TD 7. Algorithmes probabilistes

Exercice 1. *Échantillonnage par rejet*

On souhaite tirer un point aléatoire dans un disque $\mathcal{D}$ de centre $(0,0)$ et de rayon 1. Pour cela, on tire deux réels $x, y \in [-1,1]$ uniformément : si le point (x,y) appartient à $\mathcal{D}$, on le renvoie ; sinon on recommence.

1. **i.** Comment tester si $(x,y) \in \mathcal{D}$?
   ii. Quelle est la probabilité que $(x,y) \in \mathcal{D}$?
   iii. En déduire l'espérance du nombre de tirages nécessaires pour obtenir un point dans le disque.
2. On généralise l'algorithme : soit X un ensemble dans lequel on sait tirer un élément uniformément, et $S \subset X$; supposons qu'on sache tester, pour $x \in X$, si $x \in S$; enfin, on note $p = \#S/\#X$.
   i. Identifier X, S et p dans le cas du disque.
   ii. Décrire l'algorithme permettant de tirer un élément dans S.
   iii. Quelle est l'espérance du nombre d'éléments de X à tirer avant d'avoir un élément de S ?
3. Justifier que l'algorithme (général) renvoie un point *uniforme* dans S, c'est-à-dire que chaque point de S est renvoyé avec probabilité $1/\#S$.
4. Afin de tirer moins de réels dans le cas du disque, une personne astucieuse découple le tirage de x et y : elle tire d'abord x, puis tire des y tant que (x,y) n'est pas dans $\mathcal{D}$. Est-ce une bonne idée ?
5. On souhaite tirer un nombre premier uniformément parmi les premiers $\leq N$. On suppose disposer d'un algorithme déterministe de test de primalité. *On rappelle que le nombre de premiers $\leq N$ est $\pi(N) \geq N/\ln N$ (pour $N \geq 17$).* Calculer l'espérance du nombre d'entiers à tirer avant d'en obtenir un premier.

Exercice 2. *Preuve de l'algorithme de Freivalds*

L'algorithme de Freivalds prend en entrée $A, B, C \in \mathbb{K}^{n\times n}$ pour un certain corps $\mathbb{K}$. Il tire un vecteur $\vec{v}$ aléatoire uniforme dans $\{0,1\}^n$ et répond VRAI si $A \times B \times \vec{v} = C \times \vec{v}$, FAUX sinon.

1. Justifier que la complexité de l'algorithme est $O(n^2)$.
2. Montrer que si $C = A \times B$, alors l'algorithme répond toujours VRAI.
3. On suppose que $C \neq A \times B$ et on note $D = C - A \times B$.
   i. Justifier qu'il existe (i,j) tel que $D_{i,j} \neq 0$.
   ii. Montrer que si $A \times B \times \vec{v} = C \times \vec{v}$, alors $v_j = \frac{1}{D_{i,j}} \sum_{k\neq j} D_{i,k} v_k$
   iii. En déduire que $\Pr[A \times B \times \vec{v} = C \times \vec{v}] \leq \frac{1}{2}$. Que conclut-on ?
4. Quel type d'algorithme est l'algorithme de Freivalds ?

Exercice 3. QUICKSELECT

L'algorithme QUICKSELECT est une variante du tri rapide, pour résoudre le problème suivant : étant donné un tableau T de taille n et un entier $k \in \{1,\dots,n\}$, on souhaite trouver le $k^{\text{ème}}$ plus petit élément de T, noté $T^{(k)}$. *Par exemple, si $k = 1$, $T^{(1)}$ est le minimum de T.* Le principe de l'algorithme est le même que pour le tri rapide : choisir aléatoirement un pivot dans T ; calculer T_{INF} et T_{SUP} qui contiennent les éléments de T *strictement* inférieurs et *strictement* supérieurs à p, respectivement ; si nécessaire effectuer un appel récursif sur T_{INF} ou T_{SUP}. *On suppose tous les éléments distincts dans T.*

1. Soit n_{INF} et n_{SUP} les tailles de T_{INF} et T_{SUP}.
   i. Si $n_{\text{INF}} = k-1$, que vaut $T^{(k)}$?
   ii. Si $n_{\text{INF}} \geq k$, exprimer $T^{(k)}$ comme $T^{(\ell)}_{\text{INF}}$ pour un certain ℓ.
   iii. Si $n_{\text{INF}} < k-1$, exprimer $T^{(k)}$ comme $T^{(\ell)}_{\text{SUP}}$ pour un certain ℓ.
2. Écrire complètement l'algorithme. *Il n'est pas interdit d'être délicat et d'éviter de construire des tableaux inutiles.*
3. Combien de comparaisons sont effectuées, si tous les choix probabilistes sont mauvais ?
4. On s'intéresse maintenant l'espérance du nombre C_n de comparaisons effectuées pour un tableau de taille n.

i. Montrer que $\mathbb{E}[C_n] = \sum_{j=1}^{n} \mathbb{E}[C_n | p = T^{(j)}] \Pr[p = T^{(j)}]$, où p est le pivot choisi.
ii. Montrer que si $p = T^{(j)}$, l'algorithme effectue un appel récursif sur un tableau de taille $t \leq \max(j-1, n-j)$.
iii. En déduire que $\mathbb{E}[C_n | p = T^{(j)}] \leq n + E_m$ où $m = \max(j-1, n-j)$.
iv. Montrer par récurrence sur n que $\mathbb{E}[C_n] \leq 4n$.

Exercice 4. *Compteur probabiliste*

Un compteur entier qui va de 0 à n utilise un espace mémoire $O(\log n)$. On va voir une technique probabiliste qui permet de fournir un compteur approximatif qui utilise un espace exponentiellement plus petit. L'idée est d'avoir une fonction INCRÉMENT qui soit probabiliste.

1. On définit une fonction INCRÉMENT(c) qui incrémente c avec probabilité p et qui ne fait rien avec probabilité $1-p$. On initialise c à 0 et on effectue n appels à INCRÉMENT.

i. Quelle est l'espérance de la valeur de c ?
ii. Quelle valeur faut-il renvoyer pour avoir une valeur approchée de n ?
iii. Borner la probabilité que la valeur renvoyée soit $\geq 2n$.

On modifie la fonction INCRÉMENT, pour que INCRÉMENT(c) incrémente c avec probabilité $1/2^c$ et ne fasse rien avec probabilité $1 - 1/2^c$.

2. **i.** On suppose avoir accès à une fonction BITALÉATOIRE() qui renvoie 0 avec probabilité $\frac{1}{2}$ et 1 avec probabilité $\frac{1}{2}$. Écrire explicitement la fonction INCRÉMENT.
ii. Quelle est l'espérance du nombre d'appels effectués à BITALÉATOIRE ?

On note C_k la variable aléatoire qui décrit la valeur du compteur après k appels à INCRÉMENT, et $p_{k,c} = \Pr[C_k = c]$ la probabilité que le compteur ait la valeur c après k appels à INCRÉMENT.

3. **i.** Calculer $p_{0,0}$, $p_{0,c}$ pour $c > 0$ et $p_{k,0}$ pour $k > 0$.
ii. Calculer $\sum_{c \geq 0} p_{k,c}$.
iii. Montrer que pour $k, c > 0$, $p_{k,c} = \frac{1}{2^{c-1}} p_{k-1,c-1} + (1 - \frac{1}{2^c}) p_{k-1,c}$. *Indication. Si le compteur vaut c après k appels à* INCRÉMENT*, quelle peut être sa valeur après $k-1$ appels ?*

Finalement, on décide de renvoyer la valeur $v = 2^c$. On cherche à calculer l'espérance de la valeur finale de v après n appels à INCRÉMENT. On note $V_k = 2^{C_k}$ la variable aléatoire qui décrit la valeur v renvoyée par l'algorithme.

4. **i.** Exprimer $\mathbb{E}[V_k]$ en fonction de $p_{k,c}$, pour k, $c \geq 0$.
ii. En déduire que $\mathbb{E}[V_k] = \mathbb{E}[V_{k-1}] + 1$.
iii. En déduire que $\mathbb{E}[V_n] = n$.