Université de Caen Normandie, UFR des Sciences
Licence 3

Module " Algorithmique, structures informatiques et cryptologie"
Algorithmes probabilistes ; durée : 2h30

Exercice 1. Loi de Bernoulli avec une pièce de monnaie

On dispose d'une pièce de monnaie non biaisée, telle que

$$\mathbb{P}(Pile) = \mathbb{P}(Face) = 1/2.$$

On veut réaliser une loi de Bernoulli de paramètre $p = 7/8$ en effectuant un minimum de lancers de cette pièce.

1. Montrez qu'en lançant 3 fois la pièce on peut réaliser l'événement demandé de probabilité $7/8$. Précisez l'événement que vous choisissez.
2. Vous semble-t-il nécessaire d'effectuer toujours exactement 3 lancers ? En examinant "l'arbre des possibilités", précisez dans quels cas précis on doit faire 1, 2 ou 3 lancers.
3. On note X la variable aléatoire donnant le nombre de lancers de la pièce. Calculez la moyenne (espérance) du nombre de lancers
4. Mêmes questions avec $5/8$.

Indication : Notez les égalités $7/8 = 1/2 + 1/4 + 1/8$ et $5/8 = 1/2 + 1/8$.

Exercice 2. Tirages de boules dans une urne avec remises

On tire au hasard une boule dans une urne qui contient n boules bleues, n boules jaunes, n boules rouges et n boules vertes, puis on remet la boule dans l'urne. Si la boule n'est pas rouge, on répète l'expérience jusqu'à obtenir une boule rouge.

1. Quelle est la loi de probabilité vérifiée par X ? Justifiez.
2. Donnez l'espérance $E(X)$, la variance $V(X)$ et l'écart-type $\sigma(X)$ de la variable aléatoire X.

Exercice 3. Tirages de boules dans une urne avec remises (bis)

On tire au hasard une boule dans une urne qui contient n boules numérotées $1, 2, \ldots, n$. On note le numéro de la boule, puis on remet la boule dans l'urne et on répète l'expérience jusqu'à obtenir au moins une fois chacune des n boules $1, 2, \ldots, n$.

Soit X la variable aléatoire qui représente le nombre de tirages nécessaires pour obtenir les n numéros de boules.

1. Quelle expérience aléatoire vue en cours cela vous rappelle t'il ?
2. Donnez l'espérance $E(X)$, la variance $V(X)$ et l'écart-type $\sigma(X)$ de la variable aléatoire X.
3. Donnez une valeur approchée de $E(X)$ pour $n = 100$.

Exercice 4. Bornes de Chernoff et loi binomiale

Nous avons vu en cours un algorithme probabiliste pour estimer π. Nous allons affiner les informations données en cours.

Soit X et Y deux variables aléatoires uniformes sur l'intervalle $[-1, 1]$. Soit Z la variable aléatoire qui vaut 1 si $X^2 + Y^2 \leq 1$ et 0 sinon.

1. Quelle est la loi de Z ? Donner son espérance, sa variance et son écart-type.

On répète l'expérience n fois et on note $Z_1, \ldots, Z_n$ le résultat des n expériences.

2. Quelle est la loi de $T_n = Z_1 + Z_2 + \cdots + Z_n$? Donner son espérance, sa variance et son écart-type.
3. Vers quelle valeur tend (presque-sûrement) T_n/n ? Justifiez.
4. En utilisant l'inégalité de Bienaymé-Tchebychev, quelle valeur de n faut-il choisir pour avoir une estimation de $\pi/4$ à 10^{-k} près avec une certitude de 95% ? Autrement dit, quelle valeur de n faut-il pour avoir
$$\mathbb{P}\left(\left|\frac{T_n}{n} - \frac{\pi}{4}\right| < 10^{-k}\right) > 0.95 \text{ ?}$$

En fait, il existe une borne plus fine que l'inégalité de Bienaymé-Tchebychev dans ce contexte.

Proposition (Bornes de Chernoff). *Soient $Z_1, Z_2, \ldots, Z_n$ des variables aléatoires à valeurs dans $\{0, 1\}$, indépendantes, de même espérance p. Alors pour tout $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - p\right| > \varepsilon\right) \leq 2e^{-2\varepsilon^2 n}.$$

5. En utilisant la borne de Chernoff, quelle valeur de n faut-il choisir pour avoir une estimation de $\pi/4$ à 10^{-k} près avec une certitude de $c \in]0, 1[$?
6. Pour $c = 0.95$, quel est le gain de complexité par rapport à l'inégalité de Bienaymé-Tchebychev ?

Exercice 5. **Sélection de la médiane**

Soit un ensemble S de n entiers présenté sous forme d'un tableau $S[1..n]$ *non trié* dont les n éléments (entiers) sont supposés tous distincts pour simplifier. On appelle *médiane* de S l'élément m de S tel que le nombre d'éléments de S qui sont strictement plus grands que m est exactement $\lfloor n/2 \rfloor$ (où $\lfloor x \rfloor$ désigne la partie entière de x) ; autrement dit, m est le k-ième élément de S si $n = 2k$ (ce qu'on notera $rang_S(m) = k$) et le $k+1$-ième si $n = 2k+1$. Cela revient aussi à dire que $rang_S(m) = \lceil \frac{n}{2} \rceil$ (où $\lceil x \rceil$ désigne le plus petit entier supérieur ou égal à x).

Nous supposons que chaque ensemble est représenté par un tableau. Ainsi, l'accès à un élément de l'ensemble se fait en tant constant $O(1)$, par son indice dans le tableau, quand on connaît cet indice.

Voici l'algorithme (appelé aussi procédure) pour le calcul de la médiane.

Entrées: Un ensemble S constitué de n éléments
Sorties: la médiane de S ou ECHEC

1. Soit R un ensemble de $\lfloor n^{3/4} \rfloor$ éléments de S tirés aléatoirement et uniformément avec remise (possibilité d'avoir des doublons)
2. Trier R et trouver a et b tels que

$$rang_R(a) = \max\left(1, \lfloor \frac{n^{3/4}}{2} - \sqrt{n} \rfloor\right), \qquad rang_R(b) = \min\left(\lfloor \frac{n^{3/4}}{2} + \sqrt{n} \rfloor, \lfloor n^{3/4} \rfloor\right)$$

où $rang_R(a) = t$ si a est le t-ième élément le plus petit de R.
3. En comparant chaque élément de S avec a et b, calculer en un seul parcours de S :

$$rang_S(a), \quad rang_S(b), \quad P = \{x \in S,\, a \leq x \leq b\} \quad \text{et} \quad \text{Card(P)}.$$

4. Si $rang_S(a) > \lceil \frac{n}{2} \rceil$ **retourner** ECHEC
5. Si $rang_S(b) < \lceil \frac{n}{2} \rceil$ **retourner** ECHEC
6. Si $\text{Card}(P) \geq 4n^{3/4}$ **retourner** ECHEC
7. Trier P

retourner 8. c tel que $rang_P(c) = \lceil \frac{n}{2} \rceil - rang_S(a) + 1$

Question 0. (correction de l'algorithme) Montrer que les inégalités suivantes sont vraies à l'étape 7 :

$$\text{Card}(P) = rang_S(b) - rang_S(a) + 1 < 4n^{3/4} \qquad \text{et} \qquad rang_S(a) \leq \lceil \frac{n}{2} \rceil \leq rang_S(b).$$

En déduire que la valeur retournée à l'étape 8 est la médiane de S.

Nous nous intéressons maintenant à la complexité de l'algorithme.

Question 1. Sachant qu'un tri de k éléments peut se faire en $O(k \ln k)$ comparaisons dans le pire des cas, quelle est dans le pire des cas le nombre de comparaisons nécessaires pour réaliser les étapes 2 et 7 de l'algorithme (en fonction de n) ?

Question 2. Quelle est la complexité dans le pire des cas de l'étape 1 ? Même question avec l'étape 3. En déduire la complexité dans le pire des cas de la procédure complète.

Question 3. Nous notons p (qui dépend de la taille de S) la probabilité que la procédure échoue. En répétant au plus 42 fois la procédure, quelle est la probabilité d'obtenir 42 échecs ?

Question 4. Donner un algorithme probabiliste, basé sur cette procédure, qui quand il s'arrête, retourne toujours le bon résultat. Quelle est sa complexité moyenne en fonction de p et n ?

Les questions suivantes visent à mesurer la probabilité p d'échec de l'algorithme et à montrer que $p = O(n^{-1/4})$ avec n le nombre d'éléments de S. Dans la suite, pour simplifier les calculs, on supposera que l'entier n est pair, donc $\lceil \frac{n}{2} \rceil = \frac{n}{2}$. Remarquez dans ce cas qu'exactement la moitié des éléments de S sont inférieurs ou égaux à la médiane.

Question 5. On note X_i la variable aléatoire égale à 1 si le i-ième élément mis dans R lors de sa construction (et choisi au hasard dans S) est inférieur ou égal à la médiane m de S, et égale à 0 sinon.
Quelle loi de probabilité les variables aléatoires X_i suivent-elles ? Donnez l'espérance et la variance des X_i.

Question 6. Le nombre d'éléments de R inférieurs ou égaux à la médiane m de S est donné par la

formule

$$X = \sum_{i=1}^{\lfloor n^{3/4} \rfloor} X_i.$$

En remarquant que les variables aléatoires X_i sont mutuellement indépendantes, donnez la loi de probabilité de X ? Donnez l'espérance et la variance de X (on pourra omettre les parties entières $\lfloor \cdot \rfloor$ pour simplifier les calculs).

Question 7. Montrez que si $rang_S(a) > n/2$, alors $X \leq \frac{n^{3/4}}{2} - \sqrt{n}$.
(On omet une fois de plus les parties entières pour simplifier les calculs).
Indication : Constater l'inégalité suivante : le nombre d'éléments de R qui sont $\leq m$ est majoré par le nombre d'éléments de R qui sont $\leq a$, lui-même égal à $\frac{n^{3/4}}{2} - \sqrt{n}$, par définition de a.

Question 8. Rappelons que l'inégalité de Chebyshev est donnée par :

$$\Pr\left[|X - E(X)| \geq v\right] \leq \frac{var(X)}{v^2}$$

avec $E(X)$ et $var(X)$ l'espérance et la variance de X. En appliquant l'inégalité de Chebyshev, déduire l'inégalité suivante :

$$\Pr\left[rang_S(a) > \frac{n}{2}\right] \leq \Pr\left[X \leq \frac{n^{3/4}}{2} - \sqrt{n}\right] = O(n^{-1/4})$$

(on omet une fois de plus les parties entières pour simplifier les calculs).

En appliquant la même méthode, il est également possible de borner la probabilité des autres cas d'échec. Nous admettrons les bornes suivantes :

$$\Pr\left[rang_S(b) < \lfloor \frac{n}{2} \rfloor\right] = O(n^{-1/4}) \qquad \text{et} \qquad \Pr\left[\text{Card}(P) \geq 4n^{3/4}\right] = O(n^{-1/4}).$$

Question 9. Quelle est la probabilité p d'échec de la procédure. En déduire la complexité de l'algorithme décrit à la question 4 ?

Exercice 6 : Coloriage de graphes

Nous notons K_n le graphe complet à n sommets. Pour un sous-ensemble de sommets A de K_n, nous notons K_A le sous-graphe (complet) de K_n restreint aux sommets de A.

1. Trouver une coloration des arêtes du graphe K_5 en bleu et rouge telle qu'il n'y ait pas de triangle rouge ou bleu.
2. Montrez qu'une telle coloration est impossible pour K_6.

On se fixe un réel $p \in [0, 1]$. Pour chaque arête du graphe K_n, celle-ci est colorée en rouge avec une probabilité p et en bleu avec une probabilité $1 - p$, de manière indépendante des autres sommets.

3. Soit k le cardinal de A. Quelle est la probabilité que K_A soit entièrement rouge. Même question avec la couleur bleue.
4. Soit $k \leq n$. Donner une majoration de la probabilité de l'évènement suivant : "Il existe un sous-graphe complet à k sommets monochromatique".
5. Soit $k \geq 4$ et $n \leq 2^{k/2}$. Trouver une valeur de p pour laquelle l'évènement précédent a une probabilité strictement plus petite que 1.
6. En déduire l'existence d'un coloriage de K_n sans sous-graphe à k sommets monochromatique.
7. Proposer un algorithme probabiliste pour construire une telle coloration.
8. Estimer la complexité de cet algorithme.
9. Comparer cette complexité à celle de l'algorithme naïf qui passe en revue toutes les colorations possibles.