1}-1])$. Ainsi, en fin de boucle interne $x \sim \mathcal{U}([0, 2^k-1])$.

Lors de la sortie de la boucle externe, on a $x < n$ et pour $i \in [0, n-1]$,

$$\mathbb{P}(x = i \mid x < n) = \frac{\mathbb{P}(x = i \wedge x < n)}{\mathbb{P}(x < n)}$$

Or $\mathbb{P}(x < n) = \frac{n}{2^k}$. Comme $x = i \Rightarrow x < n$ alors

$$\mathbb{P}(x = i \mid x < n) = \frac{\frac{1}{2^k}}{\frac{n}{2^k}} = \frac{1}{n}$$

(2) Le temps d'exécution T a pour espérance : $\mathbb{E}(T) = k + \frac{2^k - n}{2^k}\mathbb{E}(T)$ soit

$$\mathbb{E}(T) = k \cdot \frac{2^k}{n}$$

Quand n est une puissance de 2, on a $T = k$ constant.

(3) Si T est borné, alors on peut supposer, quitte à réaliser davantage de tirages inutiles, que l'on réalise un nombre constant k de tirages. Ainsi, on a une fonction $f : \{0,1\}^k \to [1, n]$ telle que $\forall i \ \frac{|f^{-1}(\{i\})|}{2^k} = \frac{1}{n}$ ainsi $2^k = |f^{-1}(\{i\})| \cdot n$, ce qui impose $|f^{-1}(\{i\})|$ et n puissance de 2.

2 Entropie

Question 3 : Correction : On rappelle que $H(X) = -\sum_{i=1}^{n} p_i \log(p_i)$ où X est une v.a. à valeurs dans $\{e_1, \ldots, e_n\}$ et $p_i = \mathbb{P}(x = e_i)$.

En général, $H(f(X)) \leq H(X)$.

En effet, soit $y \in \mathrm{Img}(f)$ fixé et $p_y = \sum_{x|f(x)=y} p_x$. Alors $\forall x \ f(x) = y \Rightarrow p_x \log p_y \geq p_x \log p_x$. D'où en sommant :

$$H(Y) = -\sum_y p_y \log(p_y) = -\sum_y \Big(\sum_{x|f(x)=y} p_x \Big) \log(p_y) \quad \leq -\sum_y \sum_{x|f(x)=y} p_x \log(p_x) = H(X)$$

avec égalité ssi f est injective.

Question 4 : Correction : D'une part $H((p_i)_i) = \sum_i \underbrace{-p_i \log p_i}_{\geq 0} \geq 0$. La valeur 0 est atteinte pour tout vecteur $v = 1_{\{i\}}(.)$ (un seul 1 à la position i, le reste des 0). De plus, il s'agit de seules vecteurs possibles atteignant 0 car une somme de termes positifs est nulle si et seulement si tous ses termes sont nuls. Ainsi, on doit avoir $\forall i \ p_i \log p_i = 0$ c'est-à-dire $\forall i \ p_i \in \{0, 1\}$ ainsi que $\sum_i p_i = 1$.

Question 5 : Correction : (a) X prend des valeurs dans $\mathbb{N}^*$ et $\mathbb{P}(X = i) = p_i = \frac{1}{2^i}$, d'où $H(X) = -\sum_{i\geq 1} \frac{1}{2^i} \log \frac{1}{2^i} = \sum_{i\geq 1} \frac{i}{2^i}$. Or c'est le même calcul que l'espérance, donc $H(X) = 2H(X) - H(X) = \sum_{i\geq 1} \frac{i}{2^{i-1}} - \sum_{i\geq 1} \frac{i}{2^i} = \sum_{i\geq 0} \frac{1}{2^i} = 2$.

(b) Les requêtes optimales sont $S_i = [1, i]$ alors

$$\mathbb{P}(X \in S_i \mid X \notin S_{i-1}) = \frac{\mathbb{P}(X \in S_i \wedge X \notin S_{i-1})}{\mathbb{P}(X \notin S_{i-1})}$$

mais

$$\mathbb{P}(X \in S_i) = \sum_{j=1}^{i-1} \frac{1}{2^j} = \sum_{j=0}^{i-1} \frac{1}{2^j} - 1$$
$$= \frac{1 - \frac{1}{2}^i}{1 - \frac{1}{2}} - 1 = 1 - \frac{1}{2^{i-1}}$$

d'où

$$\mathbb{P}(X \in S_i \mid X \notin S_{i-1}) = \frac{\frac{1}{2^i}}{1 - (1 - \frac{1}{2^{i-1}})} = \frac{1}{2}$$

L'espérance du nombre de requêtes nécessaires est $2 = H(X)$.

Question 6 : Correction : Soit $k(y)$ le nombre de valeurs de X de longueur y, c'est-à-dire le nombre de séries de matchs dont le vainqueur est déterminé à la fin du y-ième match.

Par définition, $X[y]$ désigne toujours le vainqueur, et $X[1, y-1]$ est une séquence de matchs où le vainqueur final a déjà gagné exactement 3 fois. Il y a donc 3 parmi $y-1$ telles séquences, si l'on fixe $X[y]$, qui lui peut prendre deux valeurs. Ainsi $k(y) = 2 \times \binom{y-1}{3} = \frac{(y-1)(y-2)(y-3)}{2}$.

Un match de longueur y a probabilité 2^{-y} de se produire. Ainsi :

$$H(X) = -\sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) \qquad \text{en regroupant en termes de même longueur}$$
$$= -\sum_{y=4}^{7} k(y) \frac{1}{2^y} \log \frac{1}{2^y} = \sum_{y=4}^{7} k(y) \frac{y}{2^y} = 93/16$$

Pour $H(Y)$, on fait le calcul, sachant que $\mathbb{P}(Y = y) = k(y)/2^y$.

Comme Y est une fonction de X, on a $H((X, Y)) = H(X)$, puis $H(Y|X) = H(X, Y) - H(X) = 0$ et $H(X|Y) = H((X, Y)) - H(Y) = H(X) - H(Y)$.

3 Entropie dans les arbres

Question 7 : Correction : On note $A_k \subseteq [1, N]$ les indices des noeuds internes à la profondeur k. Clairement, si $i \neq j \in A_k$, les noeuds i et j sont des évènements disjoints. Ainsi,

$$\mathbb{E}(L) = \sum_{k\geq 1} k\mathbb{P}(L = k) = \sum_{k\geq 1} \mathbb{P}(L \geq k) \sum_{k\geq 1} \sum_{i\in A_k} P_i = \sum_{i=1}^{N} P_i$$

Question 8 : Correction : Exprimons la loi de probabilité de la v.a. $(B[i] = j \mid W[i] = l)$:

$$\mathbb{P}(B[i] = j \mid W[i] = l) = \frac{\mathbb{P}(B[i] = j \wedge W[i] = l)}{\mathbb{P}(W[i] = l)} = \frac{q_{l,j}}{P_l}$$

Son entropie vaut ainsi

$$H_l = -\sum_{j=1}^{D} \frac{q_{l,j}}{P_l} \log\left(\frac{q_{l,j}}{P_l}\right)$$

Question 9 : Correction : On remarque $\sum_{j=1}^{D} q_{l,j} = P_l$ ainsi

$$P_l \cdot H_l = \sum_{j=1}^{D} q_{l,j} \left(\log P_l - \log q_{l,j}\right) = P_l \log P_l - \sum_{j=1}^{D} q_{l,j} \log q_{l,j}$$

On remarque en faisant $\sum_{l=1}^{N} P_l \cdot H_l$ que les termes $P_i \log P_i$ s'annulent pour $i > 1$, il reste alors plus que $P_1 \log P_1 - \sum_{i=1}^{n} p_i \log p_i$, c'est-à-dire $H(F)$.

Question 10 : Correction : Si $\mathbb{E}(L) = \infty$, le résultat est immédiat. On suppose désormais $\mathbb{E}(L) < \infty$. L'algorithme peut s'exprimer comme une fonction $f : \nu \to R$ où R est l'ensemble des résultats et ν est l'ensemble des séquences finies de tirages qui permettent de terminer l'algorithme. Ainsi, on a par définition $Y = f(X_1 \dots X_L)$ donc (cf question 1) $H(Y) \leq H((X_1 \dots X_L))$.

On fixe $k \in \mathbb{N}$. Considérons l'arbre des préfixes de ν de longueurs inférieures ou égales à k, noté $A_k = \mathrm{Pref}(\nu) \cap \Sigma^{\leq k}$ et on pose $L_k = \min(L, k)$. Ainsi, L_k correspond à la profondeur de la feuille $X_1 \dots X_{L_k}$ de l'arbre A_k. Les noeuds internes correspondent à des tirages de l'algorithme, qui ont tous la même entropie de branchement, $H_l = H(\mu)$. On a donc d'après l'exercice précédent $H((X_1 \dots X_{L_k})) = \sum_l P_l \cdot H_l = \mathbb{E}(L_k) \cdot H(\mu)$.

Or $\lim_{k\to\infty} \mathbb{E}(L_k) = \mathbb{E}(L)$ et $\lim_{k\to\infty} H(X_1 \dots X_{L_k}) = H(X_1 \dots X_L)$ (ces deux passages à la limite se prouvent avec l'hypothèse $\mathbb{E}(L) < \infty$).

NB : Le cas particulier où L est borné se prouve sans passage à la limite (il suffit de prendre k égal à une borne sur L). Un cas encore plus particulier consiste à supposer L constant, auquel cas on a $H((X_1 \dots X_L)) = L \cdot H(\mu)$ par indépendance des variables aléatoires.