Algorithmes On-lines

Sélectionner k éléments d'un ensemble S de n éléments distincts.

L'objectif de cet exercice est de concevoir un algorithme $\mathcal{A}$ qui sélectionne k éléments parmi un ensemble $S = \{s_1, s_2, \cdots, s_n\}$ de n éléments distincts de façon uniforme. Nous noterons par $\mathcal{P}_k(S)$ l'ensemble des parties de S ayant exactement k éléments. Nous supposerons connu que $C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ correspond aux nombres de parties de k éléments d'un ensemble de n éléments distincts. En d'autres termes $|\mathcal{P}_k(S)| = \binom{n}{k}$.

Soit $\mathcal{A}$ un algorithme qui retourne une partie de k éléments parmi S de façon uniforme. Nous noterons X pour la variable aléatoire qui désigne la partie de S retournée par $\mathcal{A}$.

Question 1. Soit C une partie de S de k éléments. Que vaut la probabilité $Pr(X = C)$?

Question 2. Soit s un élément de S. Notons par $s \in X$ l'événement correspondant au fait que s est un élément de X, partie retournée par l'algorithme $\mathcal{A}$. Calculer les probabilités suivantes : $Pr(s \in X | X = C)$, $Pr(s \in X)$.

Nous utiliserons une fonction $Bernoulli(p)$ qui retourne 1 avec probabilité p sinon retourne la valeur 0.

Considérons l'algorithme $\mathcal{D}$:

Entrée : Un ensemble $S = \{s_1, s_2, \cdots, s_n\}$ de n éléments distincts.
Sortie : une partie X ayant k éléments de S.

1. $X \leftarrow \emptyset$;
2. pour i allant 1 à n faire
   (a) Si $i \leq k$, alors $X \leftarrow X \cup \{s_i\}$;
   (b) sinon
      i. $Z_i \leftarrow Bernoulli(\frac{k}{i})$;
      ii. Si $Z_i = 1$ alors
         – Choisir un élément ℓ de X de façon uniforme ;
         – $X \leftarrow X \backslash \{\ell\}$;
         – $X \leftarrow X \cup \{s_i\}$;
3. retourner X

Notons par $\mathcal{E}(e, i)$ l'évènement que l'élément e soit dans l'ensemble désigné par la variable X à la fin de l'itération i. L'objectif est de calculer la probabilité que l'élément e soit dans X à la fin de l'itération i, en d'autres termes, de calculer la probabilité $\Pr(\mathcal{E}(e, i))$.

Question 3. Quelle est la probabilité $(\Pr(\mathcal{E}(e, k)))$ que l'élément e soit dans X à la fin de l'itération k (c'est-à-dire lorsque i vaut k dans l'algorithme) ?

Question 4. Soit s_i le $i$ème élément de S avec $i > k$.

1. Donner les probabilités suivantes : $\Pr(Z_i = 1)$ et $\Pr(Z_i = 0)$.
2. Quelle est la probabilité que s_i soit dans X à la fin de l'itération i ? (Calculer $\Pr(\mathcal{E}(s_i, i))$)

Question 5. Soit e un élément dans X à la fin de l'itération k. Quelle est la probabilité $(\Pr(\mathcal{E}(e, k+1)|\mathcal{E}(e, k)))$ qu'il reste dans X à la fin de l'itération $k+1$? Pour cela, calculer cette probabilité dans les deux cas suivants :

1. sachant que $(Z_{k+1} = 0)$
2. sachant que $(Z_{k+1} = 1)$

Question 6. Montrer que la probabilité que l'élément e soit dans X à la fin de l'itération $k+1$ vaut $\frac{k}{k+1}$?

Nous allons prouver ***par récurrence*** grâce aux prochaines questions que la probabilité qu'un élément e dans $\{s_1, \cdots, s_i\}$ soit dans X à la fin de l'itération i vaut $\frac{k}{i}$: autrement dit, $\Pr(\mathcal{E}(e, i)) = \frac{k}{i}$. La question 6 montre que l'hypothèse de récurrence est vrai pour $i = k+1$. Supposons que l'hypothèse de récurrence est vraie pour pour i. Par la suite nous allons prouver qu'elle est vraie pour $i+1$.

Question 7. Soit e un élément dans X à la fin de l'itération i. Quelle est la probabilité qu'il reste dans X à la fin de l'itération $i+1$? Pour cela, calculer cette probabilité dans les deux cas suivants :

1. sachant que $(Z_{i+1} = 0)$
2. sachant que $(Z_{i+1} = 1)$

Question 8. Quelle est la probabilité que l'élément e soit dans X à la fin de l'itération $i+1$?

Question 9. Quelle est la probabilité que l'élément e soit dans X à la fin de l'itération n ?

Question 10. Peut-on adapter l'algorithme $\mathcal{D}$ en un algorithme on-line ? Justifier votre réponse (si c'est oui, donner l'algorithme, si c'est non, donnez vos raisons) ?