Problème de Mathématiques

Référence pp1714 — Version du 31 décembre 2025

Une urne contient n boules numérotées de 1 à n. On effectue alors deux tirages avec le protocole suivant :

— On tire une première boule au hasard ;

— Si la première boule tirée porte le numéro $1 \leqslant k \leqslant n$, on la remet dans l'urne avec k nouvelles boules numérotées k et on tire une deuxième boule au hasard dans l'urne.

Par exemple, si on a tiré la boule numéro 3, on remet quatre boules portant le numéro 3 dans l'urne avant de procéder au second tirage : la boule numéro 3 qui a été extraite lors du premier tirage et trois autres boules de numéro 3. On procède alors au second tirage dans une urne qui contient $(n+3)$ boules.

On note X (resp. Y), le numéro de la boule choisie lors du premier tirage (resp. du second tirage).

Pour modéliser l'expérience aléatoire, on considère que X et Y sont deux variables aléatoires discrètes définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, à valeurs dans $E_n = \{1, \dots, n\}$.

On rappelle que toute l'information disponible sur le couple de variables aléatoires (X, Y) est contenue dans sa loi conjointe, qui est (par définition) une mesure de probabilité discrète sur l'ensemble fini $E_n \times E_n$.

1. Comment la loi conjointe d'un couple de variables aléatoires à valeurs dans E_n est-elle caractérisée habituellement ?

2. a. Pourquoi est-il légitime de supposer que la variable aléatoire X suit la loi uniforme sur E_n ?

2. b. Calculer l'espérance de X.

3. Soit $1 \leqslant k \leqslant n$.

3. a. Expliquer pourquoi il est légitime de poser

$$\forall\, \ell \in E_n \setminus \{k\}, \quad \mathbf{P}(Y = \ell \mid X = k) = \frac{1}{n+k}.$$

3. b. Comment peut-on en déduire $\mathbf{P}(Y = k \mid X = k)$?

4. En déduire une caractérisation de la loi de Y.

5. Démontrer que Y est une variable aléatoire d'espérance finie et déduire de ce qui précède que

$$\mathbf{E}(Y) = \frac{n+1}{2} \sum_{k=1}^{n} \frac{1}{n+k} + \frac{1}{n} \sum_{k=1}^{n} \frac{k^2}{n+k}.$$

En déduire un équivalent de $\mathbf{E}(Y)$ lorsque n tend vers $+\infty$.

6. Expliquer le sens du code Python suivant et démontrer que l'instruction `simulation(10)` permet d'estimer l'espérance de Y dans le cas $n = 10$.

```python
from random import randint as unif

def tirages(n):
  X = unif(1, n)
  L = list(range(1,n+1))+[X]*X
  i = unif(1, len(L))
  Y = L[i-1]
  return [X, Y]

def simulation(n, N=10**5):
  ech_XY = [tirages(n) for k in range(N)]
  ech_Y = [t[1] for t in ech_XY]
  s = 0
  for Y in ech_Y:
    s += Y
  return s/N
```

Solution ❀ Un modèle d'urne

1. La loi conjointe de (X, Y) est caractérisée par la famille

$$\big(\mathbf{P}(X=i, Y=j)\big)_{1\leqslant i,j\leqslant n}.$$

2. a. La loi de X est une mesure de probabilité sur E_n qui indique avec quelles fréquences la variable X prend ses valeurs. Comme rien, dans l'énoncé, n'indique qu'une des boules a plus de chance qu'une autre d'être tirée, il est légitime de considérer que X suit la loi uniforme sur E_n.

2. b. Comme X est une variable aléatoire bornée, elle est d'espérance finie et comme, pour tout $1 \leqslant k \leqslant n$,

$$\mathbf{P}(X=k) = \frac{1}{n}$$

alors

$$\mathbf{E}(X) = \sum_{k=1}^{n} \frac{k}{n} = \frac{n+1}{2}.$$

3. a. Comme $\mathbf{P}(X=k) = 1/n > 0$, les probabilités conditionnelles $\mathbf{P}(Y=\ell \mid X=k)$ sont bien définies.

Conditionner par $[X=k]$ revient à supposer que la variable X a pris la valeur k, c'est-à-dire qu'on a tiré la boule numérotée k lors du premier tirage. Par conséquent, on va tirer la seconde boule dans une urne qui contient $(n+k)$ boules :

— une boule numérotée i pour tout indice $1 \leqslant i \leqslant n$ différent de k ;

— et $(k+1)$ boules numérotées k.

Une fois encore, rien n'indique qu'une boule a plus de chance qu'une autre d'être tirée, donc chaque boule numérotée $i \in E_n \setminus \{k\}$ a une chance sur $(n+k)$ d'être choisie.

Autrement dit,

$$\forall\, \ell \in E_n \setminus \{k\}, \quad \mathbf{P}(Y=\ell \mid X=k) = \frac{1}{n+k}.$$

3. b. La famille $([Y=\ell])_{\ell\in E_n}$ est un système complet d'événements (puisque Y est une variable aléatoire à valeurs dans E_n) et $\mathbf{P}_{[X=k]}$ est une mesure de probabilité sur $\mathcal{A}$, donc

$$\sum_{\ell=1}^{n} \mathbf{P}_{X=k}(Y=\ell) = 1.$$

D'après la question précédente,

$$\mathbf{P}_{X=k}(Y=k) + \sum_{\ell\in E_n\setminus\{k\}} \frac{1}{n+k} = 1$$

et donc

$$\mathbf{P}(Y=k \mid X=k) = 1 - \frac{n-1}{n+k} = \frac{k+1}{n+k}.$$

4. D'après la formule des probabilités totales et les questions précédentes,

$$\begin{aligned}\forall\, \ell \in E_n, \quad \mathbf{P}(Y=\ell) = \sum_{k=1}^{n} \mathbf{P}(Y=\ell \mid X=k)\,\mathbf{P}(X=k) &= \frac{1}{n}\sum_{k\in E_n\setminus\{\ell\}} \frac{1}{n+k} + \frac{1}{n}\frac{\ell+1}{n+\ell} \\ &= \frac{\ell}{n(n+\ell)} + \frac{1}{n}\sum_{k=1}^{n}\frac{1}{n+k}\end{aligned}$$

et comme Y est une variable aléatoire à valeurs dans E_n, on sait que la famille de probabilités

$$\big(\mathbf{P}(Y=\ell)\big)_{\ell\in E_n}$$

caractérise la loi de Y.

5. Comme la variable aléatoire Y est bornée, elle est d'espérance finie et d'après ce qui précède,

$$\begin{aligned}\mathbf{E}(Y) = \sum_{\ell=1}^{n} \ell\,\mathbf{P}(Y=\ell) &= \sum_{\ell=1}^{n}\sum_{k=1}^{n} \frac{\ell}{n(n+k)} + \sum_{\ell=1}^{n} \frac{\ell^2}{n(n+\ell)} \\ &= \frac{n+1}{2}\sum_{k=1}^{n}\frac{1}{n+k} + \frac{1}{n}\sum_{\ell=1}^{n}\frac{\ell^2}{n+\ell}.\end{aligned}$$

On peut expliciter deux sommes de Riemann.

$$\mathbf{E}(Y) = \frac{n+1}{2} \cdot \frac{1}{n}\sum_{k=1}^{n} \frac{1}{1+k/n} + n \cdot \frac{1}{n}\sum_{k=1}^{n} \frac{k^2/n^2}{1+k/n}.$$

Les fonctions $\left[t \mapsto \frac{1}{1+t}\right]$ et $\left[t \mapsto \frac{t^2}{1+t}\right]$ sont continues sur le segment $[0, 1]$, donc

$$\frac{1}{n}\sum_{k=1}^{n} \frac{1}{1+k/n} \to \int_0^1 \frac{dt}{1+t} = \ln 2$$

et

$$\frac{1}{n}\sum_{k=1}^{n} \frac{k^2/n^2}{1+k/n} \to \int_0^1 \frac{t^2\,dt}{1+t} = \int_0^1 t - 1 + \frac{1}{1+t}\,dt = \ln 2 - \frac{1}{2}.$$

Par conséquent,

$$\mathbf{E}(Y) \sim \left(\frac{3}{2}\ln 2 - \frac{1}{2}\right) \cdot n$$

lorsque n tend vers $+\infty$.

6. Étude de la fonction `tirages`

On donne à `X` une valeur comprise entre 1 et n au sens large (choisie au hasard suivant la loi uniforme).

En notant k, la valeur attribuée à `X`, on construit ensuite la liste `L` égale à

$$(1, 2, \dots, k, \dots, n, \underbrace{k, \dots, k}_{k \text{ exemplaires}}).$$

La longueur de cette liste est donc égale à $n + k$.

On choisit ensuite un indice `i` compris entre 1 et $n + k$ au sens large (choisi au hasard suivant la loi uniforme) et on attribue à `Y` la valeur de `L[i-1]`.

La valeur de `Y` est donc égale à $\ell \neq k$ lorsque `i` prend la valeur $\ell + 1$ et égal à k lorsque `i` prend la valeur $k + 1$ ou une valeur comprise entre $n + 1$ et $n + k$.

Donc `Y` prend une valeur $\ell \neq k$ avec probabilité $\frac{1}{n+k}$ et prend la valeur k (la valeur de X) avec probabilité $\frac{k+1}{n+k}$.

La liste `[X,Y]` retournée par `tirages` est une simulation de (X, Y) qui est conforme à la loi conjointe du couple (X, Y) définie plus haut.

❧ Étude de la fonction `simulation`

La fonction `simulation` effectue N appels successifs à la fonction `tirages` avec $N = 10^5$ par défaut.

La liste `ech_XY` contient donc une réalisation de l'échantillon

$$\big((X_n, Y_n)\big)_{0 \leqslant n < N}$$

et la liste `ech_Y` une réalisation de l'échantillon

$$(Y_n)_{0 \leqslant n < N}$$

où les variables Y_n sont indépendantes et suivent toutes la loi de Y.

La variable `s` sert à calculer la somme des éléments de la liste `ech_Y` et la fonction `simulation` retourne le quotient de cette somme par le nombre `N` de termes, c'est-à-dire la *moyenne* de l'échantillon `ech_Y`.

Comme la variable Y est bornée, elle est de carré intégrable et la Loi des grands nombres peut s'appliquer : pour tout $\varepsilon > 0$,

$$\mathbf{P}\left(\left|\frac{Y_0 + \cdots + Y_{N-1}}{N} - \mathbf{E}(Y)\right| \geqslant \varepsilon\right) \xrightarrow[N \to +\infty]{} 0.$$

En admettant que $N = 10^5$ soit assez grand, on en déduit que la moyenne de l'échantillon retournée par la fonction `simulation` est proche de $\mathbf{E}(Y)$ avec une probabilité assez forte.

On peut rendre la discussion précédente plus précise grâce à l'inégalité de Bienaymé-Tchebychev.

Pour $n = 10$, *on a* $\mathbf{E}(Y) \approx 5,866$ *et en exécutant dix fois l'instruction* `simulation(10)`, *on obtient les résultats suivants.*

5,865	5,863	5,872	5,879	5,859
5,883	5,870	5,873	5,855	5,855

C'est probant !