Stratégies de recrutement optimal

Hugo Laurençon et Nicolas Podvin

Dans le cadre du projet MOPSI de l'Ecole nationale des ponts et chaussées
Sous la direction de Pierre-André Zitt

École des Ponts ParisTech

Introduction

Dans cette étude, on s'intéresse à un recrutement de personnes avec décision immédiate d'engager ou non. Nous allons étudier plusieurs stratégies (recrutement au dessus du meilleur score, de la moyenne ou de la médiane des candidats déjà recrutés) et les comparer en fonction du nombre de personnes recrutées après n entretiens et de la qualité moyenne de l'équipe récrutée.

Notations On modélise les scores des candidats par une suite de variables aléatoires réelles $(X_n)_{n\geq 1}$ i.i.d de fonction de répartition F continue et de qualité maximale x_F. On note R_n le rang relatif de X_n dans $X_1, \dots, X_n$, Z_n le nombre de recrutés après n interview, Y_i le score du i-ème candidat recruté et T_i le temps où on le recrute, et M_i la moyenne et C_i la médiane des $Y_1, \dots, Y_i$.

Elitisme

On recrute une personne si elle est meilleure que toutes les personnes déjà recrutées. On exprime le nombre de recruté après n interview en fonction des rangs relatifs.

$$R_n = 1 + \sum_{k=1}^{n} \mathbf{1}_{\{X_n > X_k\}} \quad \text{d'où} \quad Z_n = \sum_{i=1}^{n} \mathbf{1}_{\{R_i = i\}}$$

Proposition 1 Les variables aléatoires $(R_n)_{n\geq 1}$ sont indépendantes et R_n suit la loi uniforme sur $\{1, \dots, n\}$.
On obtient donc

$$\mathbb{E}(Z_n) = \sum_{k=1}^{n} \frac{1}{k} \sim \log(n)$$

$$Var(Z_n) = \sum_{k=1}^{n} \frac{k-1}{k^2} \sim \log(n)$$

Plus précisément, nous avons le résultat du Théorème 2.

Théorème 1 (Lindeberg-Lévy) Considérons $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires réelles indépendantes centrées de carré intégrables, avec σ_i leur écart-type. Notons $S_n = X_1 + \cdots + X_n$, $s_n = \sqrt{Var(S_n)}$. Si la condition de Lindeberg-Lévy (notée L-L) est vérifiée :

$$\forall t \geqslant 0 \quad \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}[X_i^2 \mathbf{1}_{\{|X_i| \geq t s_n\}}] \underset{n\to\infty}{\longrightarrow} 0, \quad \text{alors} \quad \frac{S_n}{s_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Théorème 2 (Comportement asymptotique du nombre de records)

$$\frac{Z_n}{\log(n)} \xrightarrow[n\to\infty]{p.s} 1 \quad et \quad \frac{Z_n - \log(n)}{\sqrt{\log(n)}} \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,1)$$

Ce résultat repose le théorème des séries centrées et le théorème de Lindeberg-Levy (généralisation du théorème central limite au cas où les variables aléatoires ne sont plus identiquement distribuées, voir le théorème 3 et [2]).

Proposition 2 On suppose $X_1 \sim U(0,1)$ et on note $G_n = 1 - Y_n$ le gap. On a

$$\mathbb{E}(G_n) = \frac{G_0}{2^n}, \quad \mathbb{E}(M_n) = 1 - \Theta(1/n) \quad \text{et} \quad \frac{\log(T_n)}{n} \xrightarrow[n\to\infty]{p.s} 1.$$

Au dessus de la moyenne

On recrute un individu s'il est meilleur que la moyenne des candidats déjà recrutés.

Proposition 3 (Convergence vers la qualité maximale)

$$M_n \xrightarrow[n\to\infty]{p.s} x_F$$

Théorème 3 (Séries centrées) Soit $(Q_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et centrées et une suite (déterministe) $(b_n)_{n\geq 1}$ strictement croissante non bornée de réels strictement positifs.

$$\text{Si} \quad \sum_{n=1}^{+\infty} \frac{\mathbb{E}(Q_n^2)}{b_n^2}, \quad \text{alors} \quad \frac{1}{b_n} \sum_{k=1}^{n} Q_k \xrightarrow[n\to\infty]{p.s} 0$$

Théorème 4 (Comportement asymptotique des auditionnés) Notons $P_n = 1 - F(M_{n-1})$ la probabilité d'engager le n-ème candidat recruté. Sous une hypothèse de régularité sur cette probabilité,

$$\exists \alpha \geq 0, \quad W \quad | \quad \mathbb{P}(0 < W < +\infty) = 1 \quad et \quad n^\alpha P_n \xrightarrow[n\to\infty]{p.s} W,$$

nous avons

$$\frac{T_n}{n^{\alpha+1}} \xrightarrow[n\to\infty]{p.s} \frac{1}{(\alpha+1)W}$$

Pour obtenir ce théorème, on utilise à nouveau le théorème des séries centrées, cette fois ci conditionnellement à la tribu $\mathcal{F}$ engendrée par la suite $(M_n)_{n\geq 1}$.

Simulations numériques pour $X \sim Exp(1)$ Choix de la loi exponentielle : On sait théoriquement que $\alpha = 2$ et W suit une loi de Gumbel (voir [3]), et l'absence de mémoire rend les calculs plus rapides (de 10h à 7sec). Résultat : Nous obtenons après une régression linéaire $\hat{\alpha} = 1.999$ et la fonction de répartition de la loi de Gumbel et celle de la variable simulée se superposent.

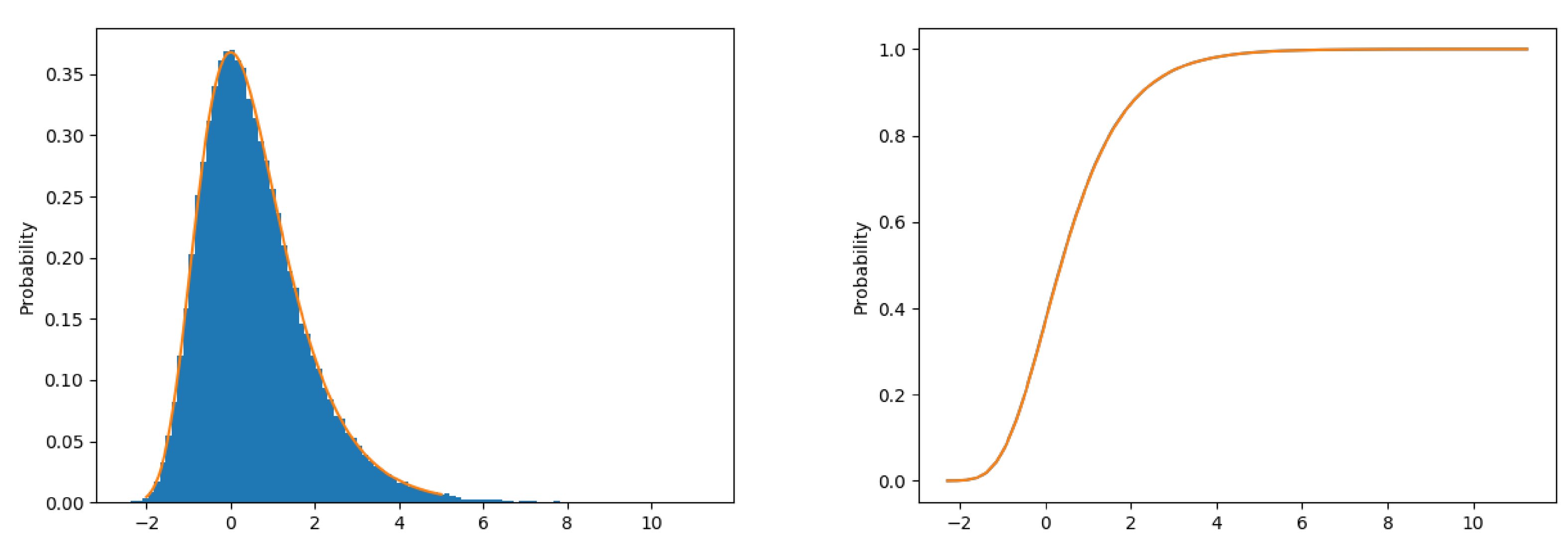

Fig. 1: A gauche : en orange, la densité de la loi de Gumbel, en bleu, l'histogramme des valeurs de $n^\alpha P_n$. A droite : en orange, la fonction de répartition de la loi de Gumbel, en bleu, celle de la variable simulée.

Proposition 4 [1] On suppose $X_1 \sim U(0,1)$ et on note $G_n = 1 - M_n$ le gap. Conditionnellement à G_0,

$$G_n \sim G_0 \prod_{i=1}^{n} \left(1 - \frac{U(0,1)}{i+1}\right) \quad \text{et} \quad T_n - T_{n-1} \sim Geo(G_{n-1}).$$

On obtient

$$\mathbb{E}(M_n) = 1 - \mathbb{E}(G_n) = 1 - \Theta(1/\sqrt{n}) \quad \text{et} \quad \mathbb{E}(T_n) = \Theta(n^{3/2}).$$

Au dessus de la médiane

On commence avec un employé. Si un candidat a un score inférieur à la médiane, on ne le recrute pas. Sinon, on le recrute temporairement et définitivement seulement lorsqu'un autre candidat obtient un score supérieur à la médiane (on le recrute lui aussi). On a ainsi toujours un nombre impair d'employés.

Proposition 5 On suppose $X_1 \sim U(0,1)$ et on note $G'_n = 1 - C_n$ le gap. Conditionnellement à G'_0,

$$G'_n \sim G'_0 \prod_{i=1}^{n} Beta(i+1, 1).$$

La loi Beta(i,j) est la loi de la i-ème plus petite valeur d'un échantillon de $i+j-1$ variables aléatoires indépendantes suivant $U(0,1)$, d'où l'apparition naturelle.
On obtient

$$\mathbb{E}(G'_n) = \Theta(1/n), \quad \mathbb{E}(M_n) = 1 - \Theta(\log n/n) \quad \text{et} \quad \mathbb{E}(T_n) = \Theta(n^2).$$

> Recruter au dessus de la médiane donne une équipe de qualité moyenne meilleure que recruter au dessus de la moyenne, au prix d'un recrutement plus long.

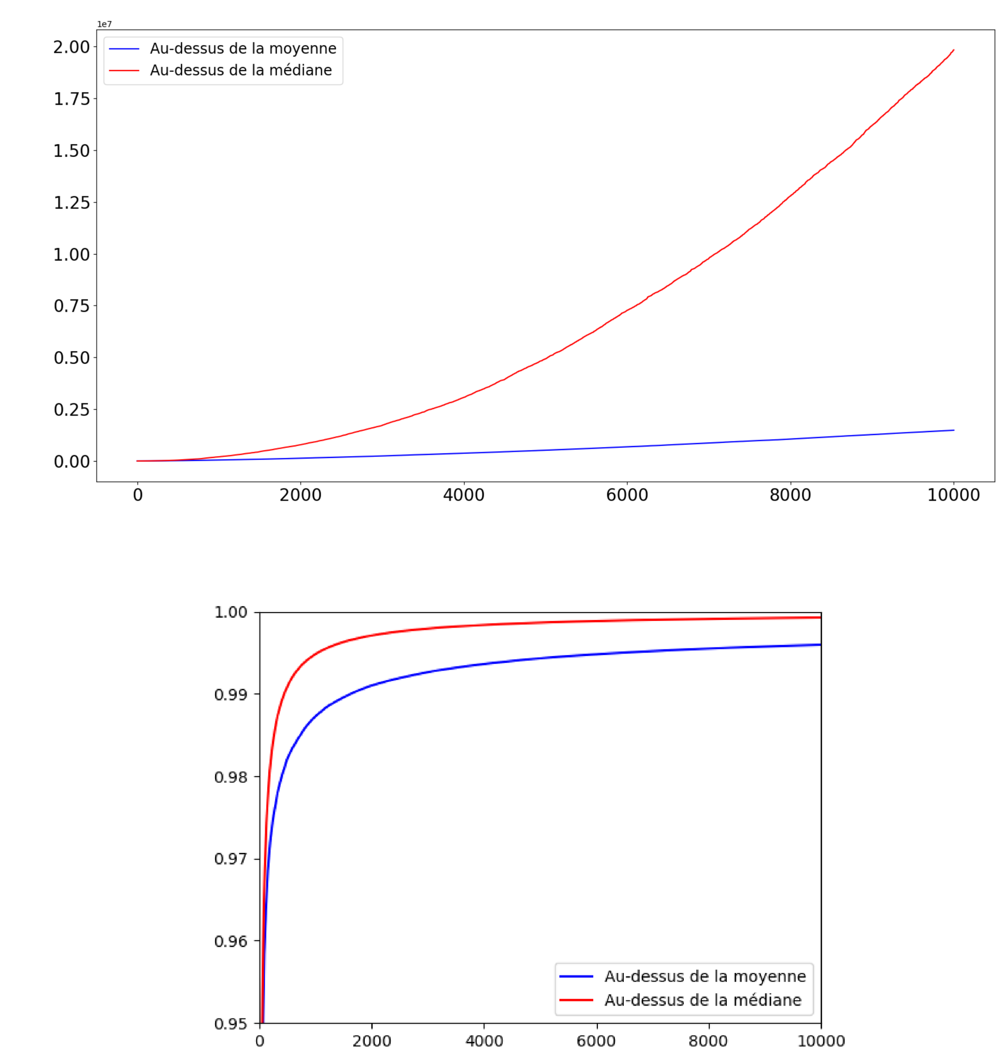

Fig. 2: T_n puis M_n en fonction de n. En bleu : stratégie au dessus de la moyenne. En rouge : stratégie au dessus de la médiane.

Références

[1] Andrei Z. Broder et Adam Kirsch. *The hiring problem and Lake Wobegon strategies.* 2009.

[2] William Feller. *An introduction to probability theory and its applications.* 1991, pp. 238–239, 518–521.

[3] Djalil Chafaï et Florent Malrieu. *Recueil de modèles aléatoires.* 2016, pp. 129–137.