Problème de Mathématiques

Référence pp1713 — Version du 31 décembre 2025

L'énoncé suivant est conforme au sujet original. Sans discuter l'intérêt mathématique de l'étude, on n'hésitera pas à formuler des critiques sur cet énoncé.

Un fabricant fournit des kits d'entretien composés d'un produit détartrant (produit A) et d'un produit dégraissant (produit B). Pour rendre service à ses clients, un détaillant accepte de leur vendre ces produits séparément.

On fait les hypothèses suivantes :

1. Au début de la journée, il ne reste aucune boîte entamée ;
2. Chaque client qui se présente chez le détaillent n'achète qu'un seul des deux produits ;
3. Chaque client achète le produit A avec la probabilité $0 < p < 1$ et le produit B avec la probabilité $q = 1 - p$;
4. Les choix des clients (produit A ou produit B) sont indépendants.

On note X (resp. Y), le nombre de produits A (resp. de produits B) vendus dans la journée et on pose

$$Z = \max\{X,\ Y\}.$$

1. On considère une journée où 4 clients se sont présentés.
1. a. Déterminer la loi de X, la loi de Y et les espérances de ces variables aléatoires.
1. b. Que représente la variable aléatoire Z ? Déterminer la loi de Z.

Dans la suite du sujet, on suppose que le nombre de clients qui se présentent chez le commerçant durant une journée est une variable aléatoire N qui suit la loi de Poisson de paramètre $\lambda > 0$.
2. Soit $n \in \mathbb{N}$. Quelle est la loi conditionnelle de X sachant que l'événement $[N = n]$ est réalisé ?
3. Déterminer la loi conjointe du couple (X, N).
4. En déduire la loi de X. Donner (sans calcul) les valeurs de $\mathbf{E}(X)$ et $\mathbf{E}(Y)$.
5. Démontrer que les variables aléatoires X et Y sont indépendantes.
6. En utilisant la relation $N = X + Y$, calculer $\mathbf{Cov}(X, N)$.
7. Pour $k \in \mathbb{N}$ et $x \in \mathbb{R}$, on note

$$S(k, x) = \sum_{j=0}^{k} \frac{x^j}{j!}.$$

Exprimer $\mathbf{P}(Z \leqslant k)$ en fonction de λ, de $S(k, \lambda p)$ et de $S(k, \lambda q)$.
8. Dans cette question, on utilise le langage de programmation *Python*.

On suppose que $p = 1/2$, que $\lambda = 10$ et qu'au début de la journée, le détaillant dispose de cinq kits d'entretien exactement, aucun d'eux n'étant entamé.
8. a. Proposer une fonction `S(k,x)` qui calcule la valeur de $S(k, x)$ en fonction des données k et x.
8. b. Écrire des instructions Python permettant de calculer la probabilité pour que le détaillant tombe en rupture de stock au cours de la journée.
9. Écrire une fonction `simulation(n)` en langage Python qui retourne un échantillon

$$T = \big[\big(X_k(\omega), Y_k(\omega), Z_k(\omega)\big)\big]_{0 \leqslant k < n} \in \mathfrak{M}_{n,3}(\mathbb{N})$$

de n réalisations indépendantes de l'expérience.
10. Comment utiliser le tableau T pour représenter graphiquement la loi de X ? la loi de Z ? pour estimer graphiquement $\mathbf{Cov}(X, N)$?

Solution ❀ Épuisement d'un stock

1. a.

> *On ne peut pas* déterminer *les lois de* X *et de* Y *car l'énoncé ne donne pas de modèle probabiliste ! On pourra les* déduire *d'un modèle probabiliste qu'on jugera conforme aux hypothèses de l'énoncé.*

Premier modèle probabiliste

On considère une famille $(B_k)_{1\leqslant k\leqslant 4}$ de variables aléatoires de Bernoulli, définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, indépendantes, de paramètre p.

On convient que l'événement $[B_k = 1]$ est réalisé lorsque le k-ième client achète le produit A et que l'événement contraire $[B_k = 0]$ est réalisé lorsque le k-ième client achète le produit B. Cela est conforme aux hypothèses (2.) et (3.).

D'autre part, l'indépendance des variables aléatoires B_k traduit l'hypothèse (4.).

Dans ces conditions,

$$X = \sum_{k=1}^{4} B_k,$$
$$Y = \sum_{k=1}^{4} \mathbb{1}_{[B_k=0]} = \sum_{k=1}^{4} (1 - B_k) = 4 - \sum_{k=1}^{4} B_k.$$

❧ En tant que somme de variables aléatoires indépendantes qui suivent la loi $\mathcal{B}(p)$, la fonction X est bien une variable aléatoire et suit la loi binomiale $\mathcal{B}(4, p)$.

Les variables aléatoires $1 - B_k$ sont indépendantes (lemme des coalitions) et suivent toutes la loi $\mathcal{B}(q)$ (elles sont liées aux événements contraires des variables B_k), donc la fonction Y est bien une variable aléatoire et suit la loi binomiale $\mathcal{B}(4, q)$.

D'après le cours,

$$\mathbf{E}(X) = 4p \qquad \text{et} \qquad \mathbf{E}(Y) = 4q.$$

1. b. Comme X et Y sont des variables aléatoires, la fonction Z définie par

$$Z = \max\{X, Y\}$$

est bien une variable aléatoire. D'après l'hypothèse (1.), elle compte le nombre de boîtes ouvertes par le commerçant (il n'y a pas de boîte entamée au début de la journée et on n'ouvre une nouvelle que lorsque le commerçant ne peut servir un client avec les boîtes entamées).

❧ Comme le commerçant doit servir quatre clients, la variable Z ne peut prendre que trois valeurs. On a :

- $Z = 2$ lorsque deux clients achètent le produit A (et les deux autres clients achètent le produit B) ;
- $Z = 4$ lorsque quatre clients achètent le produit A ou que quatre clients achètent le produit B (ce qui signifie qu'aucun client n'achète le produit A) ;
- $Z = 3$ dans les autres cas (lorsque trois clients achètent un même produit et qu'un seul client achète l'autre produit).

En termes booléens,

$$[Z = 2] = [X = 2] = [Y = 2]$$
$$[Z = 4] = [X = 4] \cup [Y = 4] = [X = 4] \sqcup [X = 0]$$
$$[Z = 3] = ([Z = 2] \sqcup [Z = 4]) = [X = 1] \sqcup [X = 3]$$

et comme la loi de X est connue, on en déduit que

$$\mathbf{P}(Z = 2) = 6p^2q^2,$$
$$\mathbf{P}(Z = 3) = 4pq(p^2 + q^2),$$
$$\mathbf{P}(Z = 4) = p^4 + q^4.$$

2.

> *Là encore, il est impossible de calculer la loi conditionnelle de* X *sachant* $[N = n]$ *car l'énoncé donne la loi de la variable aléatoire* N *mais s'abstient de préciser la loi du couple* (X, N) *!*

Second modèle probabiliste

On suppose que X et N sont deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et, *par analogie* avec la première étude, on *suppose* que, pour tout $n \in \mathbb{N}$, la loi conditionnelle de X sachant $[N = n]$ est la loi binomiale $\mathcal{B}(n, p)$.

En particulier, si $n = 0$,

$$\mathbf{P}(X = 0 \mid N = 0) = 1.$$

La loi de N étant connue, cette hypothèse va permettre de *calculer* la loi du couple (X, N) à la question suivante et le reste de l'étude reposera sur cette seule hypothèse.

En particulier, la variable Y sera *définie* par

$$Y = N - X.$$

Comme Y et Z sont des fonctions de N et de X, la loi du quadruplet (N, X, Y, Z) pourra se déduire de la loi du couple (X, N).

Troisième modèle probabiliste

On généralise le modèle étudié dans la première question en considèrant une suite $(B_n)_{n \geqslant 1}$ de variables aléatoires qui suivent toutes la loi de Bernoulli $\mathcal{B}(p)$ et on suppose que la famille

$$(N, B_1, \dots, B_n, \dots),$$

définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, est une famille de variables aléatoires indépendantes.

On pose alors

$$\forall\, \omega \in \Omega, \quad X(\omega) = \sum_{k=1}^{N(\omega)} B_k(\omega)$$

$$Y(\omega) = \sum_{k=1}^{N(\omega)} [1 - B_k(\omega)] = N(\omega) - X(\omega)$$

si l'entier $N(\omega)$ est supérieur à 1 et

$$X(\omega) = Y(\omega) = 0$$

si l'entier $N(\omega)$ est nul.

Avec ce modèle, il faut *démontrer* que X est bien une variable aléatoire (alors qu'avec le second moèle, on *supposait* que X était une variable aléatoire).

Comme les B_k sont des variables aléatoires, les sommes $B_1 + \cdots + B_n$ sont des variables aléatoires pour tout $n \in \mathbb{N}^*$, donc

$$\forall\, n \in \mathbb{N}^*,\ \forall\, x \in \mathbb{N}, \quad [B_1 + \cdots + B_n = x] \in \mathcal{A}.$$

Comme $\mathcal{A}$ est une tribu, elle est stable par intersection et par union dénombrable. On en déduit que, pour tout $x \in \mathbb{N}^*$,

$$[X = x] = \bigsqcup_{n \in \mathbb{N}^*} \left([B_1 + \cdots + B_n = x] \cap [N = n]\right) \in \mathcal{A}$$

et que

$$[X = 0] = [N = 0] \sqcup \left[\bigsqcup_{n \in \mathbb{N}^*} \left([B_1 + \cdots + B_n = 0] \cap [N = n]\right)\right] \in \mathcal{A}$$

donc X est bien une variable aléatoire discrète.

Comme $Y = N - X$ et que X et N sont des variables aléatoires discrètes, alors Y est aussi une variable aléatoire discrète.

❧ *Calculons* maintenant la loi conditionnelle de X sachant $[N = n]$.

Dans le cas particulier $n = 0$, on sait que

$$[N = 0] \subset [X = 0]$$

et par conséquent,

$$\mathbf{P}(X = 0 \mid N = 0) = 1.$$

Dans le cas général $n \geqslant 1$,

$$[X = x] \cap [N = n] = [B_1 + \cdots + B_n = x] \cap [N = n]$$

pour tout $x \in \mathbb{N}$. D'après le lemme des coalitions, les variables N et $B_1 + \cdots + B_n$ sont indépendantes, donc

$$\mathbf{P}(X = x \mid N = n) = \frac{\mathbf{P}(X = x \cap N = n)}{\mathbf{P}(N = n)} = \frac{\mathbf{P}(B_1 + \cdots + B_n = x)\,\mathbf{P}(N = n)}{\mathbf{P}(N = n)}$$
$$= \mathbf{P}(B_1 + \cdots + B_n = x).$$

Comme les B_k sont des variables de Bernoulli indépendantes de paramètre p, la somme $B_1 + \cdots + B_n$ suit la loi binomiale $\mathcal{B}(n, p)$. Par conséquent,

$$\forall\, 0 \leqslant x \leqslant n, \quad \mathbf{P}(X = x \mid N = n) = \binom{n}{x} p^x q^{n-x}$$

et

$$\forall\, x > n, \quad \mathbf{P}(X = x \mid N = n) = 0.$$

Le troisième modèle est donc compatible avec le second modèle puisque la loi conditionnelle de X sachant $[N = n]$ est la même dans les deux cas.

3. Par définition des probabilités conditionnelles,

$$\mathbf{P}(X = x, N = n) = \mathbf{P}(X = x \mid N = n)\,\mathbf{P}(N = n)$$

et d'après le modèle défini à la question précédente,

$$\mathbf{P}(X = x, N = n) = \left|\begin{array}{ll} e^{-\lambda}\dfrac{\lambda^n}{n!}\dbinom{n}{x} p^x q^{n-x} & \text{si } 0 \leqslant x \leqslant n, \\ \qquad\quad 0 & \text{si } x > n. \end{array}\right.$$

4. Comme N est une variable aléatoire à valeurs dans $\mathbb{N}$,

$$\forall\, x \in \mathbb{N}, \quad [X = x] = \bigsqcup_{n \in \mathbb{N}} [X = x, N = n]$$

et par σ-additivité de $\mathbf{P}$,

$$\mathbf{P}(X = x) = \sum_{n=0}^{+\infty} \mathbf{P}(X = x, N = n) = \sum_{n=x}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}\binom{n}{x} p^x q^{n-x} = \frac{e^{-\lambda}(\lambda p)^x}{x!} \sum_{n=x}^{+\infty} \frac{\lambda^{n-x} q^{n-x}}{(n-x)!}$$
$$= e^{-\lambda}\frac{(\lambda p)^x}{x!} e^{\lambda q}$$

donc X suit la loi de Poisson $\mathcal{P}(\lambda p)$.

❧ D'après le cours, l'espérance d'une variable qui suit une loi de Poisson est égale au paramètre de la loi. Donc $\mathbf{E}(X) = \lambda p$ et $\mathbf{E}(N) = \lambda$. Comme $Y = N - X$ est la différence de deux variables aléatoires d'espérance finie, elle est aussi une variable aléatoire d'espérance finie et

$$\mathbf{E}(Y) = \mathbf{E}(N) - \mathbf{E}(X) = \lambda - \lambda p = \lambda q.$$

‖ *Il est ici inutile de calculer la loi de* Y *pour en déduire* $\mathbf{E}(Y)$.

5. Soient x et y, deux entiers naturels. Comme $N = X + Y$,

$$[X = x, Y = y] = [X = x, N = x + y]$$

donc

$$\mathbf{P}(X = x, Y = y) = \frac{(\lambda p)^x (\lambda q)^y}{x!\,y!} e^{-\lambda} = \mathbf{P}(X = x)\, e^{-\lambda q}\frac{(\lambda q)^y}{y!}.$$

Comme X est une variable aléatoire à valeurs dans $\mathbb{N}$, on sait que

$$[Y = y] = \bigsqcup_{x \in \mathbb{N}} [X = x, Y = y]$$

et par σ-additivité de $\mathbf{P}$,

$$\mathbf{P}(Y = y) = \sum_{x=0}^{+\infty} \mathbf{P}(X = x, Y = y) = e^{-\lambda q} \frac{(\lambda q)^y}{y!}$$

donc Y suit la loi de Poisson $\mathcal{P}(\lambda q)$ (ce qui explique la valeur de $\mathbf{E}(Y)$ trouvée à la question précédente) et

$$\forall\, x, y \in \mathbb{N}, \quad \mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\,\mathbf{P}(Y = y)$$

donc les variables X et Y sont bien indépendantes.

> *Ce résultat est assez paradoxal, puisqu'il montre en particulier que* $\mathbf{Cov}(X, Y) = 0$ *alors que dans la première question (cas de* 4 *clients) :*
>
> $$\mathbf{Cov}(X, Y) = \mathbf{Cov}(X, 4 - X) = 4\,\mathbf{Cov}(X, 1) - \mathbf{Cov}(X, X) = -\mathbf{V}(X) < 0.$$

6. Par linéarité à droite de la covariance,

$$\mathbf{Cov}(X, N) = \mathbf{Cov}(X, X + Y) = \mathbf{Cov}(X, X) + \mathbf{Cov}(X, Y).$$

Comme X et Y sont indépendantes,

$$\mathbf{Cov}(X, N) = \mathbf{V}(X)$$

et comme X suit la loi de Poisson $\mathcal{P}(\lambda p)$,

$$\mathbf{Cov}(X, N) = \lambda p.$$

7. Comme $Z = \max\{X, Y\}$,

$$[Z \leqslant k] = [X \leqslant k] \cap [Y \leqslant k]$$

et comme les variables aléatoires X et Y sont indépendantes,

$$\mathbf{P}(Z \leqslant k) = \mathbf{P}(X \leqslant k)\,\mathbf{P}(Y \leqslant k).$$

Comme X est une variable aléatoire à valeurs dans $\mathbb{N}$,

$$[X \leqslant k] = \bigsqcup_{x=0}^{k} [X = x]$$

et comme elle suit la loi de Poisson $\mathcal{P}(\lambda p)$,

$$\mathbf{P}(X \leqslant k) = \sum_{x=0}^{k} \mathbf{P}(X = x) = e^{-\lambda p} \sum_{x=0}^{k} \frac{(\lambda p)^x}{x!}.$$

Comme Y suit la loi de Poisson $\mathcal{P}(\lambda q)$, on en déduit que

$$\mathbf{P}(Z \leqslant k) = e^{-\lambda p} S(k, \lambda p) e^{-\lambda q} S(k, \lambda q) = e^{-\lambda} S(k, \lambda p) S(k, \lambda q).$$

8. a. On va calculer la somme en utilisant les relations de récurrence suivantes :

$$\begin{aligned} S(0, x) &= 1 \\ \forall\, 0 \leqslant i < k, \quad S(i + 1, x) &= S(i, x) + u_{i+1}(x) \\ u_0(x) &= 1 \\ \forall\, 0 \leqslant i < k, \quad u_{i+1}(x) &= u_i(x) \times \frac{x}{i+1}. \end{aligned}$$

Le code Python suivant est la traduction littérale de ce qui précède.

```python
def S(k,x):
  s, u = 1, 1            # S(0,x), u_0(x)!
  for i in range(k):     # ∀ 0 ⩽ i < k
    u = u*x/(i+1)        # u_{i+1}(x)
    s = s+u              # S(i+1,x) = S(i,x) + u_{i+1}(x)
  return s               # S(k,x)
```

8. b. Il y a rupture de stock lorsque la demande excède l'offre. L'offre maximale est ici égale à 5, donc il y a rupture de stock lorsque plus de 5 clients demandent le produit A ou lorsque plus de 5 clients demandent le produit B. En d'autres termes, l'événement *Il y a rupture de stock* est égal à $[Z > 5]$.

D'après ce qui précède,

$$\mathbf{P}(Z > 5) = 1 - \mathbf{P}(Z \leqslant 5) = 1 - e^{-10}S(5,5)^2.$$

Le code suivant montre que la probabilité d'une rupture de stock est proche de 62% (ce qui est assez élevé).

```
from math import exp as exp
PrRupture = 1 - exp(-10)*S(5,5)**2
```

9. On crée un tableau $T \in \mathfrak{M}_{n,3}(\mathbb{N})$ dont tous les coefficients sont nuls.

Pour chaque ligne (`for i in range(n)`), on simule la réalisation de N (qui suit la loi de Poisson de paramètre $\lambda = 10$), puis celle de X (dont la loi conditionnelle sachant la valeur de N est connue) et on en déduit enfin les valeurs de Y et de Z.

```
import numpy as np
from numpy.random import binomial as binomiale
from numpy.random import poisson as Poisson

def simulation(n):
  p, lbd = 1/2, 10
  T = np.zeros((n, 3))
  for i in range(n):
    N = Poisson(lbd)
    X = binomiale(N, p)
    Y = N-X
    T[i,0] = X
    T[i,1] = Y
    T[i,2] = np.max([X, Y])
  return T
```

On peut procéder un peu différemment parce que les variables X et Y sont indépendantes.

```
def simulation(n):
  p, lbd = 1/2, 10
  T = np.zeros((n,3))
  for i in range(n):
    T[i,0] = Poisson(lbd*p)
    T[i,1] = Poisson(lbd*(1-p))
    T[i,2] = np.max(T[i,:2])
  return T
```

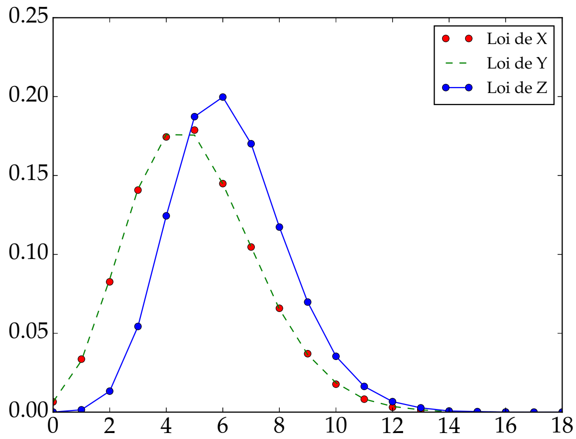
> *Si, dans le modèle probabiliste étudié, les variables X et Y n'étaient pas indépendantes, il faudrait transformer le générateur aléatoire pour tenir compte de la corrélation entre ces variables lors de la simulation (utilisation d'une* ***copule****).*

10. On réalise la simulation avec une valeur de n assez grande pour être significative.

```
T = simulation(10**5)
X, Y, Z = T[:,0], T[:,1], T[:,2]
N = X+Y
```

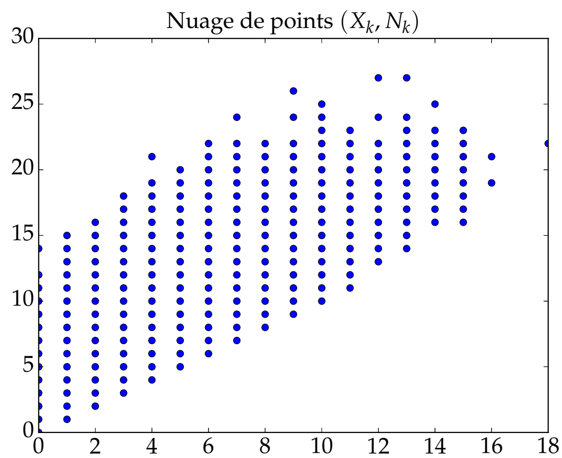
Pour représenter graphiquement les lois de X et de Z, le plus simple est de tracer un histogramme.

```
import matplotlib.pyplot as plt
b = list(range(20))
x, t = np.histogram(X, bins=b, density=True)
plt.plot(t[:-1], x, 'r-o', label="Loi de X")
z, t = np.histogram(Z, bins=b, density=True)
plt.plot(t[:-1], z, 'b-o', label="Loi de Z")
plt.ylim(0, 0.25)
plt.legend()
```



Pour estimer la corrélation des variables X et N, le plus simple est de tracer le nuage de points (X_k, N_k) pour $0 \leqslant k < n$.

```
plt.plot(X, N, 'o')  # Nuage de points
plt.ylim(0, 30)
plt.title("Nuage de points $(X_k, N_k)$", size=20)
```

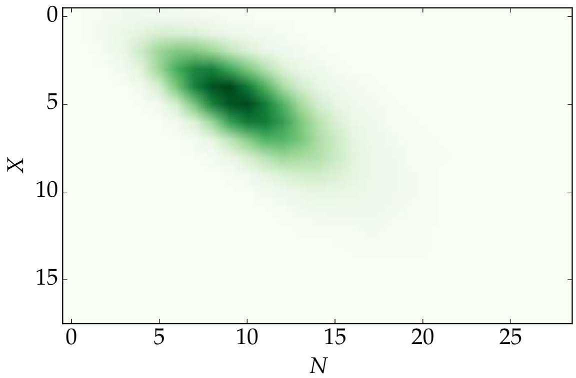


On voit bien que N a tendance à être une fonction croissante de X, ce qui justifie d'avoir trouvé une covariance strictement positive.

L'inconvénient majeur de cette méthode est d'accorder la même importance à tous les points du nuage. Il ne faut pas oublier que ce graphique présente n points et que, pour nous, $n = 10^5$! Certains points doivent donc être comptés beaucoup plus que d'autres.

L'idée est alors de compter chaque point et d'attribuer une couleur plus intense aux points les plus fréquemment rencontrés. On a besoin pour cela d'une **palette de couleurs** (*color map*) : la palette `Greens` attribue des verts très pâles (presque blancs) aux plus faibles valeurs et des verts très sombres aux plus grandes valeurs.

```
Xmax = (int)(np.max(X))
Nmax = (int)(np.max(N))
F = np.zeros((Xmax+1, Nmax+1))
# On compte le nombre d'occurrences de chaque point du nuage
for x, n in zip(X, N):
  F[x,n] += 1
plt.imshow(F, cmap='Greens')
plt.xlabel("N")
plt.ylabel("X")
```



L'inconvénient majeur de cette méthode tient à la commande `imshow` : pour des raisons techniques, la première coordonnée est portée sur l'axe vertical qui est orienté en décroissant et la seconde coordonnée est portée sur l'axe horizontal (qui est orienté, comme d'habitude, de la gauche vers la droite).

Il faut donc réfléchir un peu avant de conclure que, visiblement, N a tendance à varier *dans le même sens* que X : il est normal que la covariance soit strictement positive.