Exercices sur les algorithmes probabilités
Nom du rédacteur : Johanne Cohen
`https://www.lri.fr/~jcohen/pages/enseignement.html`

Probabilité : analyse probabiliste

Nous utiliserons une fonction $random(k)$ qui tire un entier entre 1 et k de façon uniforme.

Sélectionner une partie d'un ensemble S de n éléments distincts.

Nous considérons un ensemble $S = \{s_1, s_2, \cdots, s_n\}$ de n éléments distincts.

Un premier algorithme : Considérons l'algorithme $\mathcal{B}$ suivant :
Entrée : Un ensemble $S = \{s_1, s_2, \cdots, s_n\}$ de n éléments et un entier k avec $k \leq n$.
Sortie : une partie C de S.

1. $C \leftarrow \emptyset$
2. pour i allant 1 à n faire
   (a) $Z_i \leftarrow random\_Bernoulli(\frac{k}{n})$
   (b) Si $Z_i = 1$ alors $C \leftarrow C \cup \{s_i\}$
3. retourner C

Question 1. Soit Y la variable aléatoire désignant la taille de C.

1. Exprimer Y en fonction des variables aléatoires Z_i.
2. La loi Y est-elle une loi de Bernoulli ? une loi géométrique ? une loi binomiale ?
3. Calculer $E[Y]$.

   1. $Y = \sum_{i=1}^{n} Z_i$. L'algorithme insère un élément dès que $Z_i = 1$. Donc le nombre d'insertion correspond à la somme des valeurs de Z_i.
   2. Y mesure le nombre de succès lors qu'il y a n tirages de la loi de Bernoulli. Donc c'est une loi binomiale.
   3. $E[Y] = E[\sum_{i=1}^{n} Z_i] = \sum_{i=1}^{n} E[Z_i] = n\frac{k}{n} = k$.

Question 2. L'algorithme $\mathcal{B}$ retourne-t-il toujours une partie de S de cardinalité k ?

Non car la probabilité que $Y = 1$ est non nulle (par exemple).

Un second algorithme. Considérons l'algorithme $\mathcal{D}$ suivant :
Entrée : Un ensemble $S = \{s_1, s_2, \cdots, s_n\}$ de n éléments.
Sortie : une partie X de S.

1. $X \leftarrow \emptyset$
2. pour i allant 1 à n faire
   (a) $Z_i \leftarrow Bernoulli(\frac{1}{2})$
   (b) Si $Z_i = 1$ alors $X \leftarrow X \cup \{s_i\}$
3. retourner X

Question 3. Montrer que $Pr(X = \{s_1\}) = Pr(X = \emptyset) = \frac{1}{2}$ à la fin de la première itération de la boucle (c'est-à-dire lorsque $i = 1$)

Il suffit de remarquer que l'évènement "X vaut $\{s_1\}$ à la fin de la première itération de la boucle“" se réalise si l'évènement $Z_1 = 1$ se réalise.

$$Pr(X = \{s_1\}) = \frac{1}{2}$$

En utilisant le même argument : l'évènement "X vaut $\emptyset$ à la fin de la première itération de la boucle“" se réalise si l'évènement $Z_1 = 0$ se réalise.

$$Pr(X = \emptyset) = \frac{1}{2}$$

Nous noterons par $\mathcal{P}(\{s_1, \ldots, s_i\})$ l'ensemble des parties de l'ensemble $\{s_1, \ldots, s_i\}$. Soit C une partie du sous-ensemble $\{s_1, \ldots, s_i\}$ de i éléments (avec $i \leq k$).

Question 4. Montrer que

1. $\mathcal{P}(\{s_1, \ldots, s_{i-1}\}) \subset \mathcal{P}(\{s_1, \ldots, s_i\})$
2. si $C \in \mathcal{P}(\{s_1, \ldots, s_i\}) \backslash \mathcal{P}(\{s_1, \ldots, s_{i-1}\})$, alors C contient s_i
3. le nombre de parties de l'ensemble $\{s_1, \ldots, s_i\}$ contenant l'élément s_i est égale à $|\mathcal{P}(\{s_1, \ldots, s_{i-1}\})|$

   1. Si A est une partie de $\{s_1, \ldots, s_{i-1}\}$ alors elle est aussi une partie de $\{s_1, \ldots, _i\}$. Donc
   $$\mathcal{P}(\{s_1, \ldots, s_{i-1}\}) \subset \mathcal{P}(\{s_1, \ldots, s_i\})$$
   2. Prouvons le point 2 par contradiction. Supposons que $C \in \mathcal{P}(\{s_1, \ldots, s_i\}) \backslash \mathcal{P}(\{s_1, \ldots, s_{i-1}\})$. Supposons que C ne contient pas de s_i. cela signifie que $C \subset \{s_1, \ldots, s_{i-1}\}$, Donc $C \in \mathcal{P}(\{s_1, \ldots, s_{i-1}\})$. Ce qui est en contradiction avec le fait que $C \in \mathcal{P}(\{s_1, \ldots, s_i\}) \backslash \mathcal{P}(\{s_1, \ldots, s_{i-1}\})$
   3. Notons $\mathcal{V}$ l'ensemble de parties de l'ensemble $\{s_1, \ldots, s_i\}$ contenant l'élément s_i. Nous allons construire une fonction $f : \mathcal{V} \to \mathcal{P}(\{s_1, \ldots, s_{i-1}\})$ de la façon suivante : $f(C) = C'$ si $C' = C \backslash \{s_i\}$. Il suffit de prouver que f est une bijection,

Question 5. Soit C une partie du sous-ensemble $\{s_1, \ldots, s_i\}$ de i éléments (avec $i \leq k$).

Montrer (par récurrence) que $Pr(X = C) = \frac{1}{|\mathcal{P}(\{s_1,\ldots,s_i\})|}$ à la fin de la $i$ème itération de la boucle.

Il suffit de montrer par récurrence sur le nombre d'itérations. Notons par X_i la variable aléatoire correspondant à la partie de S à la fin de $i$ème itération. Nous allons prouver que X_i est une partie après la $i$ème itération générée de façon uniforme : c'est-à-dire $Pr(X_i = C_i) = \frac{i!(n-i)!}{n!}$ en considérant C_i une partie de S de cardinalité i.

Si $i = 1$, alors $Pr(X_1 = \{s_1\}) = \frac{1}{2}$ (d'après les questions précédentes).

Supposons que pour i, l'algorithme $\mathcal{B}$ génère une partie X_i à la fin de $i$ème itération générée de façon uniforme. Prouvons que cela est vrai aussi pour $i+1$. Pour cela, il faut prouver que

$$Pr(X_{i+1} = C_{i+1}) = \frac{1}{\binom{n}{i+1}} \text{ avec } C_{i+1} \in P(\{s_1, \ldots, s_{i+1}\}).$$

Notons les évènements suivants

- $\mathcal{E}(s_{i+1} \in X_{i+1})$ signifie que s_{i+1} est dans X après la $i+1$ itération de la boucle
- $\mathcal{E}(C \in X_{i+1})$ signifie que C est un sous ensemble de X après la $i+1$ itération de la boucle

Supposons que $s_{i+1} \in C_{i+1}$. Pour que l'évènement $X_{i+1} = C_{i+1}$ se réalise, il faut $\mathcal{E}(s_{i+1} \in X_{i+1})$ et $\mathcal{E}(C_{i+1} \backslash \{s_{i+1}\} \in X_{i+1})$. Comme ce sont deux évènements indépendants, nous avons

$$Pr(X_{i+1} = C) = \Pr\left(\mathcal{E}(s_{i+1} \in X_{i+1})\right) \cdot \Pr\left(\mathcal{E}(C_{i+1} \backslash \{s_{i+1}\} \in X_{i+1}\right)$$

Rappellons que $\Pr\left(\mathcal{E}(s_{i+1} \in X_{i+1})\right) = Pr(Z_{i+1} = 1) = 1/2$ et que par hypothèse de récurrence $\Pr\left(\mathcal{E}(C_{i+1} \backslash \{s_{i+1}\} \in X_i\right) = \Pr\left(\mathcal{E}(C_{i+1} \backslash \{s_{i+1}\} \in X_{i+1}\right) = \frac{1}{|\mathcal{P}(\{s_1,\ldots,s_i\})|}$

$$\begin{aligned} Pr(X_{i+1} = C) &= \frac{1}{2} \cdot \frac{1}{|\mathcal{P}(\{s_1, \ldots, s_i\})|} \\ &= \frac{1}{|\mathcal{P}(\{s_1, \ldots, s_{i+1}\})|} \end{aligned}$$

Même raisonnement pour le cas où $s_{i+1} \notin C_{i+1}$.

Question 6. Que déduisez-vous sur l'algorithme $\mathcal{D}$.

L'algorithme $\mathcal{D}$ génère une partie de S de façon uniformément

Tableaux de n éléments contenant un ensemble S de n entiers distincts.

Nous considérons que le tableau T de n éléments contenant tous les entiers de S est généré aléatoirement de façon uniforme. Considérons l'algorithme $\mathcal{A}$ suivant :
Entrée : un tableau T de n éléments contenant les entiers de S, un entier ℓ.
Sortie : un entier i

1. $i \leftarrow 1$;
2. Tant que $[(T[i] \neq \ell)$ et $(i < n+1)]$ faire $i \leftarrow i+1$;
3. Si $(i \leq n)$ alors renvoyer i sinon renvoyer -1

Question 7. Dire ce que l'algorithme $\mathcal{A}$ retourne.

L'algorithme retourne -1 si ℓ n'est pas dans T sinon il retourne sa position.

Question 8. Donner le nombre de comparaisons que l'algorithme $\mathcal{A}$ effectue dans le pire des cas.

Pour vérifier la condition du **tant que**, il faut faire 2 comparaisons. Pour que l'algorithme fasse le plus d"opérations, il faut faire le plus d'itérations dans la boucle **tant que**. Pour cela, il faut que l'entier ℓ ne soit pas contenu dans le tableau. Si l'entier ℓ ne soit pas contenu dans le tableau, il y a $n+1$ tests pour la boucle **tant que**. Ensuite, il y a une comparaison supplémentaire due à la dernière instruction.

En tout, il est nécessaire de faire $2n+3$ opérations.

Question 9. Donner le nombre moyen d'opérations de l'algorithme en supposant les faits 2 suivants :

1. l'entrée de l'algorithme est générée aléatoirement de façon uniforme.
2. ℓ est dans T.

Le nombre moyen d'opérations de l'algorithme dépend la position moyenne de l'élément ℓ dans T. Soit Y la variable aléatoire correspondant à la position de l'élément ℓ dans T : $Y = i$ signifie que $T[i] = \ell$. Comme le tableau est généré de façon uniforme, et que le tableau contient l'élément ℓ, $Pr(Y = i) = \frac{1}{n}$.
Calculons le nombre d'opérations nécessaires à faire lorsque $T[i] = \ell$. En appliquant le même raisonnement que la question précédente, l'algorithme nécessite $2i + 1$ comparaisons lorsque $T[i] = \ell$.

$$E[Y] = \sum_{i=1}^{n}(2i+1)Pr(Y=i) = \sum_{i=1}^{n}(2i)Pr(Y=i) + \sum_{i=1}^{n} 1Pr(Y=i) = \frac{2}{n}\frac{n(n+1)}{2} + 1$$

Donc, le nombre moyen d'opérations est $E[Y] = (n+1)+1$

Question 10. Donner le nombre moyen de comparaisons de l'algorithme en supposant les faits 2 suivants :

- l'entrée de l'algorithme est générée aléatoirement de façon uniforme.
- ℓ est dans T avec une probabilité $1/2$.
  - Si ℓ n'est pas dans T, l'algorithme doit faire $2n+3$ opérations.
  $$E[Y|\ell \text{ n'est pas dans } T] = 2n+3$$
  - Si ℓ est dans T, le nombre moyen d'opérations de l'algorithme est $n+2$.
  $$E[Y|\ell \text{ est dans } T] = n+2$$

  Donc, notons A, l'événement correspondant au fait que ℓ est dans T.
  $$E[Y] = Pr(non\ A) \cdot E[Y|non\ A] + Pr(A) \cdot E[Y|A]$$
  Le nombre moyen d'opérations est $\frac{1}{2}(2n+3) + \frac{1}{2}(n+2) = \frac{3n+5}{2}$.

Question 11. Donner le nombre moyen de comparaisons de l'algorithme en supposant les faits 2 suivants :

- l'entrée de l'algorithme est générée aléatoirement de façon uniforme.
- ℓ est dans T avec une probabilité x.
  - Si ℓ n'est pas dans T, l'algorithme doit faire $2n+3$ opérations.
  $$E[Y|\ell \text{ n'est pas dans } T] = 2n+3$$
  - Si ℓ est dans T, le nombre moyen d'opérations de l'algorithme est $n+2$.
  $$E[Y|\ell \text{ est dans } T] = n+2$$

  Donc, notons A, l'événement correspondant au fait que ℓ est dans T.
  $$E[Y] = Pr(non\ A) \cdot E[Y|non\ A] + Pr(A) \cdot E[Y|A]$$
  Le nombre moyen d'opérations est $(1-x)(2n+3) + x(n+2) = n(2-x) + \mathcal{O}(1)$.