Problème de Mathématiques

Référence pp1901 — Version du 31 décembre 2025

Toutes les variables aléatoires de ce problème sont définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$.

Partie A.

Soit $n \in \mathbb{N}^*$. On dispose de n urnes initialement vides et de N boules qu'on dépose une à une au hasard dans ces urnes. Pour $1 \leqslant k \leqslant N$, on note U_k, le numéro $(0 \leqslant U_k < n)$ de l'urne dans laquelle la k-ième boule est déposée. On note enfin X_n, le rang du premier tirage pour lequel l'une des urnes contient deux boules.

1. à quelle condition sur N la variable X_n est-elle bien définie ? Quelles sont les valeurs possibles pour cette variable ?

2. On réalise une simulation de cette expérience aléatoire au moyen du code Python suivant.

```
from numpy.random import random as random

def tirage(n):
  urnes = [0]*n
  X = 1
  choix = (int)(n*random())
  while urnes[choix]==0:
    urnes[choix] = urnes[choix]+1
    choix = (int)(n*random())
    X = X+1
  return X
```

2. a. Quelle est le type de la variable `urnes` ? Que représente-t-elle ?

2. b. Expliquer la condition d'arrêt de la boucle *while*.

2. c. D'après ce code, quelle est la loi de la famille $(U_k)_{1 \leqslant k \leqslant N}$?

3. On suppose ici que $n = 1$. Quelle est la loi de X_1 ? Que valent son espérance et sa variance ?

4. On suppose ici que $n = 2$. Quelle est la loi de X_2 ? Que valent son espérance et sa variance ?

5. On se place dans le cas général.

5. a. Justifier que

$$\forall\, 2 \leqslant k \leqslant n+1, \quad \mathbf{P}(X_n = k) = \frac{(k-1)n!}{n^k(n-k+1)!}.$$

5. b. Démontrer que la variable X_n est d'espérance finie.

5. c. Écrire en Python une fonction `esperance(n)` d'argument $n \in \mathbb{N}^*$ qui retourne l'espérance de la variable aléatoire X_n.

NB : *On s'abstiendra d'utiliser la fonction* `factorial`.

Partie B.

Pour $n \in \mathbb{N}^*$ et $m \in \mathbb{N}$, on pose

$$\alpha(n, m) = \sum_{k=0}^{m} \ell n\Big(1 - \frac{k}{n}\Big).$$

6. Démontrer que

$$\forall\, x \in [0, 1/2], \quad -x - x^2 \leqslant \ell n(1 - x) \leqslant -x.$$

7. En déduire que

$$-\frac{m(m+1)(3n+2m+1)}{6n^2} \leqslant \alpha(n, m) \leqslant -\frac{m(m+1)}{2n}$$

pour tout $n \geqslant 1$ et tout $m \leqslant n/2$.

8. On suppose dans cette question que $x > 0$ (réel fixé).

8. a. Calculer un équivalent simple de $\lfloor \sqrt{n}x \rfloor$ lorsque n tend vers $+\infty$.

8. b. Justifier l'existence d'un entier N tel que

$$\forall\, n \geqslant N, \qquad \lfloor \sqrt{n}x \rfloor \leqslant \frac{n}{2}.$$

8. c. Démontrer que, pour tout $n \geqslant 1$,

$$\forall\, 2 \leqslant k \leqslant n+1, \quad \mathbf{P}(X_n = k) = \frac{k-1}{n} \prod_{i=0}^{k-2} \Big(1 - \frac{i}{n}\Big).$$

8. d. En déduire que

$$\mathbf{P}(X_n = \lfloor \sqrt{n} x \rfloor) = \frac{\lfloor \sqrt{n} x \rfloor - 1}{n} \exp\big[\alpha(n,\ \lfloor \sqrt{n} x \rfloor - 2)\big]$$

pour tout $n \geqslant N$.

8. e. En déduire le comportement de $\mathbf{P}(X_n = \lfloor \sqrt{n} x \rfloor)$ lorsque n tend vers $+\infty$.

9. Que dire de $\mathbf{P}(X_n = \lfloor \sqrt{n} x \rfloor)$ pour $x < 0$?

Solution ❀ Tirages aléatoires

Partie A.

1. Il *suffit* que $N > n$ pour que X_n soit bien définie : dans ce cas, le nombre de boules est strictement supérieur au nombre d'urnes et quand on aura fini de tirer les N boules, il y aura nécessairement une urne qui contiendra au moins deux boules.

❧ Pour qu'une urne contienne deux boules, il faut bien tirer au moins deux boules. En tirant n boules, il se pourrait que chaque urne contienne *exactement une* boule. Mais en tirant $(n+1)$ boules, on est sûr qu'une des urnes au moins contienne au moins deux boules. Par conséquent, la variable aléatoire X_n prend des valeurs entières comprises entre 2 et $(n+1)$ au sens large.

2. a. La variable `urnes` est de type *liste*.

Initialement, elle contient n fois la valeur 0.

à chaque itération, on choisit une valeur entière $0 \leqslant i < n$ et on incrémente d'une unité le contenu de `urnes[i]` (= on dépose une boule dans l'urne i).

à l'issue de la boucle, la liste `urnes` contient donc le nombre de boules contenues dans chaque urne.

2. b. Le code s'exécute *tant qu*'on dépose des boules dans des urnes vides (ce qui se traduit par le fait que la valeur de `urnes[i]` est nulle) ; il s'arrête donc dès qu'on dépose une boule dans une urne qui n'est pas vide.

à l'issue de la boucle, la liste `urnes` ne contient donc que des 0 (pour les urnes qui sont restées vides) et des 1 (pour les urnes qui ont reçu une ou deux boules). NB : Le compteur de la première urne qui reçoit une deuxième boule n'est pas incrémenté, il reste égal à 1.

2. c. Le vecteur aléatoire $(U_k)_{1\leqslant k\leqslant N}$ est ici simulé par les valeurs successives de la variable `choix`.

Des appels itérés à la fonction `random()` ont vocation à simuler des variables aléatoires *indépendantes et de même loi*.

La fonction `random()` simule une variable aléatoire de loi uniforme sur $[0, 1[$, donc `n*random()` simule une variable aléatoire de loi uniforme sur $[0, n[$ et sa partie entière `(int)(n*random())` simule une variable aléatoire de *loi uniforme* sur l'ensemble $\{0, 1, \dots, n-1\}$.

3. Pour $n = 1$, il y a une seule urne, donc les deux premières boules sont placées dans cette urne et par conséquent X_1 est une variable aléatoire *certaine*, égale à 2. En particulier,

$$\mathbf{E}(X_1) = 2 \qquad \mathbf{V}(X_1) = 0.$$

4. Par **[1.]**, les valeurs possibles de X_2 sont 2 et 3 et

$$[X_2 = 2] = [U_1 = U_2 = 0] \sqcup [U_1 = U_2 = 1]$$

(on s'arrête dès la deuxième boule si les deux premières boules sont placées dans la même urne : l'urne 0 ou l'urne 1). D'après le modèle **[2.c.]**, les variables U_1 et U_2 sont indépendantes et suivent la loi de Bernoulli $\mathscr{B}(1/2)$. Par conséquent,

$$\mathbf{P}(X_2 = 2) = 1/2$$

et donc

$$\mathbf{P}(X_2 = 3) = 1 - \mathbf{P}(X_2 = 2) = 1/2.$$

On en déduit que

$$\mathbf{E}(X_2) = \frac{1}{2}\cdot 2 + \frac{1}{2}\cdot 3 = \frac{5}{2}$$

et comme

$$\mathbf{E}(X_2^2) = \frac{1}{2}\cdot 4 + \frac{1}{2}\cdot 9 = \frac{13}{2},$$

alors

$$\mathbf{V}(X_2) = \mathbf{E}(X_2^2) - [\mathbf{E}(X_2)]^2 = \frac{1}{4}.$$

5. a. Pour une démonstration rigoureuse, il convient d'utiliser le modèle probabiliste décrit en **[2.c.]** — à l'exclusion de toute autre justification. Il faut aussi structurer sa rédaction pour ne pas se perdre dans l'obscurité.

❧ Pour $2 \leqslant k \leqslant n+1$, on note A_k, l'ensemble des listes

$$(i_1, i_2, \dots, i_{k-1})$$

d'entiers deux à deux distincts compris entre 0 au sens large et n au sens strict (ce sont les *arrangements* de $(k-1)$ entiers parmi n). Le cardinal de A_k (ou nombre d'arrangements) est comme on sait égal à

$$\frac{n!}{[n-(k-1)]!}.$$

❧ L'événement $[X_n = k]$ est réalisé lorsque les urnes choisies pour placer les $(k-1)$ premières boules (urnes numérotées U_1, U_2, ..., U_{k-1}) sont deux à deux distinctes et que l'urne choisie pour placer la k-ième boule (numérotée U_k) a déjà été choisie.

En termes ensemblistes,

$$[X_n = k] = [U_2 \neq U_1] \cap [U_3 \notin \{U_1, U_2\}] \cap [U_4 \notin \{U_1, U_2, U_3\}] \cap \cdots \\ \cap [U_{k-1} \notin \{U_1, U_2, \dots, U_{k-2}\}] \cap [U_k \in \{U_1, U_2, \dots, U_{k-1}\}]$$

et on peut décomposer $[X_n = k]$ au moyen du système complet d'événements

$$\big([(U_1, U_2, \dots, U_k) = (i_1, i_2, \dots, i_k)]\big)_{1 \leqslant i_1, i_2, \dots, i_k \leqslant n}$$

sous la forme

$$\bigsqcup_{(i_1, \dots, i_{k-1}) \in A_k} [U_1 = i_1, \dots, U_{k-1} = i_{k-1}, U_k \in \{i_1, \dots, i_{k-1}\}].$$

Comme les variables aléatoires U_j sont indépendantes et de loi uniforme,

$$\mathbf{P}(U_1 = i_1, \dots, U_{k-1} = i_{k-1}, U_k \in \{i_1, \dots, i_{k-1}\}) = \frac{1}{n^{k-1}} \cdot \frac{k-1}{n}$$

quels que soient les indices i_1, ..., i_{k-1} et par additivité de $\mathbf{P}$, on en déduit que

$$\mathbf{P}(X_n = k) = \#(A_k) \times \frac{1}{n^{k-1}} \cdot \frac{k-1}{n} = \frac{(k-1)n!}{n^k(n-k+1)!}.$$

5. b. Comme la variable aléatoire X_n est bornée (elle ne prend qu'un nombre fini de valeurs), elle est d'espérance finie et, d'après le cours, son espérance est égale à

$$\sum_{k=2}^{n+1} k\,\mathbf{P}(X_n = k) = \sum_{k=2}^{n+1} \frac{k(k-1)n!}{n^k(n-k+1)!}.$$

5. c. Il convient d'analyser l'expression de l'espérance pour simplifier autant que possible les calculs à effectuer. Au besoin en s'inspirant de **[8.c.]**, il s'agit de calculer la somme des

$$(k+2)(k+1)\prod_{i=0}^{k}\Big(1 - \frac{i}{n}\Big)$$

pour $0 \leqslant k \leqslant n-1$ (on a décalé les indices de deux unités).

Ces produits peuvent être calculés économiquement par récurrence puisque

$$\prod_{i=0}^{k}\Big(1 - \frac{i}{n}\Big) = \Big(1 - \frac{k}{n}\Big)\prod_{i=0}^{k-1}\Big(1 - \frac{i}{n}\Big)$$

et ce produit est égal à 1 pour $k = 0$.

On en déduit alors le code suivant (où le nombre de multiplications est proportionnel à l'argument n : complexité linéaire).

```python
def esperance(n):
  espce = 0
  prod = 1
  for k in range(n):
    prod *= 1-k/n
    espce += (k+2)*(k+1)*prod/n
  return espce
```

On retrouve ainsi $\mathbf{E}(X_1) = 2$, $\mathbf{E}(X_2) = 5/2$, puis $\mathbf{E}(X_5) \approx 3,51$, $\mathbf{E}(X_{10}) \approx 4,66$, $\mathbf{E}(X_{20}) \approx 6,29\dots$

Partie B.

6. La fonction $[x \mapsto \ell n(1-x)]$ est concave sur $]-\infty, -1[$, donc

$$\forall\, x < 1, \quad \ell n(1-x) \leqslant -x$$

(comparaison à la tangente au point d'abscisse 0).

La fonction g définie par

$$g(x) = \ell n(1-x) + x + x^2$$

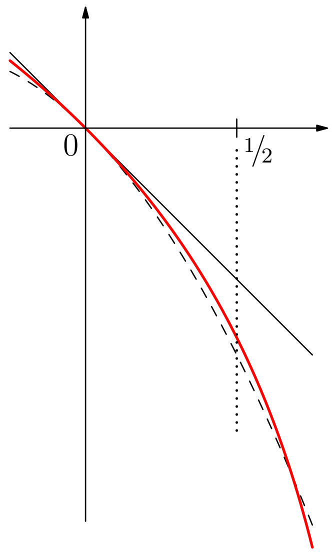
est de classe $\mathscr{C}^1$ et

$$\forall\, x < 1, \quad g'(x) = x \cdot \frac{2x-1}{1-x}.$$

On en déduit que $g'(x) \geqslant 0$ sur le segment $[0, 1/2]$ et donc que g est croissante sur cet intervalle. Or $g(0) = 0$, donc g est positive sur cet intervalle.

Finalement,

$$\forall\, x \in [0, 1/2], \quad -x - x^2 \leqslant \ell n(1-x) \leqslant -x.$$



7. Si $n \geqslant 1$ et $m \leqslant n/2$, alors

$$\forall\, 0 \leqslant k \leqslant m, \quad 0 \leqslant \frac{k}{n} \leqslant \frac{1}{2}$$

ce qui permet d'utiliser l'encadrement précédent en prenant $x = k/n$.

En sommant ces encadrements de $k = 0$ à $k = m$, on en déduit d'une part que

$$\frac{-1}{n} \cdot \frac{m(m+1)}{2} - \frac{1}{n^2} \cdot \frac{m(2m+1)(m+1)}{6} \leqslant \alpha(n, m)$$

et d'autre part que

$$\alpha(n, m) \leqslant \frac{-m(m+1)}{2n}.$$

Le résultat attendu s'en déduit par réduction au même dénominateur.

8. a. Par définition de la partie entière,

$$\sqrt{n}\, x - 1 < \lfloor \sqrt{n} x \rfloor \leqslant \sqrt{n}\, x.$$

Comme $x > 0$ est fixé, on en déduit que

$$\forall\, n \geqslant 1, \quad 1 - \frac{1}{\sqrt{n}\, x} < \frac{\lfloor \sqrt{n} x \rfloor}{\sqrt{n}\, x} \leqslant 1$$

et donc, lorsque n tend vers $+\infty$,

$$\lfloor \sqrt{n} x \rfloor \sim \sqrt{n}\, x$$

d'après le théorème d'encadrement.

8. b. D'après la question précédente,

$$\frac{\lfloor\sqrt{n}x\rfloor}{n} \sim \frac{x}{\sqrt{n}}$$

tend vers 0 lorsque n tend vers $+\infty$, donc il existe un rang N (qui dépend de x) à partir duquel le quotient est inférieur à $1/2$ (grandeur strictement supérieure à la limite). On en déduit que

$$\forall\, n \geqslant N, \quad \lfloor\sqrt{n}x\rfloor \leqslant \frac{n}{2}$$

en multipliant par $n \geqslant 1 > 0$.

8. c. Hypothèse de récurrence : il existe $2 \leqslant k \leqslant n$ tel que

$$\mathbf{P}(X_n = k) = \frac{k-1}{n}\prod_{i=0}^{k-2}\Big(1-\frac{i}{n}\Big).$$

Initialisation pour $k = 2$. D'après **[5.a.]**, $\mathbf{P}(X_n = 2) = 1/n$ et

$$\frac{2-1}{n}\prod_{i=0}^{0}\Big(1-\frac{i}{n}\Big) = \frac{1}{n}$$

Hérédité : on suppose que l'HR est vraie pour un entier $2 \leqslant k \leqslant n$. D'après **[5.a.]**,

$$\mathbf{P}(X_n = k+1) = \frac{k\cdot n!}{n\cdot n^k(n-k)!} = \frac{k(n-k+1)}{(k-1)n}\,\mathbf{P}(X_n = k)$$

et de même

$$k\prod_{i=0}^{k-1}\Big(1-\frac{i}{n}\Big) = k\Big(1-\frac{(k-1)}{n}\Big)\prod_{i=0}^{k-2}\Big(1-\frac{i}{n}\Big) = \frac{k(n-k+1)}{(k-1)n}\prod_{i=0}^{k-2}\Big(1-\frac{i}{n}\Big)$$

L'HR est donc vraie pour $(k+1)$.

L'HR est ainsi démontrée pour tout entier $2 \leqslant k \leqslant n+1$.

Il s'agit d'une récurrence sur un nombre fini de termes : puisque l'HR est héréditaire pour $2 \leqslant k \leqslant n$ et vraie en particulier pour $k = 2$, elle est vraie pour $2 \leqslant k \leqslant n+1$.

On pouvait s'abstenir de raisonner par récurrence en jonglant sur les factorielles.

8. d. Pour tout $2 \leqslant k \leqslant n+1$,

$$\mathbf{P}(X_n = k) = \frac{k-1}{n}\prod_{i=0}^{k-2}\exp \ell n\Big(1-\frac{i}{n}\Big) = \frac{k-1}{n}\exp\sum_{i=0}^{k-2}\ell n\Big(1-\frac{i}{n}\Big) = \frac{k-1}{n}\exp\alpha(n, k-2)$$

et si $n \geqslant N$, on peut appliquer en particulier cette relation pour $k = \lfloor\sqrt{n}x\rfloor$ (puisque dans ce cas, $0 \leqslant k \leqslant n/2$ d'après **[8.b.]**).

8. e. Pour $n \geqslant N$, on pose $m = \lfloor\sqrt{n}x\rfloor - 2 \leqslant n/2$ par **[8.b.]**, ce qui permet d'utiliser l'encadrement du **[7.]** D'après l'équivalent du **[8.a.]**,

$$\frac{-m(m+1)}{2n} \xrightarrow[n\to+\infty]{} \frac{-x^2}{2}$$

et

$$\frac{3n+2m+1}{3n} = 1 + \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big)$$

donc

$$\exp\big[\alpha(n,\ \lfloor\sqrt{n}x\rfloor - 2)\big] \xrightarrow[n\to+\infty]{} \exp(-x^2/2)$$

ce qui prouve que

$$\mathbf{P}(X_n = \lfloor\sqrt{n}x\rfloor) \sim \frac{x}{\sqrt{n}}\exp(-x^2/2)$$

lorsque n tend vers $+\infty$.

On peut vérifier cet équivalent au moyen du code python suivant.

```
from math import *

def prob(n, k):
  p = (k-1)/n
  for i in range(k-1):
    p *= 1-i/n
  return p

def u(n, x):
  return sqrt(n)*prob(n, round(sqrt(n)*x))

def v(x):
  return x*exp(-x**2/2)
```

On a pris soin de calculer $\mathbf{P}(X_n = k)$ *en évitant de calculer des grands entiers (qui seraient cause de débordements de capacité ou d'erreurs d'arrondi excessives). On peut constater que l'approximation est de bonne qualité pour* $1 \leqslant x \leqslant 4$.

9. Pour $x < 0$, l'entier $\lfloor \sqrt{n}x \rfloor$ est négatif, donc l'événement $[X_n = \lfloor \sqrt{n}x \rfloor]$ est impossible et sa probabilité est donc nulle.