Une loi faible des grands nombres pour un problème d'allocation

Erik Thomas*

Résumé

Le but de cette note est donner une loi faible des grands nombres pour un problème d'allocation.

1 Introduction

1.1 Position du problème

Dans toute cette note, les variables aléatoires sont définies sur un univers probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$.

On considère $n \in \mathbf{N}^\star$ urnes dans lesquelles on jette n boules, au hasard et de manière indépendante (chaque urne peut contenir un nombre de boules compris entre 0 et n). On note X_n le nombre d'urnes non vides.

Nous allons montrer la proposition suivante, à rapprocher de la loi faible des grands nombres.

Proposition 1.1. *On a* $\lim\limits_{n\to+\infty} \dfrac{X_n}{n} \underset{loi}{=} Z$ *où Z est une variable aléatoire déterministe égale presque sûrement égale à* $1 - 1/\mathrm{e}$.

Ce genre de résultat n'est pas vraiment surprenant. Dans [1], les auteurs montrent déjà que la suite $(X_n)_{n\in\mathbf{N}^\star}$ correctement normalisée se comporte comme une loi normale. Plus précisément, on a le théorème central limite suivant :

Proposition 1.2. *La suite* $\left(\dfrac{X_n - \mathbf{E}(X_n)}{\sqrt{n}}\right)_{n\in\mathbf{N}^\star}$ *converge en loi vers une loi normale centrée.*

La proposition 1.1 peut se réécrire comme une loi faible des grands nombres « classique » en faisant intervenir des variables de Bernoulli non indépendantes. Pour tout $i \in [\![1, n]\!]$, on pose $X_{i,n}$ la variable aléatoire valant 1 si l'urne i n'est pas vide et 0 sinon. On a clairement $X_n = \sum\limits_{i=1}^{n} X_{i,n}$ et si $i, j \in [\![1, n]\!]$ avec $i \neq j$,

$$\operatorname{cov}(X_{i,n}, X_{j,n}) = \left(1 - \left(1 - \frac{2}{n}\right)^n\right) - \left(1 - \left(1 - \frac{1}{n}\right)^n\right)^2 \neq 0.$$

La proposition 1.1 se reformule en

$$\lim_{n\to+\infty} \frac{X_{1,n} + \cdots + X_{n,n}}{n} \underset{\text{loi}}{=} 1 - 1/\mathrm{e}.$$

1.2 Simulation avec Python

On commence par simuler la variable aléatoire X_n.

```
import random as rd
import math as mt
import matplotlib.pyplot as plt
def simulation(n):
    l=[mt.floor(n*rd.random())+1 for i in range(n)]
    ll=[]
```

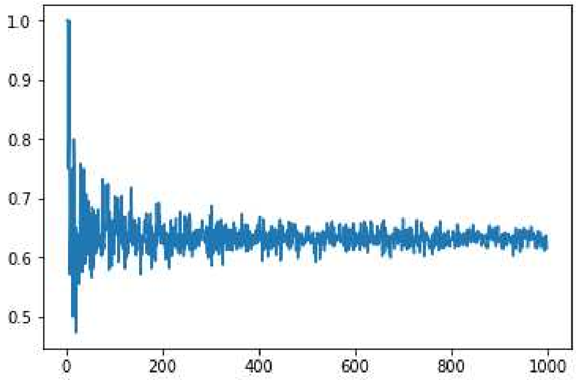
*erik.thomas@ens-rennes.fr

```
    compt=0
    for k in range(n):
        if l[k] not in ll:
            ll.append(l[k])
            compt+=1
    return compt
```

Le programme suivant permet de suivre l'évolution des sommes partielles lorsque n augmente.

```
def lfgn(n):
    x=[]
    y=[]
    for k in range(1,n):
        x.append(k)
        y.append(simulation(k)/k)
    plt.plot(x,y)
    plt.show()
```

Voici un graphe que l'on peut obtenir pour $n = 1000$.



La suite $(X_n/n)_{n\in\mathbf{N}^\star}$ semble converger effectivement vers $1 - 1/\mathrm{e} \approx 0{,}63$, mais la convergence semble lente. Peut-on alors quantifier ces oscillations ? On peut se poser la question d'une loi du log-itéré pour la suite $(X_n)_{n\in\mathbf{N}^\star}$ i.e., a-t-on

$$\limsup_{n\to+\infty} \frac{\sqrt{n}\left(\dfrac{X_n}{n} - \mathbf{E}(X_n)\right)}{\sqrt{\ln(\ln(n))}} = c \quad \text{et} \quad \liminf_{n\to+\infty} \frac{\sqrt{n}\left(\dfrac{X_n}{n} - \mathbf{E}(X_n)\right)}{\sqrt{\ln(\ln(n))}} = -c$$

presque sûrement et avec $c > 0$? La question est ouverte.

2 Preuve du résultat principal

Nous aurons besoin du résultat suivant qui précise la loi de X_n.

Lemme 2.1. *On a $X_n(\Omega) = [\![1, n]\!]$ et*

$$\forall k \in [\![1, n]\!], \quad \mathbf{P}(X_n = k) = \frac{1}{n^n}\binom{n}{k}\sum_{j=1}^{k}(-1)^{k-j}\binom{k}{j}j^n.$$

Démonstration. Soit $k \in [\![1, n]\!]$. Réaliser l'événement $(X_n = k)$ c'est choisir les k urnes non vides ($\binom{n}{k}$ façons) et choisir une surjection au hasard entre les n boules et les k urnes choisies, ainsi

$$\mathbf{P}(X_n = k) = \frac{1}{n^n}\binom{n}{k}S_{n,k},$$

où $S_{n,k}$ est le nombre de surjections entre un ensemble à n éléments et un ensemble à k éléments. Or, on a $S_{n,k} = \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n$, ce qui termine la preuve.

□

Nous allons prouver la proposition 1.1.

Démonstration. On calcule la fonction caractéristique φ_n de la variable aléatoire $\frac{X_n}{n}$. Soit $t \in \mathbf{R}$, on a :

$$\begin{aligned}
\varphi_n(t) &= \mathbf{E}\left(\mathrm{e}^{\mathrm{i}tX_n/n}\right) \\
&= \frac{1}{n^n} \sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}tk/n} \binom{n}{k} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n \\
&= \frac{1}{n^n} \sum_{k=1}^{n} \sum_{j=1}^{k} \mathrm{e}^{\mathrm{i}kt/n} \binom{n}{k} \binom{k}{j} (-1)^{k-j} j^n \\
&= \frac{1}{n^n} \sum_{j=1}^{n} \sum_{k=j}^{n} \mathrm{e}^{\mathrm{i}kt/n} \binom{n}{j} \binom{n-j}{k-j} (-1)^{k-j} j^n \\
&= \frac{1}{n^n} \sum_{j=1}^{n} j^n \binom{n}{j} \sum_{\ell=0}^{n-j} \mathrm{e}^{\mathrm{i}(j+\ell)t/n} \binom{n-j}{\ell} (-1)^{\ell} \\
&= \frac{1}{n^n} \sum_{j=1}^{n} j^n \binom{n}{j} \mathrm{e}^{j\mathrm{i}t/n} \left(1 - \mathrm{e}^{\mathrm{i}t/n}\right)^{n-j} \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(1 - \frac{k}{n}\right)^n z^{n-k} (1-z)^k
\end{aligned}$$

où l'on a posé $z := \mathrm{e}^{\mathrm{i}t/n}$. On pose aussi

$$S_n(t) := \sum_{k=0}^{n} \binom{n}{k} \left(1 - \frac{k}{n}\right)^n z^{n-k} (1-z)^k$$

et

$$V_n(t) := \sum_{k=0}^{n} \binom{n}{k} \mathrm{e}^{-k} z^{n-k} (1-z)^k = \left(z + \mathrm{e}^{-1}(1-z)\right)^n = z^n \left(1 - \mathrm{e}^{-1}\left(1 - z^{-1}\right)\right)^n.$$

En remarquant que $\varepsilon_n := \max\left\{\left|\mathrm{e}^{-k} - \left(1 - \frac{k}{n}\right)^n\right|,\ k \in [\![0, n]\!]\right\} \underset{n\to+\infty}{=} o(1)$, on a :

$$|S_n(t) - V_n(t)| \leq \varepsilon_n \sum_{k=0}^{n} \binom{n}{k} |1-z|^{n-k} = \varepsilon_n (1 + |1-z|)^n \underset{n\to+\infty}{\longrightarrow} 0$$

car $\lim_{n\to+\infty} \varepsilon_n = 0$ et $\lim_{n\to+\infty} (1 + |1-z|)^n = \mathrm{e}^{|t|}$. Or, on a :

$$V_n(t) \underset{n\to+\infty}{=} \mathrm{e}^{\mathrm{i}t} \left(1 - \left(\frac{\mathrm{i}t}{\mathrm{e}n} + o\left(\frac{1}{n}\right)\right)\right)^n \underset{n\to+\infty}{=} \mathrm{e}^{\mathrm{i}t(1-1/\mathrm{e})} + o(1).$$

On en déduit donc :

$$\lim_{n\to+\infty} \varphi_n(t) = \mathrm{e}^{\mathrm{i}t(1-1/\mathrm{e})} = \varphi_X(t)$$

où X est une variable aléatoire déterministe presque sûrement égale à $1 - 1/\mathrm{e}$. Le théorème de Lévy permet de terminer la preuve de la proposition.

□

Références

[1] V. F. Kolchin, Boris A. Sevast'ianov, Vladimir P. Christiakov. *Random Allocations.* John Wiley & Sons Inc, 1978.