Problème des secrétaires
Rapport Final MAP471A

Sabrina Lomelino Sartori

19 novembre 2023

Table des matières

1 Introduction

Le problème des secrétaires, dans sa forma classique, est le suivant : un recruteur souhaite embaucher le meilleur parmi un nombre de candidats fixe n. Les candidats doivent être interviewés un par un, dans un ordre aléatoire où toutes les $n!$ permutations possibles sont également probables. Le recruteur peut, à tout moment, de classer par ordre de préférence les éléments qui lui ont déjà été présentés. À chaque entretien (i.e., à chaque étape du processus), il doit :

— Soit accepter le candidat, mettant ainsi fin au processus d'entretiens ;

— Soit de le refuser, et interviewer le prochain candidat, en faisant face au même problème qu'avant.

Si le dernier candidat est interviewé, il doit être accepté par le recruteur.

L'objectif du recruteur est de maximiser la probabilité que le candidat qu'il accepte soit le meilleur parmi les n candidats disponibles. Le dilemme de ce scénario est la nécessite de ne pas choisir un candidat trop tôt, alors qu'un meilleur pourrait apparaître plus tard, et de non plus prolonger trop longtemps le processus, pour finalement réaliser que le meilleur élément a été rejeté plus tôt. Cela vient du fait que, quand le recruteur choisis un candidat pour embaucher, sa décision est finale, et il n'aura jamais la possibilité de choisir ce candidat s'il lui rejette.

Tout d'abord, cette version classique du problème sera traitée, en détaillant la solution qui permet le recruteur d'avoir plus de chances de choisir le meilleur candidat. Cette solution sera ensuite comparée à d'autres stratégies non optimales du même problème. Finalement, une autre version du problème, appelée problème des k-secrétaires, sera étudiée.

2 Solution du problème

2.1 Probabilités

La solution à ce problème consiste à trouver le rang $r^* < n$ tel que les r^* premiers candidats sont rejetés et le candidat accepté est le prochain candidat qui est meilleur que tous les r^* premiers interviewés.

Soit i la position du meilleur candidat. Ce candidat sera accepté si :

— $i \geq r^*$

— Le candidat le mieux classé entre les positions $[1, i-1]$ est entre les positions $[1, r^*]$

La probabilité d'avoir le meilleur candidat à la position i est $\frac{1}{n}$ et la probabilité d'avoir la dernière condition mentionnée ci-dessus est $\frac{r^*}{i-1}$. La figure illustre cela :

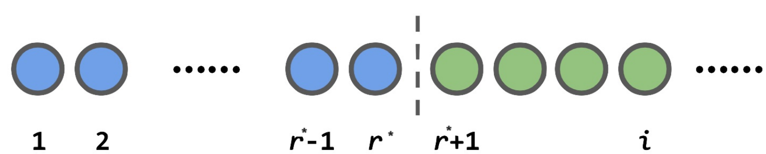

FIGURE 1 – Solution par une approche probabiliste

Considérons notre probabilité de succès, c'est-à-dire la probabilité de choisir le meilleur candidat parmi les n, comme étant $P(r)$.

Nous avons alors :

$$P(r^*) = \sum_{i=r^*+1}^{n} \frac{1}{n} \left(\frac{r^*}{i-1} \right) = \frac{r^*}{n} \sum_{i=r^*+1}^{n} \frac{1}{i-1} \tag{1}$$

On peut remarquer que $P(r^*)$ est une approximation de Riemann d'une intégrale. En posant le rapport $x = \frac{r^*}{n}$ et en prenant la limite lorsque $n \to \infty$:

$$P(r^*) = \lim_{n\to\infty} \frac{r^*}{n} \sum_{i=r^*+1}^{n} \frac{1}{i-1} = x \int_x^1 \frac{1}{t} dt = -xlnx \tag{2}$$

En maximisant l'expression ci-dessus, nous obtenons le résultat optimal :

$$x = \frac{1}{e} \Rightarrow r^* = \frac{n}{e} \tag{3}$$

Finalement, la solution optimale consiste à rejeter les $\frac{n}{e}$ premiers candidats et à choisir le premier candidat qui est meilleur que les $\frac{n}{e}$ candidats rejetés. Dans le cas où aucun candidat meilleur n'a été interviewé, choisir le dernier.

2.2 Programmation dynamique

Pour comprendre l'origine de la stratégie optimale décrite (interviewer un nombre donné de candidats et seulement après choisir), il faut décrire le processus sur la forme d'état et valeur. Soit r le nombre de candidats que le recruteur a vus jusqu'à présent et s le classement *relative* du dernier, c'est-à-dire le r-ème candidat. La fonction valeur qui on cherche est la probabilité maximale attendue de choisir le meilleur candidat *absolu* lorsque l'état est (r, s). Comme avant, il y a un total de n candidats à considérer et on doit trouver la relation de la fonction valeur entre différents états.

2.2.1 Cas particuliers

Le cas le plus simple à analyser est celui où tous les candidats ont été interviewés (i.e., $r = n$). La fonction valeur correspond à la probabilité maximale attendue de choisir le meilleur candidat dans l'état (N, s). L'unique possibilité de succès est quand $s = 1$, donc on a que $V(n, s) = 1$ si $s = 1$ et 0 sinon.

Dans le cas où tous les candidats à l'exception de un ont été interviewés, si le classement relative du $(n-1)$-ème candidat n'est pas 1, le recruteur doit continuer à interviewer le dernier candidat, espérant arriver à l'état $(N, 1)$. Puisque la une probabilité que le dernier candidat soit le meilleur est $\frac{1}{n}$, alors la probabilité de succès à l'état $(n-1, s)$, si $s \neq 1$, est :

$$V(n-1, s) = \frac{1}{n}V(n, 1), s = 2, 3, ..., n-1 \tag{4}$$

Si $s = 1$, le $(n-1)$-ème candidat interviewé a un classement relative 1, et donc la probabilité que son classement absolu soit aussi 1 est $\frac{n-1}{n}$, tandis que la probabilité que le dernier candidat soit le meilleur (et le $(n-1)$-ème soit le deuxième) est $\frac{1}{n}$. La fonction valeur dans cet état est donc :

$$V(n-1, 1) = max\left[\frac{n-1}{n}, \frac{1}{n}V(n, 1)\right] \tag{5}$$

2.3 Cas général

L'analyse des cas plus simples montre qu'on a besoin de distinguer l'état $(r, 1)$ des autres, car si $s \neq 1$, le recruteur doit toujours interviewer le prochain candidat. Comme le candidat suivant est le $(r+1)$-ème interviewé, son classement relative est entre 1 et $r+1$, chaque cas avec probabilité $\frac{1}{r+1}$. Donc, pour $s \neq 1$, on a :

$$V(r, s) = \sum_{s'=1}^{r+1} \frac{1}{r+1}V(r+1, s'), s = 2, 3, ..., r \tag{6}$$

Si $s = 1$, si le recruteur choisit le r-ème candidat, il a une probabilité de succès $\frac{r}{n}$, et s'il continue a interviewer sa probabilité de succès est donnée par l'équation 6. Alors la fonction valeur à l'état $(r, 1)$ est :

$$V(r,1) = max\left[\frac{r}{n}, \sum_{s'=1}^{r+1} \frac{1}{r+1} V(r+1,s')\right] \tag{7}$$

La solution a été trouvé par Lindley en 1961 par *backward recursion* sur r :

$$\begin{aligned} V(r,s) &= \frac{r}{n}\sum_{i=r}^{n-1}\frac{1}{i}, s \neq 1 \\ V(r,1) &= \frac{r}{n} max\left[1, \sum_{i=r}^{n-1}\frac{1}{i}\right] \end{aligned} \tag{8}$$

2.4 Stratégie

Comme le recruteur va choisir un candidat seulement s'il est dans un état du type $V(r,1)$, il faut comprendre comment la fonction $max\left[1, \sum_{i=r}^{n-1}\frac{1}{i}\right]$ se comporte quand on varie r. On vérifie que la fonction $f(r) = \sum_{i=r}^{n-1}\frac{1}{i}$ est monotone décroissante, de façon que, une fois que $f(r) \leq 1$, le recruteur doit arreter le processus et choisir le candidat dans l'état $(r,1)$. On doit donc chercher r^* tel que $f(r^*+1) \leq 1 < f(r^*)$. Cela montre que la stratégie consiste a interviewer les candidats jusqu'au r^*-ème et après choisir le premier candidat parmi les suivants dont le classement apparent est 1.

Après le r^*-ème candidat, on a une probabilité de $\frac{1}{r^*+1}$ d'arriver à l'état $(r^*+1,1)$ et, si c'est le cas, la probabilité de succès est $\frac{r^*+1}{n}$. Si le classement relative du prochain candidat (le (r^*+1)-ème) n'est pas 1 et nous sommes à l'état (r^*+1,s) le recruteur doit interviewer le (r^*+1)-ème candidat et le processus continue. En prenant la valeur espérée de cette serie, on a :

$$
\begin{aligned}
P(r^*, n) &= \frac{1}{r^*+1}\frac{r^*+1}{n} + \frac{r^*}{r^*+1}P(r^*+1, n) \\
&= \frac{1}{n} + \frac{r^*}{r^*+1}\left(\frac{1}{r^*+2}\frac{r^*+2}{n} + \frac{r^*}{r^*+2}P(r^*+2, n)\right) \\
&= \frac{r^*}{n}\left(\frac{1}{r^*} + \frac{1}{r^*+1} + ... + \frac{1}{n-1}\right) \\
&= \frac{r^*}{n}\sum_{i=r^*}^{n-1}\frac{1}{i}
\end{aligned}
\tag{9}
$$

L'expression donnée en 9 est la même que cela de l'équation 1, alors on a le même résultat.

3 Rang du secrétaire recruté

Les résultats suivants sont une simulation du problème classique, qui a pour objective maximiser la probabilité attendue de choisir le meilleur candidat. Ils ont été trouvés en considérant un nombre $n = 100$ de candidats et avec 10000 itérations.

3.1 Stratégie Optimale

La distribution des classements absolus des candidats choisis dans les simulations est donnée dans la figure 2. On vérifie que 3712 simulations ont été un succès, de façon que la probabilité estimée de choisir le meilleur candidat est 37%. La figure 3 illustre la fonction de répartition de la choix du candidat. L'inclination de la courbe montre que la stratégie optimale non seulement maximise la probabilité de choisir le candidat 1, mais aussi que, dans les cas d'échec, le recruteur est plus probable de choisir les premiers classés que les autres. Dans ce cas particulier, par exemple, la probabilité que un des 5 meilleurs candidats soit choisi est 63%.

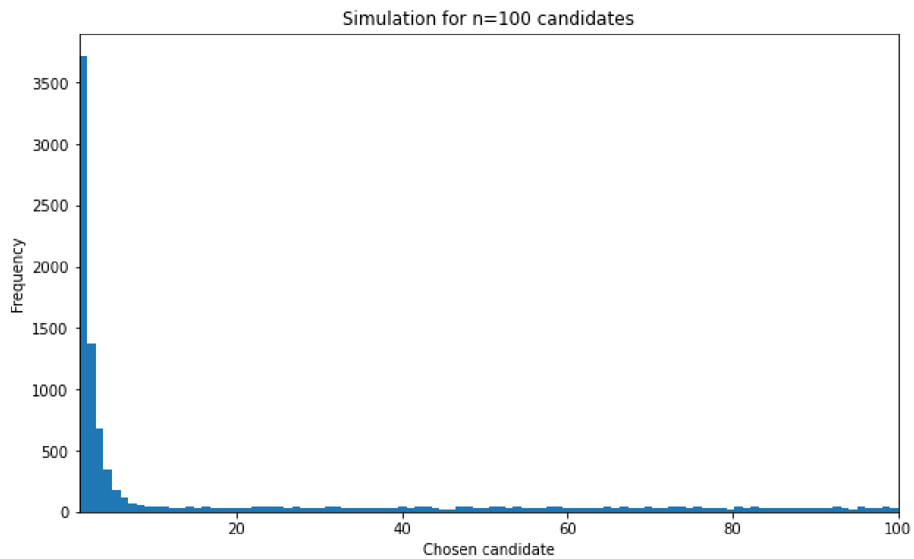

FIGURE 2 – Distribution des candidats choisis

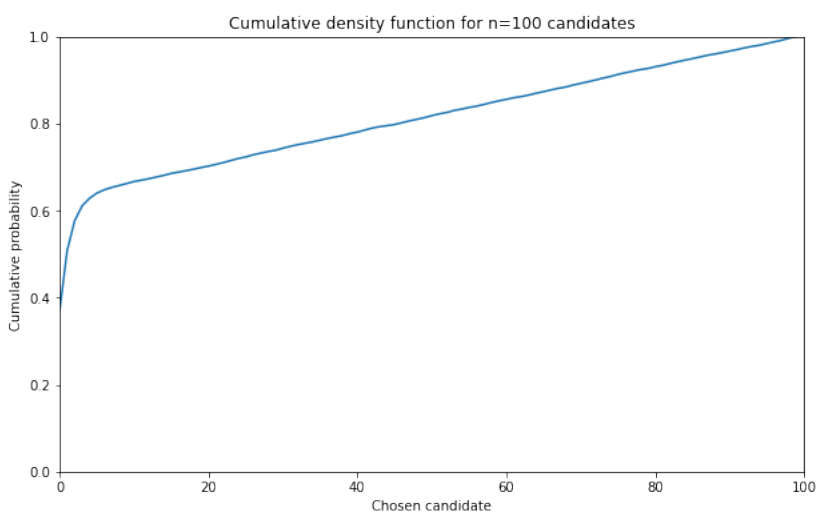

FIGURE 3 – Probabilité accumulée

3.2 D'autres stratégies

Pour étudier le comportement du choix de candidat et observer si la meilleur stratégie est de rejetter les $\frac{n}{e}$ premiers candidats interviewés, nous allons modifier la règle d'arrêt et rejeter d'autres valeurs que $\frac{1}{e}\%$ ($\approx 36.8\%$) des candidats. La fraction de candidats rejetés variera de 0.05 à 1 par incréments de 0.01. La figure 4 montre comment la probabilité de succès varie en fonction de cette fraction de candidats rejetés.

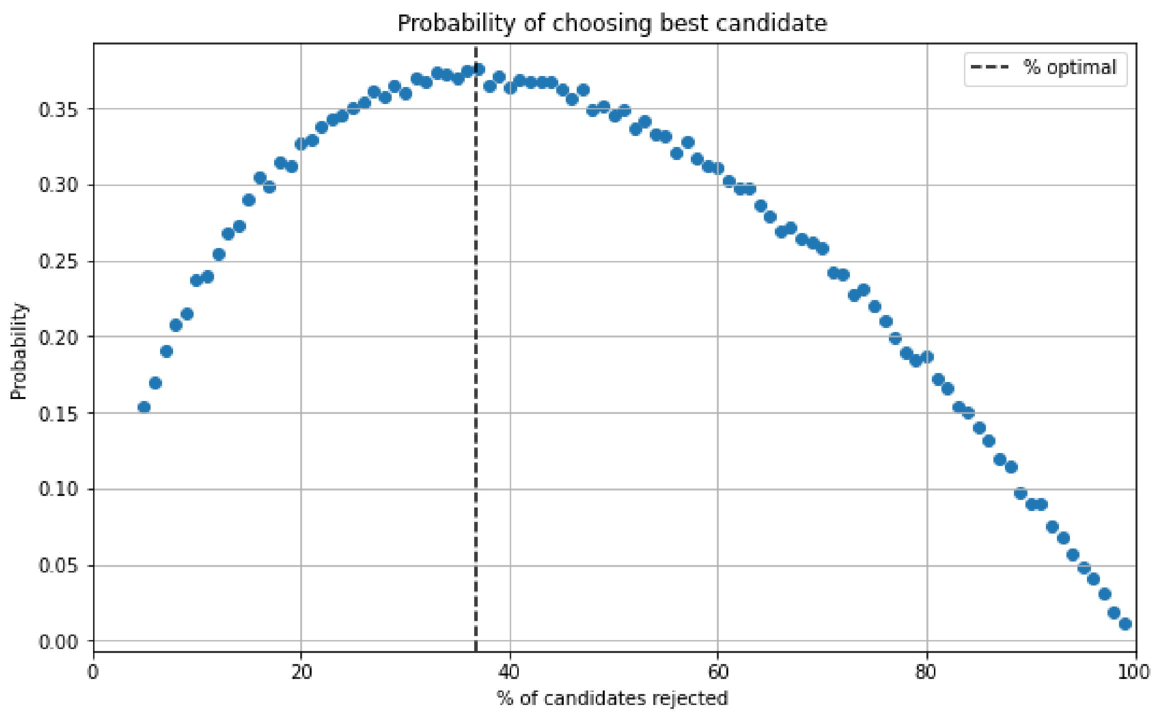

FIGURE 4 – Comparaison de stratégies

La valeur qui a donnée la plus grande probabilité est 37%, ce qui était déjà attendu d'après la stratégie optimale.

4 Chance maximale de recruter le meilleur

On va maintenant traiter une autre version du problème, appelée problème des k-secrétaires. Dans cette version, le recruteur doit accepter k candidats parmi les n. La solution à ce nouveau problème doit également maximiser la probabilité de recruter le meilleur candidat, c'est-à-dire, maximiser les chances d'avoir le candidat le mieux classé parmi les k choisis.

Ce problème a été traité la première fois par Gilbert et Mosteller en 1966. Comme dans la section 2.2, on va définir les états (r, s) et la fonction valeur $V(r, s)$. Les états caractérisent l'étape du processus où le recruteur arrive au r-ème candidat, ce candidat étant possible d'être interviewé (i.e., ne doit pas être rejeté), et le recruteur a encore besoin d'accepter s candidats. L'état (∞, s) signifie que il est arrivé au dernier candidat et doit lui rejeter, tandis que l'état $(\infty, 0)$ signifie que le recruteur a déjà choisi les k candidats.

Si nous sommes à l'état (r, s), on a deux possibilités : soit accepter le r-ème candidat, soit le rejeter. S'il est accepté et il est le meilleur candidat parmi les n, c'est déjà un succès et la probabilité de cet évènement est $\frac{r}{n}$. S'il est accepté mais à un état $(j, s-1)$ il y a un candidat meilleur que lui, la probabilité que le r-ème interviewé soit encore celui de meilleur classement relative jusqu'à l'état $(j-1, s-1)$ est $\frac{r}{j-1}$, et la probabilité que le j-ème candidat observé soit le nouveau mieux classé (i.e., soit meilleur que le r-ème) est $\frac{1}{j}$. Donc la probabilité de transition de l'état (r, s) à l'état $(j, s-1)$ est $\frac{r}{j(j-1)}$, et la probabilité de succès dans ce cas est $\sum_{j=r+1}^{n} \frac{r}{j(j-1)} V(j, s-1)$.

S'il n'est pas accepté, la probabilité de succès est seulement $\sum_{j=r+1}^{n} \frac{r}{j(j-1)} V(j, s)$ et la fonction valeur est donc :

$$V(r,s) = max\left[\frac{r}{n} + \sum_{j=r+1}^{n} \frac{r}{j(j-1)} V(j,s-1), \sum_{j=r+1}^{n} \frac{r}{j(j-1)} V(j,s)\right] \tag{10}$$

4.1 Cas avec $k = 2$

On va simuler le problème pour le cas particulier de $k = 2$. La stratégie trouvée a partir de l'équation 10 consiste à trouver les positions $r_2^{**} < r_1^{**} < n$ telles que, initialement, les premiers $r_2^{**} - 1$ candidats sont rejetés et le premier candidat accepté est le prochain candidat qui est meilleur que tous les premiers $r_2^{**} - 1$ interviewés. Après le choix du premier candidat, ceux qui n'ont pas été interviewés sont comparés aux $r_1^{**} - 1$ premiers interviewés, et le deuxième candidat choisi est le prochain qui est meilleur que tous les premiers $r_1^{**} - 1$ interviewés. Les étapes sont les suivantes :

- — Rejeter les premiers $r_2^{**} - 1$ candidats et comparer les suivants au meilleur de ceux rejetés ;
- — Choisir le premier candidat meilleur que le meilleur parmi les premiers $r_2^{**} - 1$;
  - — S'il n'y a pas de candidat meilleur, ou si le candidat choisi est le $n - 1$-ème interviewé, les deux candidats sélectionnés sont les deux derniers.
- — Si vous n'avez pas encore interviewé les premiers $r_1^{**} - 1$ candidats, rejeter ceux restants jusqu'à cette position ;
- — Choisir le premier candidat meilleur que le meilleur parmi les premiers $r_1^{**} - 1$.
  - — S'il n'y a pas de candidat meilleur, choisir le dernier.

Les valeurs numériques trouvées par Gilbert et Mosteller sont :

$$r_1^{**} - 1 = \frac{n}{e} \tag{11}$$

$$r_2^{**} - 1 = \frac{n}{e^{3/2}} \tag{12}$$

On peut voir que l'équation 11 correspond au résultat du problème classique ($k = 1$).

En simulant le problème avec $n = 100$ et 10000 itérations, on peut vérifier la distribution du classement du meilleur candidat parmi les $k = 2$ choisis, comme illustre la figure 5. On vérifie qu'il y a 5511 cas de succès parmi les 10000 itérations, donc la probabilité que le recruteur accepte le meilleur candidat est d'environ 55%.

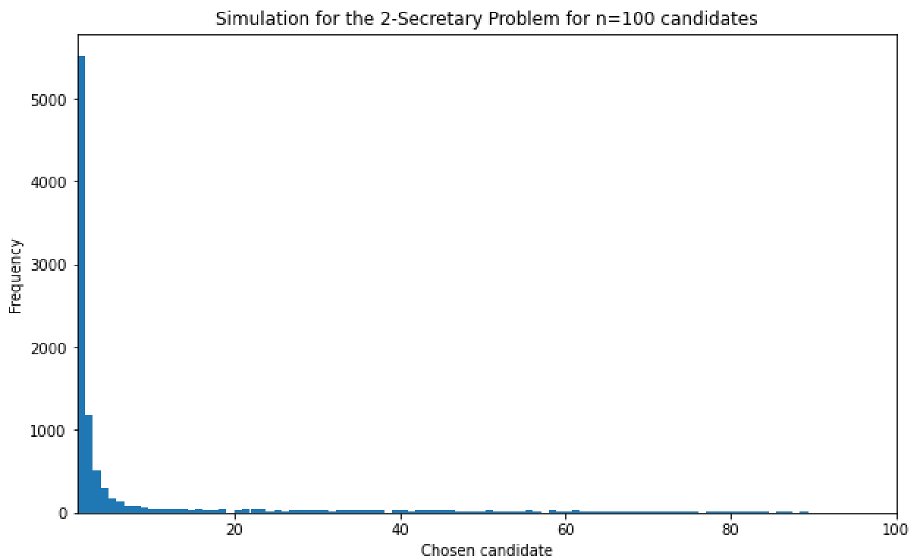

FIGURE 5 – Distribution du meilleur candidat parmi les 2 choisis

5 Conclusion

Le problème des secrétaires est en essence un problème d'arrêt optimal où il faut trouver un équilibre entre les deux défis du recruteur : la manque d'informations sur le classement absolu des candidats et le fait qu'après avoir rejeté un candidat, il n'est pas possible de revenir en arrière.

Le problème a été traité plus en détaille dans sa forme classique, où la probabilité doit être maximisée sous la contrainte d'accepter seulement un candidat. De plus, la stratégie optimale a été simulée et comparée avec d'autres possibles règles d'arrêt. Finalement, une variation du problème a été discutée, où la contrainte est de recruter deux candidats.

En somme, le problème des secrétaires a plusieurs applications dans différents domaines, comme le problème du mariage stochastique, le problème de la sélection de maisons et le problème de l'embauche de personnel. Il est important puisqu'il est un exemple clé de l'analyse des compromis entre la prise de décision immédiate et la recherche du meilleur résultat possible, en tenant compte des informations progressivement acquises.

Références

[AL20] Susanne Albers and Leon Ladewig. New results for the k-secretary problem, 2020.

[Fer89] Thomas S. Ferguson. Who Solved the Secretary Problem ? *Statistical Science*, 4(3) :282 – 289, 1989.

[Fre83] P. R. Freeman. The secretary problem and its extensions : A review. *International Statistical Review / Revue Internationale de Statistique*, 51(2) :189–206, 1983.