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

Algorithmes probabilistes : recrutement d'un collaborateur

Embauche d'un collaborateur.

Cet ennoncé de cet exercice se base sur le livre *Introduction to Algorithms* de T. Cormen, C. Leiserson, R. Rivest, and C. Stein à the MIT Press, 2 edition, (2001).

Une société veut recruter un nouvel collaborateur. Il y a n candidats possibles. Elle interviewe un candidat par jour et elle décide ensuite de l'embaucher ou non.

Question 1. La société utilise la procedure suivante :
Entrée : une liste de n candidats
Sortie : un candidat

1. La société génère une liste de passage des n candidats de façon uniforme.
2. La société interviewe le candidat 1 et elle l'embauche
3. pour tous les candidats v faire
   (a) La société interviewe le candidat v
   (b) si le candidat v est meilleur que celui embauché alors
      i. La société renvoie le candidat embauché
      ii. elle embauche le candidat v

Il y a un coût pour interviewer (coût c_i) et pour embaucher (coût c_h).

Question 1.1 Donner le coût de la procédure au pire des cas, au meilleur des cas.

Tous les candidats sont interviewés. Le coût pour interviewer pour la société est $n \times c_i$. Le coût liés aux embauches successifs de candidat dans cette procédure dépend du rang meilleur candidat.

Considérons le meilleur des cas. Le meilleur candidat est le premier de la liste. La société embauche uniquement un candidat. Donc, le coût pour embaucher pour la société est c_h.

Le coût total de cette procédure est $n \times c_i + c_h$.

Considérons le pire des cas. Il faut construire une instance où à chaque interview, la société embauche. Pour cela, la listes candidats correspond à une liste triée croissante en fonction de la valeur des candidats.

Le coût total de cette procédure est $n \times (c_i + c_h)$.

Maintenant nous allons évaluer le coût moyen de la procédure.

Question 1.2 Notons X_i la variable aléatoire indicatrice de l'événement correspondant aux fait que le candidat i soit embauché durant le processus.

1. Donner la définition formelle.

$$\begin{aligned} X_i &= 1_{\text{le candidat } i \text{ est embauché}} \\ &= \begin{cases} 1 & si \quad \text{le candidat } i \text{ est embauché} \\ 0 & si \quad \text{le candidat } i \text{ n' est pas embauché} \end{cases} \end{aligned}$$

2. Calculer $\Pr(X_i = 1)$.

   Observons que si *lecandidatiestembauché*, cela signifie qu'il est le meilleur candidat parmi les i premiers candidats. Donc $\Pr(X_i = 1) = \frac{1}{i}$.

3. Calculer $E(X_i)$.

$$E[X_i] = 0 \Pr(X_i = 0) + 1 \Pr(X_i = 1) = \frac{1}{i}.$$

Question 1.3 Donner le nombre moyen d'embauches dans la la procédure.

Soit X la variable aléatoire désignant le nombre d'embauches.

$X = \sum_{i=1}^{n} X_i$ (par définition)

$E[X] = \sum_{i=1}^{n} E[X_i]$ (par linéarité de la moyenne),

$E[X] = \sum_{i=1}^{n} \frac{1}{i}$ (par la question précédente),

$E[X] = \ln n + \mathcal{O}(1)$ (par la définition du nombre harmonique),

Question 1.4 En déduire le coût moyen de la procédure.

Le coût moyen de la procédure $= n \times c_i + E[X]c_h = n \times c_i + (\ln n + \mathcal{O}(1))c_h$

Question 2. Comme chaque interview a un coût pour la société, **la société ne souhaite pas interroger tous les candidats** afin de trouver le candidat le plus compétent. Pour cela, la société est prête à se contenter d'un candidat dont les compétences sont proches du candidat le plus compétent, en échange d'un recrutement unique.

La société utilise la procedure suivante : elle reçoit les candidats. Après avoir rencontré le candidat i, elle lui attribue un score noté $score(i)$. **Ici, nous supposons qu'aucun candidat ne reçoive le même score.** Après avoir vu k candidats, la société connaît le score le plus élevé (que l'on notera $Meilleur\_score$) des k premiers candidats. Ensuite la société sélectionnera le premier candidat qui aura un score plus grand que les k premiers candidats. Voici la procédure dans un pseudo-code :

Entrée : deux entiers k, et n correspondant le nombre de candidats

Sortie : le candidat sélectionné

1. $Meilleur\_score \leftarrow -\infty$
2. pour tout entier t allant de 1 à k faire
   (a) si $score(t) > Meilleur\_score$ alors $Meilleur\_score \leftarrow score(t)$

3. pour tout entier t allant de $k+1$ à n faire
   (a) si $score(i) > Meilleur\_score$ alors Retourner t
4. Retourner n

L'ordre du passage des candidats se fait aléatoirement et uniformément.

Question 3. Soit B_t l'événement que le candidat le plus compétent soit à la place t. Donner la probabilité $\Pr(B_t)$ tout en justifiant.

Nous pouvons considérer que l'ordre de passage peut se modéliser par un tableau généré de façon uniforme. la probabilité que le candidat le plus compétent soit à la place t correspond à la probabilité que le maximum d'un tableau généré de façon uniforme soit au rang t.

$$\Pr(B_t) = 1/n$$

Question 4. Soit O_t l'événement qu'aucun des candidats de la position k à la position $t-1$ soit retenu. Donner la probabilité $\Pr(O_t)$ tout en justifiant.

Il suffit de remarquer que O_t se réalise si le maximum du tableau $T[1,\ldots,t-1]$ se trouve en position j, avec $1 \leq j \leq k$. Donc $O_t = \bigcup_{t=1}^{k} B_t$. Comme ce sont des événements disjoints

$$\Pr(O_t) = \sum_{t=1}^{k} \Pr(B_t) = \frac{k}{n}$$

Soit S_t l'événement **que le candidat le plus compétent soit à la place t et qu'il soit retenu.**

Question 5. Donner la probabilité de l'événement S_t lorsque $1 \leq t \leq k$.

$\Pr(S_t) = 0$ car les k premiers candidats sont rejetés. En effet dans ce cas là, S_t ne se réalise jamais.

Question 6. ontrer que les événements O_t et B_t sont indépendants.

Rappellons que

- O_t l'événement qu'aucun des candidats de la position k à la position $t-1$ soit retenu.
- B_t l'événement que le candidat le plus compétent soit à la place t.

Rappellons que deux événements A et B sont indépendants si $\Pr(B) = \Pr(B|A)$.
Calculons $\Pr(O_t|B_t) = \Pr(O_t)$. Il suffit de noter que l'événement O_t dépend uniquement des scores des candidats à la position 1 à $t-1$.
Calculons $\Pr(B_t|O_t) = \Pr(B_t)$. Il suffit de noter que l'événement B_t dépend uniquement des scores de tous les candidats.

Question 7. Montrer que $\Pr(S_t) = \frac{k}{n(t-1)}$ pour $t > k$.

Tout d'abord notons que $S_t = O_t \cap B_t$ (par définition).
$\Pr(S_t) = \Pr(O_t \cap B_t) = \Pr(O_t) \cdot \Pr(B_t) = \frac{k}{n(t-1)}$.

Nous admettrons que $\ln j - \ln \ell \leq \displaystyle\sum_{i=\ell}^{j} \frac{1}{i} \leq \ln(j-1) - \ln(\ell - 1)$.

Question 8. Soit S l'événement que le candidat le plus compétent soit retenu.

1. Exprimer S et $\Pr(S)$ en fonction des S_t.

$S = \displaystyle\bigcup_{t=1}^{n} S_t$. Comme ce sont des événements disjoints, on a

$$\begin{aligned}
\Pr(S) &= \sum_{i=0}^{n} \Pr(S_i) \\
&= \sum_{i=k+1}^{n} \Pr(S_i) \quad \text{d'après la question 5}
\end{aligned}$$

2. Quelle est la probabilité $\Pr(S)$ que le meilleur candidat soit sélectionné ? Vous pouvez encadrer cette probabilité.

$$\begin{aligned}
\Pr(S) &= \sum_{i=k+1}^{n} \frac{k}{n(i-1)} \\
&= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} && \text{en sortant la constante de la somme} \\
&= \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} && \text{en supprimant le } -1 \\
&= \frac{k}{n} (\sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{k-1} \frac{1}{i}) && \text{en faisant apparaitre le nombre harmononique}
\end{aligned}$$

Comme $\ln(n) \leq \displaystyle\sum_{i=1}^{n-1} \frac{1}{i} \leq \ln(n-1) + 1$ et $-\ln(k) \geq -\displaystyle\sum_{i=1}^{k-1} \frac{1}{i} \geq -\ln(k-1) - 1$

$$\frac{k}{n}(\ln n - \ln k) \leq \Pr(S) \leq \frac{k}{n}(\ln(n-1) - \ln(k-1))$$

Question 9. En posant $k = n/3$. Donner l'ordre de grandeur de la probabilité $\Pr(S)$ en sachant que $\ln a - \ln b = \ln(\frac{a}{b})$ et $\ln 3 \approx 1.09$. Quand déduisez-vous ?

Remarquons que $\frac{k}{n} = 1/3$

$$\begin{aligned}
\frac{k}{n}(\ln n - \ln k) &\leq \Pr(S) && \text{d'après la question précédente} \\
\frac{1}{3}(\ln 3) &\leq \Pr(S) && \text{en utilisant la relation } k = n/3 \\
\frac{1}{3}1.09 &\leq \Pr(S) && \text{en utilisant la relation } \ln 3 \approx 1.09 \\
\frac{1}{3} &< \Pr(S)
\end{aligned}$$