# Problème du secrétaire

Une DRH auditionne $n$ candidats pour un poste de secrétariat. Ces $n$ candidats seront identifiés dans ce problème aux entiers $1, 2, \dots n$. On considère aussi que l'entier associé à un candidat est un indicateur de son niveau, de sorte que le candidat $n$ est le meilleur d'entre eux (le candidat 1 étant le plus mauvais).

Partie A. Modélisation de l'ordre de convocation.

Les candidats sont convoqués par la DRH dans un ordre tiré uniformément au hasard. On modélise ceci de la façon suivante. Soit $\Omega_n$ l'ensemble des $n$-arrangements à valeurs dans $[\![1, n]\!]$, c'est à dire des $n$-uplets d'entiers deux à deux distincts de $[\![1, n]\!]$. Par exemple pour $n = 3$, on a

$$\Omega_3 = \{(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1)\}.$$

On travaille avec l'espace probabilisé $(\Omega_n, P)$ où $P = P_n$ est l'équiprobabilité sur $\Omega_n$. Un élément de $\Omega_n$ donne l'ordre de convocation des candidats. Par exemple, pour $n = 3$, $(1, 3, 2)$ correspond à la situation où le candidat 1 est auditionné en premier, suivi du candidat trois, le candidat 2 étant auditionné en dernier. Dans les questions suivantes, $n$ est quelconque et $n \geq 3$.

1. Quel est le cardinal de $\Omega_n$ ?
2. Calculer la probabilité de l'événement $A$ : « le meilleur candidat est convoqué en dernier ». Celle de $B$ : « le pire candidat est convoqué en premier ».
3. Les événements $A$ et $B$ sont-ils indépendants ?

Partie B. Stratégie de recrutement.

La DRH n'a pas connaissance du niveau des candidats avant de les rencontrer et elle ne peut pas rappeler un candidat après l'avoir vu. Elle doit donc prendre une décision après chaque entretien : laisser partir le candidat à jamais ou le recruter. Elle fixe un entier $i \in [\![1, n-1]\!]$ et adopte la stratégie $\texttt{Strat}_{i,n}$ expliquée ci-dessous.

Pour un ordre de convocation $\omega = (x_1, \dots, x_n)$,

- Elle auditionne les $i$ premiers candidats $x_1, \dots x_i$ sans prendre aucune décision.
- Elle recrute parmi $x_{i+1}, \dots x_n$ le premier candidat qui est meilleur que tous les candidats $x_1, \dots x_i$.
- Si $x_{i+1}, \dots x_n$ sont tous moins bons que le meilleur des $i$ premiers, alors elle est obligée de recruter le dernier auditionné $x_n$.

On note $X(\omega)$ le candidat finalement recruté. Mathématiquement, cela donne

$$q(\omega) := \min\{j \in [\![i+1, n]\!] : x_j > \max(x_1, \dots, x_i)\}, \quad X(\omega) = x_{q(\omega)},$$

en convenant que $\min(\emptyset) = n$ (mais on pourra s'en tenir à l'explication en trois points ci-dessus). Par exemple, dans le cas $n = 3$, en appliquant $\texttt{Strat}_{1,3}$, on a $X((1, 2, 3)) = 2$ et $X((3, 2, 1)) = 1$.

1. Cas $n = 3$. On travaille sur $(\Omega_3, P)$ avec la stratégie $\texttt{Strat}_{1,3}$.
   - (a) Pour chaque élément $\omega$ de $\Omega_3$, donner $X(\omega)$. Écrire alors explicitement les événements $(X = 1)$, $(X = 2)$ et $(X = 3)$.
   - (b) Donner la loi de $X$.
   - (c) Calculer son espérance.
2. Cas $n = 4$. On travaille sur $(\Omega_4, P)$ avec la stratégie $\texttt{Strat}_{2,4}$.
   - (a) Donner la loi de $X$.
   - (b) Calculer son espérance.

Partie C. Une étude de fonction.

Après une étude de fonction soignée, donner le tableau de variations de la fonction

$$f : x \mapsto x \ln\left(\frac{1}{x}\right)$$

sur l'intervalle $]0, 1]$.

Partie D. À la recherche du meilleur.

La DRH cherche à choisir l'entier $i$ dont dépend sa stratégie de façon à optimiser la probabilité de recruter le meilleur candidat. Soit $n$ quelconque (et "grand"). On travaille sur $(\Omega_n, P)$ avec la stratégie $\texttt{Strat}_{i,n}$. Soit $Y : \Omega \to [\![1,n]\!]$ la variable donnant le rang du meilleur candidat ($n$) dans l'ordre de convocation.

1. Donner la loi de $Y$.
2. (**) Expliquer l'égalité

$$P(X=n \mid Y=k) = \begin{cases} 0 & \text{si } 0 \le k \le i \\ \frac{i}{k-1} & \text{si } i+1 \le k \le n. \end{cases}$$

**(on ne réclame pas forcément une preuve utilisant $(\Omega, P)$ : une "explication en français" suffira)**.

3. Exprimer $P(X = n)$, la probabilité de sélectionner le meilleur candidat (l'expression attendue utilise le symbole $\Sigma$).
4. (a) Justifier, pour $k \in \mathbb{N}^*$, l'inégalité $\frac{1}{k+1} \le \ln(k+1) - \ln(k) \le \frac{1}{k}$.

   (b) Justifier, en supposant $i \ge 2$, l'inégalité

$$\frac{i}{n} \ln\left(\frac{n}{i}\right) \le P(X=n) \le \frac{i}{n} \ln\left(\frac{n-1}{i-1}\right).$$

   (c) La DRH décide d'appliquer la stratégie $\texttt{Strat}_{i,n}$ en choisissant $i$ aux alentours de 37% de $n$. Approuvez-vous ce choix ?

## Corrigé

Le problème de *la* secrétaire est un classique. On cherche parfois à éviter la muflerie en préférant l'histoire d'une princesse devant se choisir un prétendant, mais est-ce vraiment mieux ? L'écriture de ce problème a été inspirée par cet article d'Étienne Ghys sur le thème de la décision en mathématiques

`http://images.math.cnrs.fr/Decision.html`.

Partie A. Modélisation de l'ordre d'arrivée.

1. D'après le cours sur le dénombrement $|\Omega_n| = n!$ (on a $n$ choix pour la première coordonnée, $n-1$ pour la seconde, etc...)
2. L'événement $A$ est l'ensemble des $n$-uplets de la forme $(x_1, \dots, x_{n-1}, n)$ où $(x_1, \dots, x_{n-1})$ est un $(n-1)$ arrangement de $[\![1, n-1]\!]$. L'événement $A$ est donc de cardinal $(n-1)!$. Puisqu'on travaille avec l'équiprobabilité, sa probabilité est

$$P(A) = \frac{|A|}{|\Omega_n|} = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

   L'événement $B$ est l'ensemble des $n$-uplets de la forme $(1, x_2, \dots, x_n)$ où $(x_2, \dots, x_n)$ est un $(n-1)$ arrangement de $[\![2, n]\!]$. L'événement $B$ a donc même cardinal que $A$ et même probabilité $\frac{1}{n}$.
3. Examinons l'événement $A \cap B$ : c'est l'ensemble des $n$-uplets $(1, x_2, \dots, x_{n-1}, n)$. tels que $(x_2, \dots, x_{n-1})$ est un $(n-2)$-arrangement de $[\![2, n-1]\!]$. Il y en a $(n-2)!$ ($(n-2)$ choix pour la deuxième coordonnée, $(n-3)$ choix pour la troisième etc...) On a donc

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega_n|} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

   Cette probabilité est différente de $P(A)P(B) = \frac{1}{n^2}$ donc les événements ne sont pas indépendants. Cela se comprend mieux en écrivant des probabilités conditionnelles : $P(A \mid B) = \frac{1}{n-1} > \frac{1}{n} = P(A)$ : sachant que le plus mauvais était convoqué en premier, on n'a plus que $n-1$ places pour convoquer le meilleur.

Partie B. Stratégie de recrutement.

1. (a) En appliquant $\texttt{Strat}_{1,3}$, on obtient

$$X((1,2,3)) = 2 \quad X((1,3,2)) = 3 \quad X((2,1,3)) = 3$$

$$X((2,3,1)) = 3 \quad X((3,1,2)) = 2 \quad X((3,2,1)) = 1$$

   On a donc $(X=1) = \{(3,2,1)\}$, $(X=2) = \{(1,2,3),(3,1,2)\}$, $(X=3) = \{(1,3,2),(2,1,3),(2,3,1)\}$.

   (b) La probabilité des événements ci-dessus s'obtient en divisant leur cardinal par celui de $\Omega_3$ qui vaut 6. On présente la loi de $X$ à l'aide du tableau

| $k$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X=k)$ | $\frac{1}{6}$ | $\frac{2}{6}$ | $\frac{3}{6}$ |

(c)

$$E(X) = \frac{1}{6}.1 + \frac{2}{6}.2 + \frac{3}{6}.3 = \frac{14}{6} = \frac{14}{6} = \frac{7}{3}.$$

(a) En appliquant $\mathtt{Strat}_{2,4}$, on calcule (au brouillon) l'image par $X$ des 24 éléments de l'univers $\Omega_4$. On trouve 4 fois 1, 4 fois 2, 6 fois 3, et 10 fois 4. La loi de $X$ est donc donnée par

| $k$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=k)$ | $\frac{4}{24}=\frac{1}{6}$ | $\frac{4}{24}=\frac{1}{6}$ | $\frac{6}{24}=\frac{1}{4}$ | $\frac{10}{24}=\frac{5}{12}$ |

(b)

$$E(X) = \frac{4}{24}.1 + \frac{4}{24}.2 + \frac{6}{24}.3 + \frac{10}{24}.4 = \frac{70}{24} = \frac{35}{12}.$$

Partie C. Une étude de fonction.

Pour tout $x \in ]0,1]$, $f(x) = -x\ln(x)$. La fonction $f$ est dérivable sur $]0,1]$ comme produit de fonctions dérivables et pour tout $x$ dans l'intervalle, $f'(x) = -\ln(x) + (-x).\frac{1}{x} = -\ln(x) - 1$, qui change de signe en $e^{-1}$.

| $x$ | $0$ | | $e^{-1}$ | | $+\infty$ |
|---|---|---|---|---|---|
| $f'(x)$ | | $+$ | $0$ | $-$ | |
| $f$ | $0$ | $\nearrow$ | $e^{-1}$ | $\searrow$ | $0$ |

Pour calculer la limite en $0_+$, on a utilisé les croissances comparées.

Partie D. À la recherche du meilleur.

1. Appuyons-nous sur l'intuition : le meilleur candidat a autant de chances d'être convoqué en premier qu'en douzième (si $n \geq 12$...) On conjecture donc que $Y$ est de loi uniforme sur $[\![1,n]\!]$. Prouvons-le. Soit $k \in [\![1,n]\!]$. L'événement $(Y=k)$ est l'ensemble des $n$-uplets tels que la $k$-eme coordonnées vaut $n$. Il est donc de cardinal $(n-1)!$ et donc de probabilité $\frac{(n-1)!}{n!} = \frac{1}{n}$, CQFD.
2. (*) Si $n$ se trouve parmi les $i$ premiers candidats convoqués, il ne sera pas recruté, ce qui donne bien que $P(X=n \mid Y=k) = 0$ pour $0 \leq k \leq i$. Supposons désormais que $k \in [\![i+1,n]\!]$. Si on sait que $Y=k$, alors la DRH va recruter le candidat $n$ si et seulement si aucun candidat meilleur que les $i$ premiers ne se présente entre $i+1$ et $k-1$. Cela signifie que le meilleur candidats parmi les $k-1$ premiers se trouve dans les $i$ premiers. Il est intuitif que le rang de ce meilleur candidat est uniforme sur $[\![1,k-1]\!]$. La probabilité qu'il fasse partie des $i$ premiers auditionnés est donc bien $\frac{i}{k-1}$.

   Donnons une « vraie » preuve. L'événement $(X=n) \cap (Y=k)$ est l'ensemble des $n$-uplets $(x_1,\dots,x_n)$ tels que $x_k = n$ et $\max(x_1,\dots,x_{k-1}) \in \{x_1,\dots,x_i\}$. On peut construire un tel $n$-uplet de la manière suivante :

   (a) On choisit $k-1$ entiers $\{a_1,\dots,a_{k-1}\}$ parmi $[\![1,n-1]\!]$ : $\underline{\binom{n-1}{k-1} \text{ choix}}$.

   (b) Ces $k-$ entiers ont un maximum. On choisit son rang dans le $n$-uplet entre 1 et $i$ : $\underline{i \text{ choix}}$.

   (c) On vient de fixer une coordonnée de $(x_1,\dots,x_{k-1})$. On fixe les autres ; cela revient à se donner un $k-2$ arrangement de $\{a_1,\dots,a_{k-1}\} \setminus \max(a_1,\dots,a_{k-1})$ : $\underline{(k-2)! \text{ choix}}$.

   (d) On fixe les coordonnées $(x_{i+1},\dots,x_n)$ : en se donnant un ordre pour les éléments de $[\![1,n-1]\!] \setminus \{a_1,\dots,a_{k-1}\}$. $\underline{(n-k)! \text{ choix}}$.

   On a donc

   $$|(X=n)\cap(Y=k)| = \binom{n-1}{k-1} \times i \times (k-2)! \times (n-k)! = \frac{(n-1)!i}{k-1},$$

   d'où

   $$P\left((X=n)\cap(Y=k)\right) = \frac{\frac{(n-1)!i}{k-1}}{n!} = \frac{i}{n(k-1)}$$

   Puisque $P(Y=k) = \frac{1}{n}$, le passage aux probabilités conditionnelles amène le résultat demandé.
3. La formule des probabilités totales amène

$$\begin{aligned} P(X=n) &= \sum_{k=1}^{n} P(X=n \mid Y=k)\cdot P(Y=k) \\ &= \sum_{k=i+1}^{n} \frac{i}{k-1}.\frac{1}{n} \\ &= \frac{i}{n}\sum_{k=i}^{n-1}\frac{1}{k}. \end{aligned}$$

4. (a) Il s'agit d'encadrer l'accroissement d'une fonction : utilisons les accroissements finis. Soit $k \geq 1$. La fonction ln est continue sur $[k, k+1]$ et dérivable sur $]k, k+1[$. D'après le théorème des accroissements finis, il existe $c \in ]k, k+1[$ tel que

$$\ln(k+1) - \ln(k) = \frac{1}{c} \in \left] \frac{1}{k+1}, \frac{1}{k} \right[,$$

ce qui implique le résultat demandé.

(b) Supposons $i \geq 2$. En sommant des inégalités, on obtient

$$\frac{i}{n} \sum_{k=i}^{n-1} (\ln(k+1) - \ln(k)) \leq \frac{i}{n} \sum_{k=i}^{n-1} \frac{1}{k} \leq \frac{i}{n} \sum_{k=i}^{n-1} (\ln(k) - \ln(k-1)),$$

ce qui amène par télescopage

$$\frac{i}{n}(\ln(n) - \ln(i)) \leq P(X=n) \leq \frac{i}{n}(\ln(n-1) - \ln(i-1)).$$
$$\frac{i}{n}(\ln(n) - \ln(i)) \leq P(X=n) \leq \frac{i}{n}(\ln(n-1) - \ln(i-1)).$$
$$\frac{i}{n} \ln\left(\frac{n}{i}\right) \leq P(X=n) \leq \frac{i}{n} \ln\left(\frac{n-1}{i-1}\right).$$

(c) Si $n$ est grand, l'inégalité précédente montre que $P(X = n)$, la probabilité de trouver le meilleur candidat, est égale environ à $f\left(\frac{i}{n}\right)$, où $f$ est la fonction étudiée en partie C. Cette fonction atteint son maximum en $e^{-1} \approx 0.37$, ce qui montre que le choix de la DRH est le bon : en choisissant l'entier $i$ aux alentours de 37% de $n$, elle optimise ses chances de trouver le meilleur candidat. Sa probabilité de trouver ce candidat est alors proche de $f(e) = e^{-1} \approx 0.37$.