Banque « Agro-véto »
C — 0521

Mathématiques

Durée : 3 heures

Rappel : **l'usage de la calculatrice est autorisé pour cette épreuve.**

Chaque candidat est responsable de la vérification de son sujet d'épreuve : pagination et impression de chaque page. Ce contrôle doit être fait en début d'épreuve. En cas de doute, il doit alerter au plus tôt le surveillant qui vérifiera et, éventuellement, remplacera le sujet.

Ce sujet est constitué de deux problèmes totalement indépendants et comporte 8 pages numérotées de 1 à 8.

Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.

Premier problème : pseudo-solution d'un système linéaire

Ce problème propose d'étudier des solutions approchées de systèmes linéaires.
Les parties sont, dans une large mesure, indépendantes.

Partie A : position du problème et notations

1. **Question préliminaire :**

   (a) Montrer que le système : $\left\{\begin{array}{ccccc} x & + & 2y & = & 0 \\ 2x & - & y & = & 1 \\ 3x & + & y & = & \dfrac{3}{2} \end{array}\right.$ d'inconnue $(x,y) \in \mathbb{R}^2$ ne possède pas de solution.

   (b) Déterminer le triplet $(a,b,c) \in \mathbb{R}^3$ pour lequel le couple $\left(\dfrac{1}{2}, -\dfrac{1}{6}\right)$ est solution du système : $\left\{\begin{array}{ccccc} x & + & 2y & = & a \\ 2x & - & y & = & b \\ 3x & + & y & = & c \end{array}\right.$ d'inconnue $(x,y) \in \mathbb{R}^2$.

Cet exemple illustre le fait qu'un système sans solution peut tout de même posséder des solutions « proches » du résultat voulu. C'est cette notion qui sera précisée et étudiée dans ce problème.

Rappels et notations

Dans tout le problème, n désigne un entier naturel non nul. Tout système linéaire de la forme : $\left\{\begin{array}{ccccc} u_1 x & + & v_1 y & = & b_1 \\ u_2 x & + & v_2 y & = & b_2 \\ \vdots & \vdots & \vdots & = & \vdots \\ u_n x & + & v_n y & = & b_n \end{array}\right.$ d'inconnue $(x,y) \in \mathbb{R}^2$ où $u_1, \ldots, u_n, v_1, \ldots, v_n \in \mathbb{R}$ peut se ré-écrire matriciellement sous la forme : $AX = B$, avec :

$$X = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{M}_{2,1}(\mathbb{R}), \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R}) \quad \text{et} \quad A = \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \\ \vdots & \vdots \\ u_n & v_n \end{pmatrix} \in \mathcal{M}_{n,2}(\mathbb{R}).$$

- **Transposition :**
  Pour tout $A \in \mathcal{M}_{n,p}(\mathbb{R})$, on note ${}^tA \in \mathcal{M}_{p,n}(\mathbb{R})$ la matrice transposée de A.
  Par ailleurs, on rappelle que l'application « transposition » est linéaire.
  De plus, pour tous $A \in \mathcal{M}_{n,p}(\mathbb{R})$ et $B \in \mathcal{M}_{p,q}(\mathbb{R})$:
  $${}^t(AB) = {}^tB\,{}^tA.$$

- **Produit scalaire :**
  Pour tout $n \in \mathbb{N}$, l'espace vectoriel $\mathbb{R}^n$ est muni de son *produit scalaire canonique*, défini pour tous $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$ par : $< X|Y >= {}^tXY = \sum_{k=1}^{n} x_i y_i.$

- **Norme :**
  Pour tout $X \in \mathbb{R}^n$, on pose : $\|X\| = \sqrt{< X|X >}$, norme du vecteur X.
  En particulier, le nombre : $\|X - Y\|$ est la distance entre X et Y.

> **Définition (Pseudo-solution d'un système linéaire).**
> Soient $A \in \mathcal{M}_{n,2}(\mathbb{R})$ et $B \in \mathcal{M}_{n,1}(\mathbb{R})$.
> On dit que $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathcal{M}_{2,1}(\mathbb{R})$ est une *pseudo-solution* du système linéaire : $AX = B$ d'inconnue $X \in \mathcal{M}_{2,1}(\mathbb{R})$ si pour tout $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{M}_{2,1}(\mathbb{R})$:
> $$\left\| A \begin{pmatrix} x \\ y \end{pmatrix} - B \right\| \geq \left\| A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - B \right\|.$$

En d'autres termes, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ est une pseudo-solution du système si $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ minimise la distance entre B et l'image de A.

Partie B : Étude d'un exemple en dimension 2

Dans cette partie, $n = 2$ et on pose $A = \begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix}$. On s'intéresse au système :

$$(\mathcal{E}) : \left\{ \begin{array}{rcrcr} x & + & 3y & = & 2 \\ x & + & 3y & = & -1 \end{array} \right.$$

d'inconnue $(x, y) \in \mathbb{R}^2$.

2. Justifier que le système $(\mathcal{E})$ n'admet pas de solution.

Dans le plan muni d'un repère orthonormé $(O; \vec{i}, \vec{j})$, on considère le point $B(2; -1)$ et on note $\mathcal{D}$ la droite vectorielle dirigée par le vecteur $\vec{u} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

3. Sur un dessin, tracer la droite $\mathcal{D}$ et placer le point B.

On note $H(x_H; y_H)$ le projeté orthogonal de B sur $\mathcal{D}$.

4. (a) Placer le point H sur le dessin.
   (b) Donner une équation cartésienne de $\mathcal{D}$.
   (c) Déterminer les coordonnées (x_H, y_H) du point H.
   On pourra remarquer que : $H \in \mathcal{D}$ *et* $\overrightarrow{HB} \perp \vec{u}$.
   (d) Résoudre le système : $\left\{ \begin{array}{rcrcl} x & + & 3y & = & x_H \\ x & + & 3y & = & y_H \end{array} \right.$ d'inconnue $(x, y) \in \mathbb{R}^2$.
5. (a) Soit $(x, y) \in \mathbb{R}^2$. Calculer le produit matriciel $A \begin{pmatrix} x \\ y \end{pmatrix}$.
   En déduire que le point de coordonnées $A \begin{pmatrix} x \\ y \end{pmatrix}$ appartient à $\mathcal{D}$.
   (b) Donner le théorème qui permet d'affirmer que pour tout $(x, y) \in \mathbb{R}^2$:
   $$\left\| A \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\| \geq HB,$$
   où HB désigne la distance entre H et B et expliquer cette inégalité.
6. En déduire la pseudo-solution du système $(\mathcal{E})$.

Partie C : étude du cas général

Soient $A \in \mathcal{M}_{n,2}(\mathbb{R})$ et $B \in \mathcal{M}_{n,1}(\mathbb{R})$. On note $(\mathcal{S})$ le système : $\quad AX = B \quad$ d'inconnue $X \in \mathcal{M}_{2,1}(\mathbb{R})$. Le but de cette partie est de montrer le théorème suivant :

> **Théorème :** Si : $\quad rg(A) = 2, \quad$ alors :
> (i) Le système $(\mathcal{S})$ possède une unique pseudo-solution.
> (ii) $H \in \mathcal{M}_{2,1}(\mathbb{R})$ est une pseudo-solution de $(\mathcal{S})$ si seulement si : $\quad {}^tAAH = {}^tAB$.

Dans cette partie, on suppose : $\quad rg(A) = 2$.

7. (a) Énoncer le théorème du rang.
   (b) En déduire $\ker A$.
   (c) Montrer que pour tout $X \in \mathcal{M}_{2,1}(\mathbb{R})$: $\quad {}^tAAX = 0 \implies X = 0$.
   On pourra remarquer que : $\quad {}^tAAX = 0 \implies \underbrace{{}^tX{}^tAAX}_{=\|AX\|^2} = 0$.
   (d) En déduire que ${}^tAA$ est inversible.
   (e) Montrer que le système : $\quad {}^tAAX = {}^tAB \quad$ d'inconnue $X \in \mathcal{M}_{2,1}(\mathbb{R})$ possède une unique solution.

On note $\begin{pmatrix} x_H \\ y_H \end{pmatrix}$ la solution du système : $\quad {}^tAAX = {}^tAB \quad$ d'inconnue $X \in \mathcal{M}_{2,1}(\mathbb{R})$.

8. (a) Montrer que pour tous $X, Y \in \mathcal{M}_{n,1}(\mathbb{R})$:
$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2 < X|Y > \qquad (\bigstar).$$
   Indication : on pourra développer $< X - Y|X - Y >$ *par bilinéarité.*
   (b) Montrer que pour tout $(x, y) \in \mathbb{R}^2$: $\left\langle A\begin{pmatrix} x_H \\ y_H \end{pmatrix} - B \middle| A\begin{pmatrix} x - x_H \\ y - y_H \end{pmatrix} \right\rangle = 0$.
   (c) En déduire que pour tout $(x, y) \in \mathbb{R}^2$:
$$\left\| A\begin{pmatrix} x \\ y \end{pmatrix} - B \right\|^2 = \left\| A\begin{pmatrix} x_H \\ y_H \end{pmatrix} - B \right\|^2 + \left\| A\begin{pmatrix} x \\ y \end{pmatrix} - A\begin{pmatrix} x_H \\ y_H \end{pmatrix} \right\|^2.$$
   Indication : on pourra écrire :
$$A\begin{pmatrix} x \\ y \end{pmatrix} - B = A\begin{pmatrix} x_H \\ y_H \end{pmatrix} - B + A\begin{pmatrix} x \\ y \end{pmatrix} - A\begin{pmatrix} x_H \\ y_H \end{pmatrix}$$
   et utiliser $(\bigstar)$.
   (d) En déduire que (x_H, y_H) est l'unique pseudo-solution du système $(\mathcal{S})$.

Partie D : méthode de régression des moindres carrés

Supposons qu'on dispose de données expérimentales $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, pour lesquelles les x_i ne sont pas tous égaux entre eux. Ces données sont supposées suivre un modèle, qui prédit que pour tout $i \in [\![1; n]\!]$: $\quad y_i = mx_i + p \quad$ pour des valeurs de m et p fixées à l'avance. Le but de cette partie est de chercher une estimation des paramètres m et p.

Pour tout $i \in [\![1, n]\!]$, on note M_i le point de coordonnées (x_i, y_i).

9. Montrer que, par un choix judicieux de $A \in \mathcal{M}_{n,2}$ et $B \in \mathcal{M}_{n,1}$, exprimés en fonction de $(x_i)_{i\in\{1,\dots,n\}}$ et $(y_i)_{i\in\{1,\dots,n\}}$, ce problème se ramène à la résolution du système $(\mathcal{R})$: $AX = B$ d'inconnue $X = \begin{pmatrix} m \\ p \end{pmatrix} \in \mathcal{M}_{2,1}(\mathbb{R})$.

10. À quelle condition sur les points $(M_i)_{i\in\{1,\dots,n\}}$ le système $(\mathcal{R})$ admet-il une solution ?

La condition n'étant pas satisfaite en général, on note $(\hat{m}, \hat{p})$ la pseudo-solution du système $(\mathcal{R})$, et on pose :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \gamma = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad \text{et} \quad \delta = \frac{1}{n}\sum_{i=1}^{p} x_i y_i.$$

11. (a) Montrer que la pseudo-solution $(\hat{m}, \hat{p})$ du système vérifie :

$$\left\{\begin{array}{ccccc} \hat{m}\gamma & + & \hat{p}\bar{x} & = & \delta \\ \hat{m}\bar{x} & + & \hat{p} & = & \bar{y} \end{array}\right.$$

On pourra commencer par calculer ${}^tAA$, puis utiliser le point (ii) du théorème démontré à la partie C.

(b) On admet que $\gamma \neq \bar{x}^2$.
Déterminer une expression de $\hat{m}$ et $\hat{p}$ en fonction de $\bar{x}$, $\bar{y}$, γ et δ.

Définition : On appelle *droite de régression par la méthode des moindres carrés* la droite d'équation : $\quad y = \hat{m}x + \hat{p}$.

Partie E : application en biologie

Dans une boîte de Pétri, on met en culture des bactéries.

Pour tout $t \geq 0$, on note $N(t)$ le nombre de bactéries par millilitre à l'instant t.

Des mesures du nombres $N_i = N(t_i)$ de bactéries sont effectuées à divers instants t_i. On obtient le tableau suivant ou $\ln(N_i)$ désigne le logarithme népérien de N_i.

Valeur de i	1	2	3	4	5	6	7
t_i en heures	0	0,5	1	1,5	2	2,5	3
$y_i = \ln(N_i)$	9,15	9,30	9,38	9,50	9,65	9,72	9,85

12. Déterminer, à l'aide de la question 11b et en expliquant votre démarche, la droite de régression par la méthode des moindres carrés du nuage de points $(t_i, y_i)_{i\in\{1,2,\dots,7\}}$.

Le nombre $N(t)$ vérifie l'équation différentielle (E) : $\quad N'(t) = \lambda N(t), \quad$ pour un certain paramètre $\lambda \in \mathbb{R}$ (inconnu). Notons N_0 le nombre de bactéries par millilitre au temps $t = 0$.

13. (a) Résoudre l'équation (E), et exprimer $N(t)$ en fonction de t, λ et N_0.

(b) En déduire, à l'aide de la question 12, une approximation des paramètres λ et N_0.

Second problème : contrôle qualité

Les parties A et B sont totalement indépendantes.
Les résultats de la partie B pourront être admis pour traiter la partie C.

Partie A : la loi géométrique tronquée

Soient $N \in \mathbb{N}^*$ un entier naturel non nul et $p \in]0,1[$.

1. **Préliminaires :** On note f la fonction $x \mapsto 1 + x + x^2 + \cdots + x^{N-1} = \displaystyle\sum_{k=0}^{N-1} x^k$ sur $\mathbb{R}$.
   (a) Montrer par récurrence que pour tous $N \in \mathbb{N}^*$ et $x \in]0,1[$: $\displaystyle\sum_{k=0}^{N-1} x^k = \frac{1-x^N}{1-x}$.
   (b) Justifier que f est dérivable sur $]0,1[$.
   (c) En déduire pour tout $x \in]0,1[$ une expression sous la forme d'un quotient de :
   $$\sum_{k=0}^{N-1} kx^{k-1} = 1 + 2x + 3x^2 + \cdots + (N-1)x^{N-2}.$$

Un fabricant met sur le marché des composants électroniques. Chaque jour, un composant peut tomber en panne avec probabilité p, et a une durée de fonctionnement maximale de N jours (obsolescence programmée : au $(N+1)^{\text{ième}}$ jour, le composant est nécessairement en panne). On note V la variable aléatoire donnant la durée de fonctionnement d'un composant.

2. **Loi de V :**
   (a) Montrer que pour tout $k \in [\![1, N-1]\!]$: $\mathbb{P}[V = k] = (1-p)^{k-1}p$.
   (b) Justifier que : $\mathbb{P}[V = N] = (1-p)^{N-1}$.
   (c) Vérifier que : $\displaystyle\sum_{k=1}^{N} \mathbb{P}[V = k] = 1$.
   (d) Montrer que pour tout $r \in [\![1, N]\!]$: $\mathbb{P}[V \geq r] = (1-p)^{r-1}$.

   On note alors : $V \sim \mathcal{G}(N,p)$, et on dit que V suit une loi *géométrique tronquée de paramètres N et p.*

   (e) Montrer finalement que : $\mathbb{E}[V] = \dfrac{1}{p}\big(1 - (1-p)^N\big)$.
3. **Application :**
   (a) Le fabricant assure une durée de vie maximale de $N = 100$ jours avec une durée de vie moyenne de 10 jours. Déterminer à la calculatrice une valeur approchée de p arrondie au dixième. *Vous expliquerez brièvement votre démarche.*
   (b) Un organisme de contrôle indépendant décide de faire une étude de ces composants, et vérifie, au dixième jour, le fonctionnement de 1000 d'entre eux : 44% sont encore en état de marche. Ce résultat est-il surprenant ? *Vous justifierez votre réponse.*

Partie B : des résultats intermédiaires

Les trois questions de cette partie sont indépendantes.

4. **Inégalité de Markov :** Soit X une variable aléatoire réelle prenant un nombre fini de valeurs :
$$0 < x_1 < x_2 < \cdots < x_n.$$
On suppose que l'une au moins des valeurs est supérieure ou égale à 1 et on pose :
$$m = \min\{k \in [\![1,n]\!] \mid \quad x_k \geq 1\},$$
de sorte que :
$$0 < x_1 < x_2 < \cdots < x_{m-1} < 1 \leq x_m < \cdots < x_n.$$
   (a) Exprimer $\mathbb{P}[X \geq 1]$ et $\mathbb{E}[X]$ en fonction de $\big(\mathbb{P}[X = x_k]\big)_{k\in[\![1,n]\!]}$.
   (b) En déduire l'inégalité de Markov : $\quad \mathbb{P}[X \geq 1] \leq \mathbb{E}[X]$.
5. **Fonction génératrice :** Soient $n \in \mathbb{N}^*$, $q \in]0,1[$ et $t \geq 0$. On considère S_n une variable aléatoire binomiale de paramètres n et q, et on pose : $\quad Y_n = e^{tS_n}$.
   (a) Justifier que pour tout $k \in [\![0,n]\!]$: $\quad \mathbb{P}[Y_n = e^{tk}] = \binom{n}{k} q^k (1-q)^{n-k}$.
   (b) Montrer que : $\quad \mathbb{E}[Y_n] = \displaystyle\sum_{k=0}^{n} e^{tk} \binom{n}{k} q^k (1-q)^{n-k}$.
   (c) En déduire finalement que : $\quad \mathbb{E}[Y_n] = (1 - q + qe^t)^n$.
6. Soient $\varepsilon \in]0,1[$ et $q \in]0, 1-\varepsilon[$.
On note $g : t \mapsto t(q+\varepsilon) - \ln(1 - q + qe^t)$ définie sur $[0,+\infty[$.
   (a) Exprimer $g'(t)$ sur $[0,+\infty[$.
   (b) Résoudre algébriquement l'inéquation : $\quad g'(t) \geq 0 \quad$ d'inconnue $t \in [0,+\infty[$.
   On note α la solution de l'équation $g'(t) = 0$ et on admet que
$$g(\alpha) = (q+\varepsilon)\ln\frac{q+\varepsilon}{q} + (1-q-\varepsilon)\ln\frac{1-q-\varepsilon}{1-q}.$$
   (c) Dresser le tableau de variation de g (les limites ne sont pas demandées).
   On pourra exprimer des valeurs en fonction de α et $g(\alpha)$.
   (d) En déduire que $g(\alpha) > 0$ et que pour tout $t \geq 0$:
$$g(t) \leq (q+\varepsilon)\ln\frac{q+\varepsilon}{q} + (1-q-\varepsilon)\ln\frac{1-q-\varepsilon}{1-q}. \tag{1}$$

Partie C : estimation des grandes déviations

Revenons au problème du contrôle qualité.

On pose : $\quad n = 1000 \quad$ et pour tout $i \in [\![1,n]\!]$ on appelle V_i la variable aléatoire correspondant à la durée de vie du i-ème composant. $(V_i)_{i\in[\![1,n]\!]}$ est donc une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{G}(100; 0{,}1)$.

Pour tout $i \in [\![1,n]\!]$, on pose :
$$Z_i = \begin{cases} 0 & \text{si } V_i < 10 \\ 1 & \text{si } V_i \geq 10 \end{cases} \qquad \text{et} \qquad S_n = \sum_{k=1}^{n} Z_i.$$

7. (a) Déterminer la loi de Z_i pour tout $i \in [\![1, n]\!]$.
   (b) En déduire que S_n suit une loi binomiale de paramètres n et q, pour un certain $q \in]0, 1[$ qu'on précisera.
8. En appliquant l'inégalité de Markov (question 4b) à une variable aléatoire judicieusement choisie, montrer que pour tous $t > 0$ et $\varepsilon > 0$:

$$\mathbb{P}\left[\frac{S_n}{n} \geq q + \varepsilon\right] \leq \mathbb{E}\left[e^{t(S_n - n(q+\varepsilon))}\right].$$

9. En utilisant la question 5c, justifier que pour tous $t > 0$ et $\varepsilon > 0$:

$$\mathbb{E}\left[e^{t(S_n - nq - n\varepsilon)}\right] = \exp\big(-ng(t)\big).$$

   où g est la fonction définie en 6.
10. En déduire finalement qu'il existe un réel $h > 0$, qu'on déterminera, pour lequel :

$$\mathbb{P}\left[\frac{S_n}{n} \geq q + \varepsilon\right] \leq e^{-nh}.$$

   On pourra utiliser la question 6d.
11. Commentez les résultats de l'étude.

— Fin du sujet —