# Problème de Mathématiques

Référence pp2109 — Version du 31 décembre 2025

---

Un entier $n \geqslant 2$ étant donné, on dispose d'une urne contenant $2n$ boules numérotées de 1 à $n$, chaque numéro apparaissant deux fois. On effectue une succession de tirages "au hasard" dans cette urne, en tirant à chaque fois deux boules simultanément selon le protocole suivant :

— si les deux boules tirées portent le même numéro, alors on ne remet pas les boules dans l'urne et on dit qu'une paire est reconstituée ;

— si les deux boules portent des numéros distincts, alors on les remet dans l'urne avant de procéder au tirage suivant.

**1.** Proposer un modèle probabiliste de cette expérience aléatoire construit sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ convenable.

**2.** Pour tout $1 \leqslant i \leqslant n$ et pour tout $k \in \mathbb{N}^*$, on note

$$[T_i = k]$$

le fait que la $i$-ème paire soit reconstituée lors du $k$-ième tirage. On pose alors $X_1 = T_1$ et

$$\forall\, 2 \leqslant i \leqslant n, \quad X_i = T_i - T_{i-1}.$$

**2. a.** Comment déduire la famille $(X_i)_{1\leqslant i\leqslant n}$ du modèle précédent ?

**2. b.** Cette construction des $X_i$, $1 \leqslant i \leqslant n$, permet-elle de justifier que les $X_i$ sont des variables aléatoires indépendantes ?

Quelle est alors la loi de la variable aléatoire $X_i$ ?

**2. c.** En déduire que $\mathbf{E}(T_n) = n^2$.

**3.** On considère le code Python suivant.

```python
from numpy import random as rd

def reconstitution_des_paires(n):
  T = []
  nb_tirages = 0
  nb_paires = n
  while (nb_paires>0):
    tirer_encore = True
    while tirer_encore:
      nb_tirages += 1
      a = rd.randint(nb_paires)
      b = rd.randint(nb_paires)
      if (a==b):
        T.append(nb_tirages)
        nb_paires -= 1
      tirer_encore = False
  return T
```

**3. a.** Expliquer le fonctionnement de ce code.

**3. b.** Peut-on l'utiliser pour estimer l'espérance de $T_n$ ?

## Solution ❀ Problème d'urne

**1.** Nous allons modéliser la suite des tirages successifs par une suite $(U_i)_{i\geqslant 1}$ de variables aléatoires de Bernoulli, en convenant que l'événement $[U_i = 1]$ signifie qu'une paire est reconstituée lors du $i$-ème tirage.

De cette manière, chaque variable aléatoire

$$S_i = U_1 + \cdots + U_i$$

donne le nombre de paires reconstituées au cours des $i$ premiers tirages.

❧ Pour toute liste

$$(\varepsilon_i)_{1\leqslant i\leqslant N} \in \{0,\ 1\}^N,$$

la probabilité de l'événement

$$\bigcap_{i=1}^{N} [U_i = \varepsilon_i]$$

se déduit de la Formule des probabilités composées.

Pour pouvoir utiliser cette formule, nous devons définir la valeur de la probabilité

$$\mathbf{P}(U_1 = \varepsilon_1)$$

et celles des probabilités conditionnelles

$$\mathbf{P}(U_{i+1} = \varepsilon_{i+1} \mid U_1 = \varepsilon_1, \dots, U_i = \varepsilon_i)$$

pour tout $1 \leqslant i < N$.

❧ Initialement, l'urne contient $2n$ boules et on en extrait une paire. Convenons qu'il y a $n$ boules rouges numérotées de 1 à $n$ et $n$ boules bleues numérotées de 1 à $n$. Il y a donc $\binom{2n}{2}$ tirages possibles et seuls $n$ tirages permettent de reconstituer une paire (= les tirages de la forme $\{r_k, b_k\}$ constitués d'une boule rouge et d'une boule bleue portant le même numéro).

En admettant que ces tirages soient équiprobables, nous convenons que

$$\mathbf{P}(U_1 = 1) = \frac{n}{\binom{2n}{2}} = \frac{1}{2n-1} \tag{1}$$

et (loi de Bernoulli oblige...) que

$$\mathbf{P}(U_1 = 0) = 1 - \mathbf{P}(U_1 = 1) = \frac{2n-2}{2n-1}. \tag{2}$$

❧ Après les $i$ premiers tirages, le nombre de paires reconstituées est égal à

$$s_i = \varepsilon_1 + \cdots + \varepsilon_i \tag{3}$$

et il reste donc $2(n - s_i)$ boules dans l'urne.

En supposant alors que $s_i < n$ (= l'urne n'est pas encore vide), une analyse semblable à la précédente nous porte à convenir que

$$\mathbf{P}(U_{i+1} = 1 \mid U_1 = \varepsilon_1, \dots, U_i = \varepsilon_i) = \frac{1}{2(n-s_i)-1} \tag{4}$$

et (donc !) que

$$\mathbf{P}(U_{i+1} = 0 \mid U_1 = \varepsilon_1, \dots, U_i = \varepsilon_i) = \frac{2(n-s_i)-2}{2(n-s_i)-1} \tag{5}$$

❧ On s'est ainsi donné les moyens de calculer les probabilités

$$\mathbf{P}(U_1 = \varepsilon_1, \dots, U_N = \varepsilon_N)$$

pour toute liste

$$(\varepsilon_i)_{1\leqslant i\leqslant N} \in \{0,\ 1\}^N.$$

On **admet** (ce qui est loin d'être facile à démontrer) que ces calculs sont cohérents et qu'il existe effectivement un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et une suite $(U_i)_{i\geqslant 1}$ de variables aléatoires de Bernoulli définies sur cet espace probabilisé qui vérifient les relations (1), (2), (4) et (5).

**2. a.** Les valeurs des $T_i$ indiquent à quels tirages les paires sont reconstituées.

❧ Par conséquent, $X_1 = T_1$ donne le nombre de tirages nécessaires pour reconstituer la première paire et, pour tout $2 \leqslant i \leqslant n$, la différence

$$X_i = T_i - T_{i-1}$$

donne le nombre de tirages nécessaires pour reconstituer la $(i+1)$-ième paire **après qu'on a reconstitué la $i$-ème paire**.

Il est ainsi clair que les $X_i$ prennent leurs valeurs dans $\mathbb{N}^*$.

❧ Considérons $n$ entiers

$$x_1, \dots, x_n \in \mathbb{N}^*$$

ainsi que leurs sommes cumulées :

$$t_0 = 0 \quad \text{et} \quad \forall\, 1 \leqslant i \leqslant n, \quad t_i = \sum_{k=1}^{i} x_k.$$

De cette manière,

$$\begin{cases} X_1(\omega) = x_1 \\ X_2(\omega) = x_2 \\ \quad\vdots \\ X_i(\omega) = x_i \\ \quad\vdots \\ X_n(\omega) = x_n \end{cases} \iff \begin{cases} T_1(\omega) = t_1 \\ T_2(\omega) = t_2 \\ \quad\vdots \\ T_i(\omega) = t_i \\ \quad\vdots \\ T_n(\omega) = t_n \end{cases}$$

et on en déduit que

$$[X_1 = x_1] \cap [X_2 = x_2] \cap \cdots \cap [X_n = x_n]$$

est égal à

$$\bigcap_{i=1}^{n} \big(\underbrace{[U_{t_{i-1}+1} = 0] \cap \cdots \cap [U_{t_i-1} = 0]}_{\text{échecs}} \cap \underbrace{[U_{t_i} = 1]}_{\text{i-ème succès}}\big).$$

Comme les $U_k$ sont des variables aléatoires, les $[U_k = \varepsilon_k]$ sont des événements :

$$\forall\, k \in \mathbb{N}^*,\ \forall\, \varepsilon_k \in \{0,\ 1\}, \quad [U_k = \varepsilon_k] \in \mathcal{A}.$$

Comme une tribu est stable par intersection, on en déduit que

$$\bigcap_{i=1}^{n} \big([U_{t_{i-1}+1} = 0] \cap \cdots \cap [U_{t_i-1} = 0] \cap [U_{t_i} = 1]\big) \in \mathcal{A}.$$

Comme cela est vrai quels que soient les entiers $x_1$, ..., $x_n$, on en déduit que les $X_1$, ..., $X_n$ sont des variables aléatoires :

$$\forall\, 1 \leqslant i \leqslant n,\ \forall\, x_i \in \mathbb{N}^*, \quad [X_i = x_i] \in \mathcal{A}.$$

Toute combinaison linéaire de variables aléatoires est encore une variable aléatoire, donc, pour tout $1 \leqslant i \leqslant n$,

$$T_i = \sum_{k=1}^{i} X_i$$

est une variable aléatoire sur $(\Omega, \mathcal{A})$ :

$$\forall\, 1 \leqslant i \leqslant n,\ \forall\, t_i \in \mathbb{N}^*, \quad [T_i = t_i] \in \mathcal{A}.$$

**2. b.** On a démontré à la question précédente que $X_1$, ..., $X_n$ étaient bien des variables aléatoires en constatant que l'événement

$$[X_1 = x_1] \cap \cdots \cap [X_n = x_n]$$

était égal à

$$\bigcap_{i=1}^{n} \big([U_{t_{i-1}+1} = 0] \cap \cdots \cap [U_{t_i-1} = 0] \cap [U_{t_i} = 1]\big).$$

On calcule la probabilité de cet événement en appliquant la Formule des probabilités composées, au moyen des relations (1), (2), (4) et (5).

Il faut alors remarquer (avec les notations utilisées précédemment) que

$$\forall\ t_{i-1} < k \leqslant t_i, \quad s_k = i-1$$

(après l'instant $T_{i-1}$ et jusqu'à l'instant $T_i$ inclus, la composition de l'urne est constante : on a retiré $(i-1)$ paires reconstituées au cours des tirages précédents).

On en déduit que la probabilité

$$\mathbf{P}\big([X_1 = x_1] \cap \cdots \cap [X_n = x_n]\big)$$

est égale à

$$\prod_{i=1}^{n} \Big(1 - \frac{1}{2(n-i)+1}\Big)^{(t_i-1)-t_{i-1}} \frac{1}{2(n-i)+1}$$

c'est-à-dire à

$$\prod_{i=1}^{n} \Big(1 - \frac{1}{2(n-i)+1}\Big)^{x_i-1} \frac{1}{2(n-i)+1}.$$

Cette relation prouve que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et que $X_i$ suit la loi géométrique de paramètre

$$p_i = \frac{1}{2(n-i)+1}.$$

> *En particulier, $p_n = 1$, ce qui signifie que $\mathbf{P}(X_n = 1) = 1$. C'est tout à fait normal : on reconstitue l'avant-dernière paire à l'instant $T_{n-1}$ et il ne reste alors plus que deux boules dans l'urne. Par conséquent, la dernière paire est reconstituée dès le tirage suivant, soit*
>
> $$\forall\ \omega \in \Omega, \quad T_n(\omega) = T_{n-1}(\omega) + 1$$
>
> *et donc*
>
> $$\forall\ \omega \in \Omega, \quad X_n(\omega) = 1.$$

**2. c.** Par linéarité de l'espérance,

$$\mathbf{E}(T_n) = \sum_{i=1}^{n} \mathbf{E}(X_i) = \sum_{i=1}^{n} 2(n-i)+1 = n + 2\sum_{k=0}^{n-1} k = n + n(n-1) = n^2.$$

> *Je souligne à nouveau l'absurdité d'un tel sujet (EDHEC 2011). L'espérance que nous venons de calculer dépend du modèle retenu pour décrire les tirages successifs et ce modèle n'est pas imposé par l'énoncé, il est seulement sous-entendu par la description de l'expérience aléatoire, ce qui ramène le début de ce sujet à une devinette.* ***Et moi, j'aime pas les devinettes.***

> *Je vous invite à ce propos à découvrir le paradoxe de Bertrand en version courte*
>
> `https://fr.wikipedia.org/wiki/Paradoxe_de_Bertrand`
>
> *et si le cœur vous en dit à lire un article plus long, qui a la bonne idée d'exposer l'inutilité des simulations informatiques à ce sujet.*
>
> `https://arxiv.org/pdf/1810.07805.pdf`

**3. a.** On effectue des tirages successifs dans une urne.

La fonction `reconstitution_des_paires` est constituée de deux boucles `while` imbriquées.

La variable `nb_tirages` compte le nombre de tirages effectués : elle est initialement nulle et incrémentée à chaque itération de la boucle `while` *interne*.

La variable `nb_paires` compte le nombre de paires restantes dans l'urne : il y a initialement $n$ paires de boules, soit $2n$ boules ; à chaque itération de la boucle `while` *externe*, on décrémente la variable `nb_paires`, ce qui revient à retirer deux boules de l'urne.

La boucle `while` externe continue tant qu'il y a des boules dans l'urne (soit tant que `nb_paires` est strictement positif).

La boucle `while` interne continue tant que les deux boules tirées portent des numéros différents et s'arrête quand les variables `a` et `b` prennent la même valeur. On ajoute alors à la liste `T` la valeur de nb_tirages, si bien qu'au terme du code, la liste T contient les valeurs des variables aléatoires $T_1, \dots, T_n$.

**3. b.** Non !

Dans la boucle while interne, les valeurs de $a$ et $b$ simulent des variables aléatoires indépendantes de loi uniforme sur $\{1, ..., n-i\}$ où $2(n-i)$ est le nombre de boules qui sont encore présentes dans l'urne.

Si $A$ et $B$ sont deux variables aléatoires indépendantes de loi uniforme sur $\{1, \dots, n-i\}$, alors

$$[A = B] = \bigsqcup_{k=1}^{n-i} [A = k] \cap [B = k]$$

donc

$$\mathbf{P}(A = B) = \sum_{k=1}^{n-i} \mathbf{P}(A = k)\,\mathbf{P}(B = k) = (n-i) \cdot \frac{1}{(n-i)^2} = \frac{1}{n-i}.$$

L'analyse effectuée au **2.b.** montre alors que les variables aléatoires $X_1, \dots, X_n$ sont bien indépendantes mais cette fois la variable aléatoire $X_i$ suit la loi géométrique de paramètre

$$\frac{1}{n-i+1} \quad \text{et non plus} \quad \frac{1}{2(n-i)+1}.$$

Par conséquent, avec le modèle sous-entendu par la simulation informatique,

$$\mathbf{E}(T_n) = \sum_{i=1}^{n} (n-i+1) = n + \sum_{j=0}^{n-1} j = \sum_{j=0}^{n} j = \frac{n(n+1)}{2}$$

et non pas $\mathbf{E}(T_n) = n^2$ comme on a trouvé précédemment !

❧ Le code suivant donne une estimation de $\mathbf{E}(T_n)$ en s'appuyant sur la Loi des grands nombres.

```
def estimation_espérance(n, N=1000):
  valeurs_T = []
  for i in range(N):
    T = reconstitution_des_paires(n)
    valeurs_T.append(T[-1])
```

Pour $n = 10$, la valeur retournée est effectivement proche de 55.

❧ Concrètement, le code étudié simule une expérience analogue, mais bien distincte, de celle qui est décrite par l'énoncé.

Dans l'expérience initiale, on tire deux boules parmi $2n$ et il est possible d'obtenir deux boules de la même couleur.

Selon le code, les variables aléatoires $A$ et $B$ sont indépendantes et de même loi (uniforme sur $\{1, \dots, n-i\}$), donc tout se passe comme si on disposait d'une urne avec les boules rouges et d'une autre urne avec les boules bleues et qu'on tirait une boule de chaque urne pour comparer leurs numéros.

On comprend ainsi pourquoi l'espérance de $T_n$ est plus faible avec le second modèle qu'avec le premier : il est plus facile de reconstituer des paires de cette manière !